

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO, CONTABILIDADE E
ATUÁRIA
DEPARTAMENTO DE CONTABILIDADE E ATUÁRIA

Rafael Tavares Luz

**Aplicação de modelagem de preços hedônicos ao mercado de hambúrgueres congelados
de São Paulo**

São Paulo
2023

Prof. Dr. Carlos Gilberto Carlotti Júnior
Reitor da Universidade de São Paulo

Profa. Dra. Maria Dolores Montoya Diaz
Diretora da Faculdade de Economia, Administração, Contabilidade e Atuária

Profa. Dra. Mara Jane Contrera Malacrida
Chefe do Departamento de Contabilidade e Atuária

Prof. Dr. Renê Coppe Pimentel
Coordenador do Programa de Pós-Graduação em Controladoria e Contabilidade

RAFAEL TAVARES LUZ

Aplicação de modelagem de preços hedônicos ao mercado de hambúrgueres congelados de
São Paulo

Versão original

Dissertação de Mestrado apresentada ao Departamento de Contabilidade e Atuária da Faculdade de Economia, Administração, Contabilidade e Atuária da Universidade de São Paulo como requisito parcial para a obtenção do título de Mestre em Ciências.

Orientador: Luiz Paulo Lopes Fávero

São Paulo
2023

Catálogo na Publicação (CIP)
Ficha Catalográfica com dados inseridos pelo autor

Luz, Rafael Tavares.

Aplicação de modelagem de preços hedônicos ao mercado de hambúrgueres congelados de São Paulo / Rafael Tavares Luz. - São Paulo, 2023.

94 p.

Dissertação (Mestrado) - Universidade de São Paulo, 2023.

Orientador: Luiz Paulo Lopes Fávero.

1. Mercado de alimentos – São Paulo (cidade). 2. Preços hedônicos. I. Universidade de São Paulo. Faculdade de Economia, Administração, Contabilidade e Atuária. II. Título.

Nome: Luz, Rafael Tavares

Título: Aplicação de modelagem de preços hedônicos ao mercado de hambúrgueres congelados de São Paulo

Dissertação de Mestrado apresentada ao Departamento de Contabilidade e Atuária da Faculdade de Economia, Administração, Contabilidade e Atuária da Universidade de São Paulo como requisito parcial para a obtenção do título de Mestre em Ciências.

Aprovado em: ____ / ____ / ____

Banca Examinadora

Prof. Dr.: _____

Instituição: _____

Julgamento: _____

Profa. Dra.: _____

Instituição: _____

Julgamento: _____

Prof. Dr.: _____

Instituição: _____

Julgamento: _____

Aos meus pais, por todo amor e cuidado.

À Isabela, minha companheira de vida.

AGRADECIMENTOS

Ao professor Dr. Luiz Paulo Lopes Fávero, por toda sabedoria, conhecimento e absoluta generosidade em todas as trocas durante a construção deste trabalho. É uma inspiração tê-lo como orientador.

À professora Dra. Gabriela Scur e aos professores Dr. Nuno Manoel Martins Dias Fouto e Dr. Robson Braga pelas riquíssimas recomendações e contribuições compartilhadas no exame de qualificação desta dissertação.

À Conceição, por todas as conversas significativas e pela ajuda ao longo de todos esses anos.

Ao Renato, pelo incentivo e pela essencial ajuda no meu processo de descoberta e autoconhecimento.

À professora Adriana Silva, por sua paixão contagiante pelo conhecimento e pelo ensino.

Ao Tomás, pelo apoio e incentivo a retornar ao mundo acadêmico e buscar meus sonhos.

Aos meus familiares e amigos, por todo o apoio, companheirismo e diversão proporcionada ao longo do caminho.

Aos meus gatos Oliver, Tavinho e Matilda, pela companhia e pela leveza que trazem todos os dias.

RESUMO

Luz, R. T. (2023). *Aplicação de modelagem de preços hedônicos ao mercado de hambúrgueres congelados de São Paulo* (Dissertação de Mestrado). Faculdade de Economia, Administração, Contabilidade e Atuária, Universidade de São Paulo, São Paulo.

Este trabalho estuda o mercado de hambúrgueres congelados da região metropolitana de São Paulo a partir da aplicação da modelagem de preços hedônicos desenvolvida por Rosen. Tal modelagem apoia-se na Teoria dos Atributos de Lancaster, linha de pesquisa da Microeconomia que propõe o estudo das relações de consumo por meio das características dos bens e a utilidade que elas proporcionam ao consumidor. A partir da revisão da literatura, foram identificadas as principais características dos alimentos consideradas no ato de consumo, as quais foram posteriormente transformadas em variáveis explicativas nos modelos de regressão. A modelagem de preços hedônicos permite a quantificação do impacto de cada uma dessas características no preço final de comercialização nos produtos. A modelagem de preços hedônicos foi aplicada em uma base de dados em painel, com dados agregados com as vendas semanais de hambúrgueres em varejos da região metropolitana de São Paulo entre os anos de 2019 e 2021. Para a regressão dos dados, foram utilizados e comparados o Método dos Mínimos Quadrados Ordinários (OLS) e o Método dos Mínimos Quadrados Ponderados (WLS). No Brasil, o segmento de hambúrgueres congelados é um dos pioneiros na oferta de produtos à base de proteína vegetal como alternativa à proteína animal. Dessa maneira, o estudo realizado permitiu a estimação do impacto no preço de hambúrgueres a partir da utilização dessas proteínas alternativas. Foram estudadas, também, outras características, como a marca e os nutrientes dos produtos, permitindo o entendimento da dinâmica de mercado e formação de preços no segmento de hambúrgueres. Adicionalmente, pôde-se verificar o impacto de atributos relativos à saudabilidade e sustentabilidade no preço dos produtos do segmento, características estas que tendem a ter ainda mais relevância nos próximos anos.

Palavras-chave: Econometria. Preços hedônicos. Mercado de alimentos. Regressão múltipla. Região metropolitana de São Paulo

ABSTRACT

Luz, R. T. (2023). *Application of hedonic price modeling to the frozen packaged hamburger market* (Master's Thesis). Faculdade de Economia, Administração, Contabilidade e Atuária, Universidade de São Paulo, São Paulo.

This work analyzes the frozen hamburger market in the São Paulo Metro Area through the application of hedonic price modeling, originally developed by Rosen (1974). Such modeling is based on Lancaster's Theory of Attributes, a branch of Microeconomics research that proposes studying consumption relations through the characteristics of goods and the individual utility each one provides to the consumer. The main attributes of the food products considered during the act of consumption were identified throughout the study's literature review and subsequently transformed into explanatory variables in the regression models. Hedonic price models allow the quantification of the impact of each of these attributes on the products' final prices. In this study, the model was applied to a panel data set with weekly hamburger sales data aggregated by retailer in the São Paulo Metro Area between 2019 and 2021. Two types of regression methods were applied: the Ordinary Least Squares Method (OLS) and the Weighted Least Squares Method (WLS). In Brazil, the frozen hamburger segment is one of the pioneers in offering plant-based products as an alternative to animal protein. Thus, the use of these alternative proteins is one of the factors whose impact was analyzed. Other characteristics analyzed include the product's brand and its nutritional content, which contributed to the understanding of market dynamics and price formation in the hamburger segment. Furthermore, it allowed the quantification of the impact of attributes related to healthiness and sustainability – characteristics whose relevance is likely to increase in the near future.

Keywords: Econometrics. Hedonic prices. Food market. Multiple regression. Sao Paulo metro area.

LISTA DE FIGURAS

Figura 1 - Fronteira de características para dois bens e duas características	25
Figura 2 - Fronteira de características para três bens e duas características	26
Figura 3 - Curvas θ e p	28
Figura 4 - Curvas \emptyset e p	29
Figura 5 - Pesquisa de preços hedônicos e aplicações no setor de alimentos	31
Figura 6 - λ estimado - Variável dependente preço	42
Figura 7 - Histograma variável dependente preço	43
Figura 8 - Distribuição e correlação das variáveis de Nutrientes	49
Figura 9 - Estatística KMO para Energia_kcal, Carboidratos, Proteínas, Gorduras, Gorduras_saturadas e Fibra_alimentar	50
Figura 10 - Resíduos da regressão OLS	53
Figura 11 - Teste de Breuch-Pagan/Cook-Weisberg - OLS	59
Figura 12 - Outputs do modelo OLS	60
Figura 13 - Estatística VIF - OLS	61
Figura 14 - Teste de Breuch-Pagan/Cook-Weisberg - WLS	62
Figura 15 - Outputs do modelo WLS	62
Figura 16 - Estatística VIF - WLS	63
Figura 17 - Frequências e preços médios – Produtos Proteína Vegetal	67
Figura 18 - Frequências e Box Plot de preço por Ingrediente	86
Figura 19 - Frequências e Box Plot de preço por Marca	87
Figura 20 - Frequências e Box Plot de preço por Região	88
Figura 21 - Frequências e Box Plot de preço por ano	89
Figura 22 - Frequências e Box Plot de preço – Certificado Angus	90
Figura 23 - Frequências e Box Plot de preço – Certificado Orgânico	90
Figura 24 - Frequências e Box Plot de preço – Certificado Vegano	91
Figura 25 - Frequências e Box Plot de preço – Sem Antibióticos / Transgênicos	91
Figura 26 - Frequências e Box Plot de preço – Certificado Wagyu / Selo Sustentabilidade	92
Figura 27 - Dipersão Preço vs Peso da Embalagem	93
Figura 28 - OLS sem seleção de variável	94

LISTA DE QUADROS

Quadro 1 -	Etapas da modelagem de dados proposta	40
Quadro 2 -	Granularidade da base de dados	40
Quadro 3 -	Grupos por Tipo de proteína	45
Quadro 4 -	Variáveis explicativas - Nutrientes	46
Quadro 5 -	Estatística descritiva - Nutrientes	47
Quadro 6 -	Relação entre a estatística KMO e a adequação global da análise fatorial.	48
Quadro 7 -	Fatores PCA para Energia_kcal, Carboidratos, Proteínas, Gorduras, Gorduras_saturadas e Fibra_alimentar	51
Quadro 8 -	Score fatorial para Energia_kcal, Carboidratos, Proteínas, Gorduras, Gorduras_saturadas e Fibra_alimentar	51
Quadro 9 -	Pressupostos do modelo de regressão	55
Quadro 10 -	Variáveis explicativas do modelo	57
Quadro 11 -	Coefficientes e VIF OLS/WLS	64

LISTA DE SIGLAS

FIPE	Fundação Instituto de Pesquisas Econômicas
IPC	Índice de Preços ao Consumidor
KMO	Kaiser–Meyer–Olkin
OLS	<i>Ordinary Least Squares</i> (Mínimos Quadrados Ordinários)
PCA	Análise Fatorial por Componentes Principais
WLS	<i>Weighted Least Squares</i> (Mínimos Quadrados Ponderados)

SUMÁRIO

1	INTRODUÇÃO	17
1.1	DEFINIÇÃO DO PROBLEMA	17
1.2	JUSTIFICATIVA DO TEMA	19
1.3	OBJETIVOS	20
2	REFERENCIAL TEÓRICO	21
2.1	TEORIA DOS ATRIBUTOS	21
2.2	EQUILÍBRIO DE MERCADO E PREÇOS HEDÔNICOS	26
2.3	ESTUDOS EMPÍRICOS	30
3	PESQUISA	39
3.1	MÉTODO DE PESQUISA	39
3.2	COLETA DOS DADOS	40
3.3	UNIVERSO DE PESQUISA	41
3.4	VARIÁVEL DEPENDENTE	42
3.4.1	Transformação BOX-COX	42
3.5	VARIÁVEIS EXPLICATIVAS	43
3.5.1	Características do Produto	43
3.5.2	Nutrientes	46
3.6	ANÁLISE FATORIAL	47
3.6.1	Aplicação da análise fatorial	49
3.7	MODELO DE REGRESSÃO	52
3.7.1	Método dos mínimos quadrados ordinários (OLS)	52
3.7.2	Método dos mínimos quadrados ponderados – <i>Weighted Least Squares</i> (WLS)	53
3.7.3	Pressupostos dos modelos de regressão	55
4	APRESENTAÇÃO DOS RESULTADOS	57
4.1	CARACTERIZAÇÃO DA AMOSTRA	58
4.2	APLICAÇÃO DO MÉTODO DOS MÍNIMOS QUADRADOS ORDINÁRIOS (OLS)	59
4.3	APLICAÇÃO DO MÉTODO DOS MÍNIMOS QUADRADOS PONDERADOS (WLS)	62
4.4	DISCUSSÃO	64
5	CONCLUSÕES E RECOMENDAÇÕES	69
	REFERÊNCIAS	71
	ANEXOS	75

1 INTRODUÇÃO

1.1 DEFINIÇÃO DO PROBLEMA

O mercado de alimentos vem sofrendo constante transformação ao longo das últimas décadas. A atual geração de consumidores busca cada vez mais diversidade e diferenciação dos alimentos adquiridos de maneira geral, enquanto gerações anteriores tinham acesso a uma menor diversidade de opções. Esse movimento relaciona-se diretamente com o bem-estar do consumidor: uma maior diversidade de opções de produtos aumenta a probabilidade de que cada consumidor possa adquirir um alimento cujos atributos reflitam de maneira mais precisa suas preferências individuais (Costanigro & Mccluskey, 2011). Não coincidentemente, fabricantes de produtos alimentícios disponibilizam uma variedade cada vez maior de produtos com diferentes características, que buscam atender às necessidades fisiológicas, sociais e econômicas dos diferentes tipos de consumidor.

Paralelamente, é notório o crescimento da preocupação da sociedade com questões ambientais. Consumidores estão gradualmente mais conscientes acerca das questões ambientais e mais interessados em consumir produtos considerados sustentáveis (Satimanon & Weatherspoon, 2010). Dentre os setores da economia global, o setor alimentício é, sem dúvida, um dos mais questionados e cobrados a respeito. De acordo com Jiang et al. (2019), a preocupação com o impacto ambiental causado pela agricultura tradicional e pela produção em geral de alimentos é crescente entre consumidores e também na própria indústria, gerando uma demanda por uma cadeia sustentável de produção de alimentos.

A sustentabilidade do processo produtivo e os aspectos nutricionais figuram entre as questões que mais influenciam a decisão de consumo de alimentos na atualidade (Hoppe et al., 2012). Nesse contexto, vegetarianismo¹ e veganismo² são identificados como movimentos *bottom-up*, a partir dos quais o consumidor visa influenciar e alterar a estrutura da produção de alimentos (Ruby et al., 2013). Dessa maneira, a quantidade de indivíduos que reduz ou elimina o consumo de carne é crescente, ainda que pequena. No contexto nacional, em 2018 14% dos brasileiros se consideravam vegetarianos e estavam dispostos a consumir mais

¹ Vegetarianismo é uma escolha alimentar na qual se retiram os produtos de origem animal do cardápio.

² Veganismo é um movimento em que seus adeptos evitam, na medida do possível e do praticável, excluir todas as formas de exploração e crueldade contra os animais - seja na alimentação, vestuário ou outras esferas do consumo.

produtos veganos, segundo relatório publicado pela Agência Brasil³. Outra matéria, publicada na revista Exame em 2022⁴, apontava um crescimento de 29% para 50% dos flexitarianos⁵ no Brasil.

Em resposta a esses movimentos, a indústria de alimentos vem, nos últimos anos, diversificando seu portfólio e oferecendo ao mercado mais alternativas de produtos que tragam ao consumidor a percepção de sustentabilidade e cuidado com o processo produtivo e as questões nutricionais. Observa-se uma crescente disponibilidade de produtos à base de proteína vegetal (ou *plant-based*), que substituem os ingredientes animais de alimentos como hambúrgueres, tortas, pratos prontos, quibes, almôndegas, entre outros. Também cresce a oferta de produtos com selos referentes à sua procedência/qualidade e à sustentabilidade do processo produtivo. Um relatório publicado pela *Bloomberg* em 2021⁶ apontou que o mercado de alimentos com base vegetal representará até 7,7% do mercado global de proteínas em 2030, equivalente a \$162 Bi (contra \$29,4 Bi em 2020).

No Brasil, dados da *Euromonitor* apontam que o faturamento de produtos à base de proteína vegetal em 2020 foi de R\$ 452,5 Mi⁷ – o que representa um crescimento de 70% em relação a 2015. Destaca-se o segmento de hambúrgueres, pioneiro dentre os produtos de proteína vegetal, com vendas desde o início de 2019 e crescente participação em faturamento e no portfólio das marcas.

Do ponto de vista acadêmico, é vasta a linha de pesquisa comportamental que avalia as motivações que levam aos consumidores a buscarem alimentos de origem vegetal e com atributos relativos à sustentabilidade da cadeia produtiva ou suas questões nutricionais. Como exemplo, Hoppe et al. (2012) estudam consumo de alimentos orgânicos no Brasil, à luz da Teoria do Comportamento Planejado, enquanto Moraes et al. (2021) realizam um estudo sobre a relevância dos fatores sociais no consumo de frutas e vegetais por adolescentes.

Também é rica a pesquisa microeconômica sobre a dinâmica de consumo de produtos e seus atributos. A Teoria dos Atributos de Lancaster (1966) propõe que produtos adquiridos numa determinada atividade de consumo geram valor (ou utilidade) ao consumidor de

³ Agência Brasil (2021). No Brasil, 14% da população se considera vegetariana.

<https://agenciabrasil.ebc.com.br/saude/noticia/2021-10/no-brasil-14-da-populacao-se-considera-vegetariana>

⁴ Exame (2022). Entenda o boom do mercado de proteína plant-based. <https://exame.com/negocios/boom-do-mercado-de-proteina-plant-based/>

⁵ Pessoas que seguem uma alimentação vegetariana, com eventual consumo de proteína animal.

⁶ Bloomberg (2021). Plant-based Foods Market to Hit \$162 Billion in Next Decade, Projects Bloomberg Intelligence. <https://www.bloomberg.com/company/press/plant-based-foods-market-to-hit-162-billion-in-next-decade-projects-bloomberg-intelligence/>

⁷ Canal Agro (2021). Plant-based: proteína vegetal ganha espaço no Brasil. <https://summitagro.estadao.com.br/noticias-do-campo/plant-based-proteina-vegetal-ganha-espaco-no-brasil/>

maneira indireta, através de suas características (ou atributos) intrínsecas. Complementarmente, Rosen (1974) utiliza tais postulados de Lancaster e propõe que, com base num mercado com oferta e demanda em equilíbrio, tais atributos podem ter seus respectivos preços marginais quantificados – conceito de preço hedônico. A aplicação do modelo de preços hedônicos é bastante abrangente desde a década de 1970. No Brasil, observam-se diversas aplicações do método, como em Fouto (2004) e Fávero (2005), que estudaram os mercados de microcomputadores pessoais e de imóveis, respectivamente. Porém não foram identificadas, até o momento, aplicações ao mercado de alimentos brasileiro, apesar da notória relevância do setor.

Já em outros países, foram identificados diversos estudos relativos aos atributos de alimentos, incluindo o estudo de seus atributos nutricionais e relacionados ao seu processo produtivo. Tais trabalhos avançaram tanto na definição dos atributos hedônicos percebidos, com base na teoria comportamental (como em Baker et al. (2004); Aertsens et al. (2009); Qi & Ploeger (2019)), bem como na implementação do modelo de preços hedônicos em si e efetiva mensuração dos atributos com base nos mercados de consumo em diferentes localidades, como em Satimanon e Weatherspoon (2010) e Qi e Ploeger (2021).

1.2 JUSTIFICATIVA DO TEMA

Martínez-Garmendia (2010) aponta que a aplicação de modelos de preços hedônicos no setor alimentício foi historicamente limitada a alimentos frescos, com poucos exemplos de aplicação a bens de consumo embalados, incluindo alimentos industrializados / processados. De acordo com Martínez-Garmendia (2010), isso se deve às premissas de equilíbrio de mercado da modelagem de preços hedônicos de Rosen (1974), que podem ser mais bem observadas em mercados de transação única, como no caso do mercado imobiliário estudado por Fávero (2005), ou mercados de bens perecíveis com oferta inelástica de bens.

Sendo assim, apesar de sua riqueza enquanto campo de pesquisa, notam-se algumas lacunas relativas à aplicação da Teoria e atributos e do Modelo de Preços Hedônicos: (1) Ausência de aplicações ao mercado de alimentos no Brasil; (2) Poucas aplicações ao mercado de alimentos industrializados.

Dessa maneira, este trabalho aplica o modelo de preços hedônicos de Rosen (1974) ao mercado de hambúrgueres congelados vendidos em grandes varejos da região metropolitana de São Paulo (SP). Com isso, espera-se contribuir para a pesquisa nessas linhas pouco exploradas dentro do tema de Preços Hedônicos, bem como com o entendimento do

comportamento de consumo para proteínas vegetais e alimentos com apelos de saudabilidade e sustentabilidade – segmentos estes já significativos e com tendência de crescimento nos próximos anos.

1.3 OBJETIVOS

Este trabalho pretende trazer um panorama da evolução da pesquisa realizada a respeito de Preços Hedônicos, incluindo suas revisões e evoluções metodológicas, bem como as aplicações do modelo voltadas ao setor de alimentos.

Propõe-se, também, a definição e coleta dos atributos intrínsecos aos produtos do segmento de hambúrgueres congelados potencialmente significativos para a tomada de decisão de compra e valoração desses produtos. Esses atributos serão utilizados nos modelos estimados e terão seus respectivos preços hedônicos estimados.

Relativamente à tendência de crescimento dos segmentos de proteínas vegetais e busca por alternativas de alimentos mais saudáveis e sustentáveis, este estudo busca confirmar se, de fato, o consumidor paulistano valoriza a presença de tais características nos produtos do segmento, bem como estimar o preço hedônico atrelado a cada um desses componentes.

Como mencionado por Martínez-Garmendia (2010), o modelo de preços hedônicos não é usualmente aplicado ao mercado de alimentos industrializados / processados, em função das premissas de equilíbrio de mercado a serem respeitadas pelo modelo. Dessa maneira, pretende-se discutir a viabilidade da aplicação para o segmento de hambúrgueres, bem como realizar, caso necessário, ajustes ao modelo para adequá-lo às premissas e pressupostos da teoria subjacente.

Usualmente, as aplicações do modelo de preços hedônicos utilizam dados ao nível de transação de venda, enquanto esta pesquisa fará uso de uma base de dados agregada com transações de compras semanais. Pretende-se, dessa maneira, discutir as eventuais limitações para aplicação do modelo a dados com essas características, além de propor alternativas de modelagem que sejam adequadas.

2 REFERENCIAL TEÓRICO

Nesta seção, serão revisados os principais conceitos teóricos a serem utilizados no estudo. Isso inclui as teorias e constructos referentes à Precificação Hedônica, os determinantes pessoais no consumo de alimentos sustentáveis e orgânicos, bem como exemplos de aplicação do método em diferentes contextos e mercados no modelo de estimação dos preços hedônicos.

2.1 TEORIA DOS ATRIBUTOS

Diversos estudos foram realizados buscando entender melhor o comportamento de consumo. A Teoria do Comportamento do Consumidor foi fruto de um longo processo de refinamento desde o século XIX, em estudos de Slutsky e Hicks-Allen, posteriormente abordada por Debreu e Uzawa até meados do século XX (Lancaster, 1966). De acordo com essa linha de pesquisa, a atividade de consumo baseia-se na utilidade que um bem, ou conjunto de bens adquiridos, gera ao consumidor. Dessa maneira, o que norteia a escolha de um consumidor em uma determinada atividade de consumo é a utilidade proporcionada.

Também no início do século XX alguns trabalhos avançaram no entendimento de quais características de produtos de diferentes mercados eram significativas ao consumidor e como influenciavam o preço de produtos. Nesse sentido, o estudo de Waugh (1928) é considerado um dos pioneiros, trazendo uma abordagem quantitativa sobre quais características de aspargos, tomates e pepinos eram mais influentes em seus preços, no contexto do mercado norte-americano. Já em seu estudo, Court (1939) traz uma visão da decomposição do preço de carros no mercado norte-americano com base nas características dos veículos e a evolução desses componentes ao longo dos anos. Seguindo linha semelhante, Griliches (1961) realiza uma análise de regressão para entendimento do impacto de cada característica dos veículos em seu preço final de comercialização, denominando os coeficientes estimados em sua formulação de *preços hedônicos*.

Lancaster (1966) propõe a Teoria dos Atributos, partindo de pressupostos similares aos da Teoria do Comportamento do Consumidor, porém assumindo que o produto pode ser representado por seu conjunto de atributos. De acordo com a teoria proposta, o consumo é uma atividade na qual os bens, singularmente ou em combinação, são “entradas” a partir das quais a “saída” é um “pacote” de características (Fávero, 2005).

O referido trabalho de Lancaster é considerado um marco significativo no desenvolvimento da linha de pesquisa voltada aos atributos dos bens consumidos, a utilidade por eles gerada e seus preços implícitos. A metodologia de Lancaster parte de três pressupostos centrais:

- (1) Um produto não gera utilidade ao consumidor de maneira direta, mas sim indiretamente através dos atributos que o constituem;
- (2) Produtos possuem normalmente mais de uma característica, que podem existir também em outros produtos;
- (3) Combinações de produtos podem apresentar diferentes características daquelas observadas nos produtos individualmente.

Nesse contexto, Lancaster estabelece que o consumo de um determinado bem pode ser definido através da expressão linear que relaciona o nível de atividade k com número de bens consumidos na atividade de consumo y . Sendo assim, tem-se a seguinte expressão para o consumo do j -ésimo bem consumido x_j :

$$x_j = \sum_k a_{jk} y_k \quad (2.1)$$

O vetor do total de produtos consumidos dá-se pela seguinte expressão:

$$x = A \cdot y \quad (2.2)$$

Considerando-se que um bem é composto pelo conjunto de atributos que o constituem, assume-se que tal bem pode ser descrito a partir do vetor $z = (z_1, z_2, \dots, z_n)$, onde cada z_i se refere a um atributo, ou característica. Analogamente, tem-se a seguinte expressão para o consumo da i -ésima característica consumida z_i :

$$z_i = \sum_k b_{ik} y_k \quad (2.3)$$

O vetor do total de características consumidas dá-se pela seguinte expressão:

$$z = B y \quad (2.4)$$

Como anteriormente mencionado, o modelo de Lancaster propõe que a utilidade é proporcionada a consumidor através dos atributos dos bens consumidos. Sendo assim, pode-se definir a função $U(z)$, que é a função de utilidade gerada a partir dos atributos consumidos z . Por simplificação, assume-se que há uma correspondência de um para um entre bens e atividades de consumo. Nesse caso, a expressão (2.4) pode ser escrita da seguinte maneira:

$$z = Bx \quad (2.5)$$

Considerando que o consumidor busca a maximização da utilidade gerada, dado um determinado orçamento para a atividade de consumo, tem-se a seguinte expressão de maximização:

$$\begin{aligned} & \max U(z) && (2.6) \\ \text{Sujeito a } & px \leq k \\ \text{Dado que } & z = Bx \text{ e } z, y \geq 0 \end{aligned}$$

Nota-se que as expressões acima dão-se tanto no espaço dos bens (espaço G) quanto no espaço das características (espaço C). A função de maximização $U(z)$ está no espaço C. A restrição orçamentária $px \leq k$ está presente no espaço G. As restrições $z, y \geq 0$ estão presentes em ambos os espaços. E, por fim, a expressão $z = Bx$ ocorre no espaço C, porém, como visto anteriormente, esta expressão também poderia ser escrita como $x = Ay$, localizando-se, nesse caso, no espaço G.

A solução da equação (2.6) requer que a função objetivo e suas restrições sejam escritas dentro do mesmo espaço (G ou C). Para tal, é necessário transformar a função utilidade para o espaço G, ou transformar a restrição de orçamento $px \leq k$ para o espaço C. Em ambos os casos, a equação de transformação $z = Bx$ e a estrutura da matriz B , que relaciona os bens às características, exercem um papel central. Por sua vez, a matriz B depende diretamente da quantidade de bens e características que possui, observando três distintos casos.

No primeiro deles, quando o número de características é igual ao número de bens, tem-se uma relação de um para um entre o vetor de atividades e o vetor de características. Nesse caso, a matriz B é invertível e a equação relacionando bens e características pode ser escrita das seguintes formas:

$$z = Bx, \text{ ou } x = B^{-1}z \quad (2.7)$$

No segundo caso, o número de características excede o número de bens. Assim, o sistema de equações $z = Bx$ possui mais equações que variáveis x_i e, na maioria das vezes, de acordo com Lancaster, a solução e análise desses casos se dá melhor no espaço G.

Por fim, no terceiro caso, onde há mais bens que características, observa-se que a equação de transformação $z = Bx$ possui menos equações que variáveis e para cada vetor de características há mais de um vetor de bens. Isso permite que o consumidor seja capaz de optar pela combinação mais eficiente de bens para atingir um determinado conjunto de

características que satisfaça sua necessidade de utilidade. Nesse contexto, tem-se um problema minimização de custos, sendo z^* o vetor de características desejado:

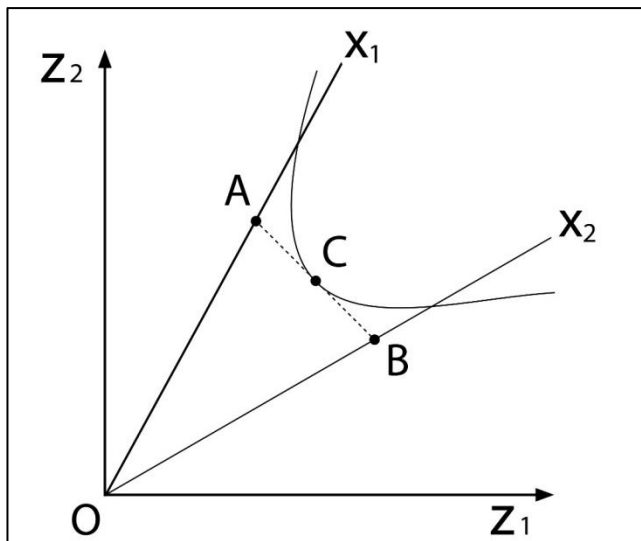
$$\begin{array}{l} \min px \\ \text{Sujeito a } Bx = z^* \\ \text{Dado que } x \geq 0 \end{array} \quad (2.8)$$

Lancaster cita que, dada a complexidade da economia da época – *que se potencializou nas últimas décadas* –, a maior parte dos mercados encontra-se nesse terceiro caso. Isso também vale para o mercado de hambúrgueres congelados proposto neste trabalho, no qual o consumidor tem diversas opções de marcas para adquirir produtos similares com diferenciação de tamanho, sabor e outras características.

Para esse tipo de formulação, Lancaster (1966) menciona que a escolha do consumidor consiste em duas partes: escolha eficiente e escolha privada. A escolha eficiente refere-se à determinação da fronteira de características e a combinação eficiente de bens a ela associada. A escolha privada consiste na definição de qual ponto da fronteira de características é preferida por um determinado consumidor.

Na Figura 1 podemos observar esses aspectos através de um exemplo simplificado com dois bens e duas características. Os eixos representam as características z_1 e z_2 – portanto estamos no espaço C. Os bens x_1 e x_2 possuem características z_1 e z_2 , porém em proporções diferentes – o que pode ser confirmado pela distinta inclinação entre as retas relativas a cada um dos bens. Nesse cenário, o segmento de reta AB representa a fronteira de características para aquisição de bens x_1 e x_2 (escolha eficiente). Já no ponto C, onde a fronteira é tangenciada pela curva de indiferença de um determinado consumidor, temos a representação da escolha privada.

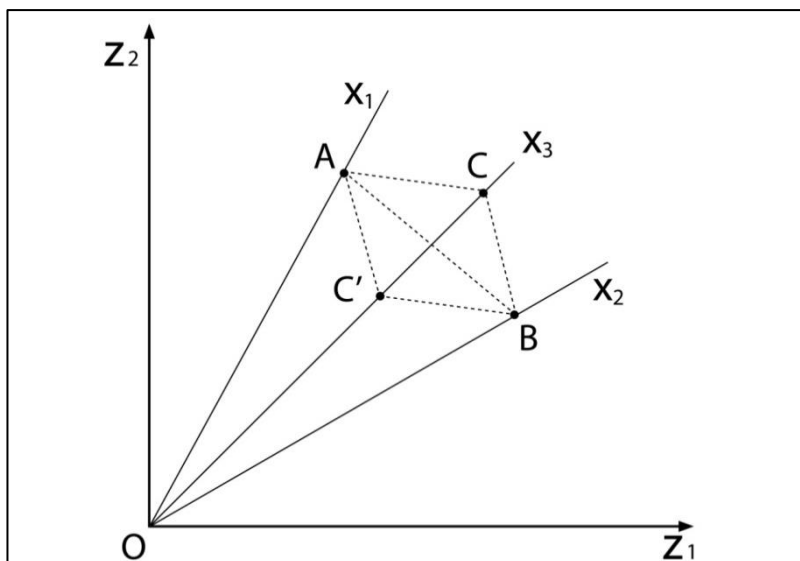
Figura 1 - Fronteira de características para dois bens e duas características



Fonte: elaborado pelo Autor.

Observa-se que, para um orçamento definido, a quantidade a ser adquirida de cada bem é limitada ao preço do bem em si, conforme equações (2.6) e (2.8). Dessa maneira, utilizando um exemplo de três bens e duas características, nota-se que o preço de cada bem pode interferir diretamente na escolha de um determinado consumidor que busca as características z_1 e z_2 . Conforme demonstrado na Figura 2, é possível comprar uma determinada quantidade do bem x_3 a um determinado preço, ilustrado pelo ponto C . Nesse cenário, o bem x_3 pertence à fronteira de características, que é composta pelos segmentos AC e CB . Porém, caso o bem x_3 tenha seu preço aumentado suficientemente, nota-se que o ponto C' (que representa a quantidade de x_3 que pode ser adquirida com o novo preço) deixa de compor a fronteira de características, que passa a ser representada pelo segmento AB .

Figura 2 - Fronteira de características para três bens e duas características



Fonte: elaborado pelo Autor.

Nesse caso, entende-se que consumidores tenderiam a adquirir combinações dos bens x_1 e x_2 caso desejassem obter as características z_1 e z_2 . De acordo com Lancaster (1966), há um efeito de substituição eficiente. Caso as alterações nos preços não sejam grandes o suficiente, não há alteração nos segmentos que compõem a fronteira de características e, conseqüentemente, o efeito de substituição eficiente não ocorre. Caso contrário, observa-se esse que efeito, que pode acarretar completa alteração entre uma atividade de compra e outra.

2.2 EQUILÍBRIO DE MERCADO E PREÇOS HEDÔNICOS

Em seu livro sobre índices de preço e mudança de qualidade, Griliches (1971) aponta a importante contribuição de Lancaster (1966) entre outros trabalhos na formulação de um modelo onde bens são representados por conjuntos de atributos. Porém, reconhece que tais estudos não trouxeram novas implicações significativas para os constructos de índices de preço e que, até o referido momento, muitas das contribuições para o tema haviam sido majoritariamente empíricas.

Em seu estudo sobre Precificação Hedônica e Mercados Implícitos, Rosen (1974) busca cobrir justamente essa lacuna, desenvolvendo um modelo teórico com potencial aplicação a mercados reais. Rosen estabelece que o valor dos produtos é definido a partir de seus atributos (ou características) que geram utilidade ao consumidor. Além disso, de acordo com Rosen, os atributos mencionados poderiam ter seu preço hedônico definido, que se

referem ao preço implícito de cada atributo a partir do qual oferta e demanda de um determinado mercado entram em equilíbrio.

Rosen traz o conceito de que, para um determinado mercado, o vetor z de características está presente (de maneira heterogênea) em seus produtos. Os componentes do vetor z podem ser objetivamente observados nos produtos, de modo que diferentes consumidores podem observar igualmente essas características nos produtos. Em contrapartida, diferentes consumidores podem valorizá-las de maneira diferente, com base na utilidade que lhes proporcionarão.

Analogamente ao trabalho de Lancaster (1966), Rosen define que um produto pode ser representado por suas coordenadas $z = z_1, z_2, \dots, z_n$ no espaço de características, onde z_i se refere à quantidade da i -ésima característica do produto. Além disso, cada ponto no espaço também está atrelado a um preço $p(z) = p(z_1, z_2, \dots, z_n)$, que direciona consumidores e produtores (oferta e demanda) em suas escolhas de conjuntos de características adquiridos e vendidos.

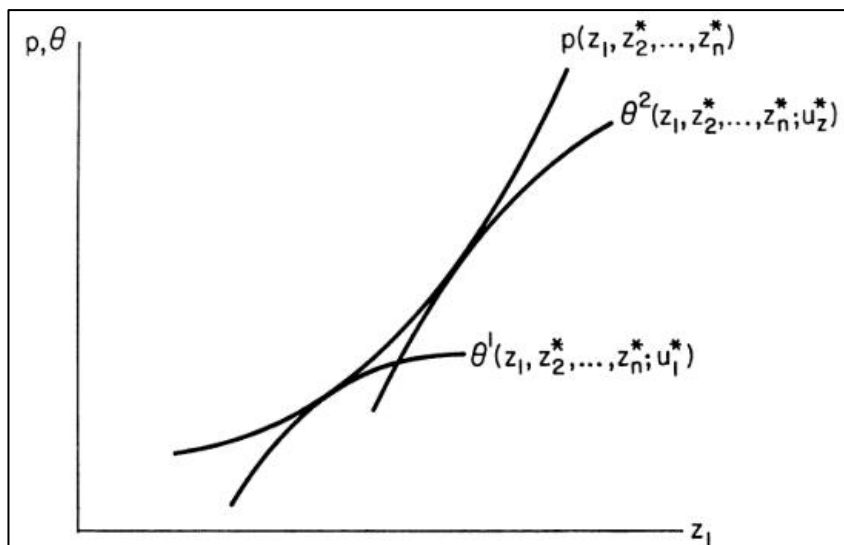
Dessa maneira, o consumidor tem, à sua disposição, um espectro de produtos e escolhe a combinação de produtos que será adquirida - considerando suas restrições orçamentárias para aquela atividade de compra e considerando sua função de utilidade $U(x, z_1, z_2, \dots, z_n)$, levando em conta os atributos do vetor z e outros produtos x adquiridos. Essa função é, por hipótese, côncava.

Assumindo-se que o preço de x é unitário, chega-se à seguinte expressão representando a renda do consumidor:

$$y = x + p(z) \tag{2.9}$$

Sendo assim, analogamente à função U , é definida a função $\theta(z; u, y)$, que representa o valor que o consumidor está disposto a pagar para obter diferentes quantidades do vetor z , considerando sua renda. A função θ define um conjunto de curvas de indiferença relacionando z_i com dinheiro e é côncava em z , dado que a função U também o é. Ou seja, o valor que um consumidor está disposto a pagar por valores incrementais do atributo z_i aumenta à medida que z_i cresce, porém a uma taxa de crescimento cada vez menor.

Sendo $p(z)$ o preço que o consumidor efetivamente terá que pagar no mercado para obter determinada quantidade das características z , observa-se que o ponto ótimo sob a ótica do consumo ocorre justamente quando a curva sua θ tangencia a curva p , conforme ilustrado na Figura 3:

Figura 3 - Curvas θ e p 

Fonte: Rosen (1974).

Nota-se que a Figura 3 é um recorte do espaço de características, fixando valores de (z_2^*, \dots, z_n^*) . Além disso, diferentes consumidores possuem diferentes funções θ , representadas pelas curvas θ^1 e θ^2 que tangenciam $p(z)$ em pontos diferentes, o que significa que o consumidor 2 tem preferências por produtos que ofereçam mais da característica z_1 . Nesse contexto, Rosen aponta que consumidores com funções de valor θ tendem a adquirir produtos com especificações semelhantes, resultado do modelo de equilíbrio espacial.

Outra conclusão importante é que esse modelo de decisão de consumo proposto é também aplicável para compras de mais unidades do bem que oferece os atributos z_i . Nesse caso, a função utilidade pode ser representada por $U(x_1, z_1, z_2, \dots, z_n, m)$, onde m é a quantidade adquirida do produto.

Do ponto de vista da produção, os custos relacionados à produção de bens com as características z são definidos a partir da função $C(M, z; \beta)$, onde $M(z)$ representa a quantidade produzida de bens com o vetor de características z e β representa outros parâmetros relativos à função de produção. Além disso, assume-se que C é convexa e o custo marginal para produção de maiores níveis de características z aumenta ao longo de z . Dessa forma, a função de lucro π se dá da seguinte maneira:

$$\pi = Mp(z) - C(M, z_1, \dots, z_n) \quad (2.10)$$

Pressupondo-se um mercado não monopolista, considera-se que M e $p(z)$ são independentes. Dessa maneira, a otimização da função lucro π requer as seguintes condições:

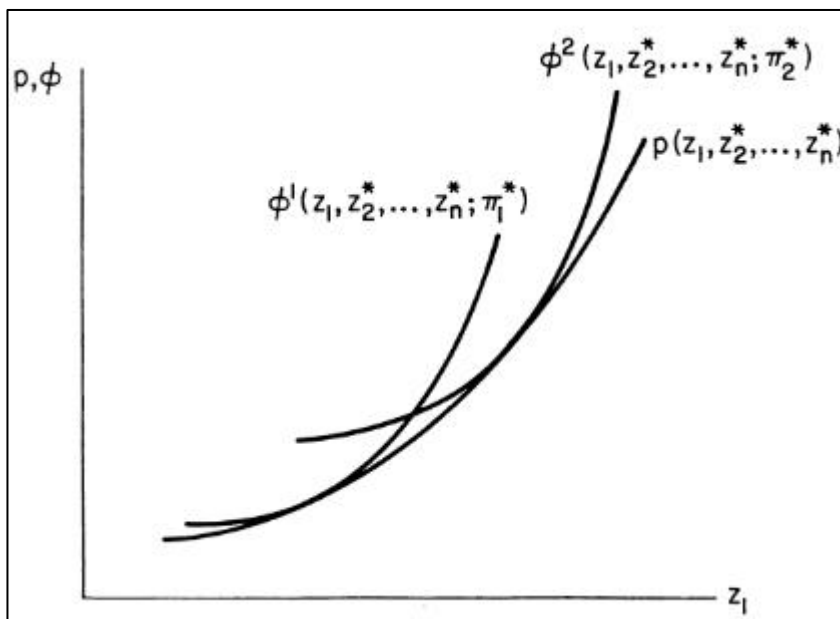
$$p_i(z) = C_{z_i}(M, z_1, \dots, z_n)/M \quad i = 1, \dots, n \quad (2.11)$$

$$p(z) = C_M(M, z_1, \dots, z_n) \quad (2.12)$$

No ponto ótimo, a receita marginal de atributos adicionais z se iguala aos seus custos unitários de produção. A partir desse ponto, os custos superam a receita marginal e, do ponto de vista de lucratividade, não é mais vantajoso para o produtor incrementar os elementos oferecidos no vetor z .

Sendo assim, analogamente à curva θ , que representa o valor que o consumidor está disposto a pagar por características z , estabelece-se a curva ϕ , que representa o valor de oferta, conforme Figura 4. Nota-se que diferentes vendedores também possuem diferentes curvas ϕ , dado que possuem diferentes parâmetros π e, dessa forma, tangenciam $p(z)$ em diferentes pontos no espaço. Isso significa que diferentes vendedores terão seus próprios pontos de maximização de lucro oferecendo diferentes quantidades do vetor z em seus produtos.

Figura 4 - Curvas ϕ e p



Fonte: Rosen (1974).

Definidas as curvas de valores de oferta θ e demanda ϕ , tem-se, então, o equilíbrio de mercado justamente no ponto em que as curvas se tangenciam. Nesse cenário, observações de $p(z)$ representam simultaneamente ambas as famílias de funções.

Uma formulação mais abrangente para o valor de oferta e demanda, considerando variáveis exógenas que influenciam tanto a curva de oferta quanto a de demanda, é a seguinte:

$$p_i(z) = F^i(z_1, \dots, z_n, Y_1, W) \quad (\text{demanda}) \quad (2.13)$$

$$p_i(z) = G^i(z_1, \dots, z_n, Y_2, W) \quad (\text{oferta}) \quad (2.14)$$

Onde Y_1 são variáveis exógenas que influenciam a curva de valor de demanda, Y_2 são variáveis exógenas que influenciam a curva de valor de oferta e W são variáveis exógenas que influenciam ambas as curvas.⁸

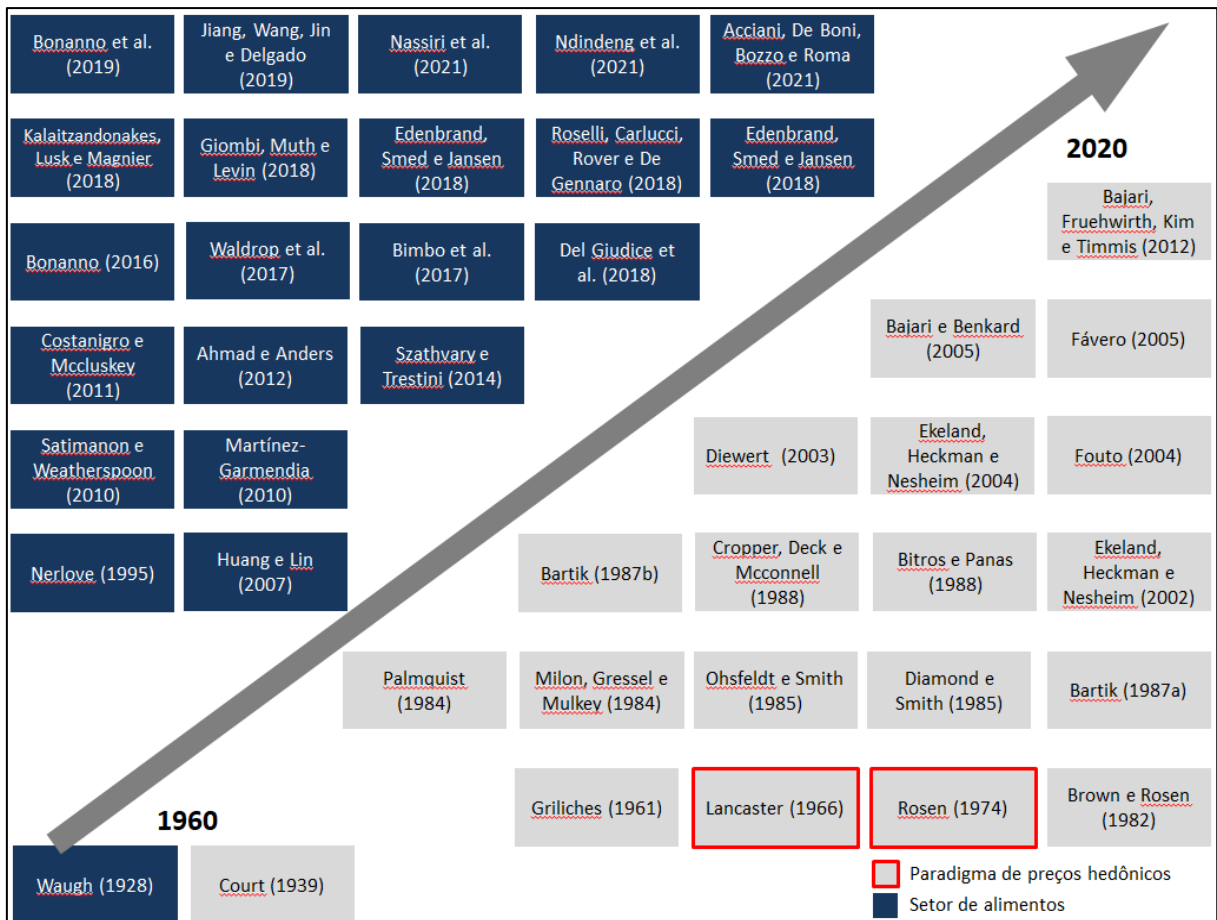
Por fim, Rosen propõe que a formulação seja aplicada em dois estágios. No primeiro estágio, deve-se estimar $\hat{p}(z)$ a partir do método hedônico usual, desconsiderando variáveis exógenas Y_1 , Y_2 e W . A partir disso, é possível calcular os preços marginais implícitos $\partial p / \partial z_i = \hat{p}_i(z)$, considerando a quantidade de cada característica efetivamente vendida. No segundo estágio, os preços marginais $\hat{p}_i(z)$ são utilizados como variável resposta nas equações de demanda (2.13) e oferta (2.14).

2.3 ESTUDOS EMPÍRICOS

Na Figura 5, observamos as principais referências nas quais o proposto trabalho basear-se-á. Trata-se de estudos teóricos posteriores aos de Rosen e Lancaster, bem como de aplicações do modelo de preços hedônicos ao setor de alimentos.

⁸ Rosen (1974) propõe as equações de oferta e demanda desconsiderando as variáveis exógenas W comuns a ambas as curvas. A formulação incluindo W foi proposta no trabalho de Fávero, 2005.

Figura 5 - Pesquisa de preços hedônicos e aplicações no setor de alimentos



Fonte: elaborado pelo Autor.

A teoria econômica sobre modelos hedônicos, preços, oferta e demanda para atributos de bens proposta por Rosen consolidou-se nas últimas décadas, porém seu modelo empírico ainda é campo para debate (Ekland, Heckman, & Nesheim, 2002). Muitos trabalhos se dedicaram à efetiva aplicação empírica do modelo em mercados reais e análise dos resultados, enquanto outros avançaram no estudo de aspectos ainda considerados frágeis da modelagem de Rosen, como as questões relativas à identificação dos modelos hedônicos, buscando melhor entendimento das técnicas quantitativas e formas funcionais mais adequadas a cada tipo de caso, ou então avançando na definição do conjunto de atributos a ser considerado na modelagem para cada tipo de mercado.

Milon, Gressel e Mulkey (1984) citam que limitar o modelo hedônico às formas funcionais convencionais pode, potencialmente, gerar estimativas enviesadas dos parâmetros das equações de Rosen. Dessa maneira, os autores propõem a adoção da transformação Box-Cox, que permitiria uma identificação confiável e mais flexível da forma funcional mais

aderente aos dados, a partir da incorporação dos parâmetros λ_1 e λ_2 ao modelo, conforme equação (2.15), onde λ_1 e λ_2 são calculados para que as variáveis dependentes e explicativas sejam, respectivamente, normalizadas.

$$P(Z)^{\lambda_1} = \beta Z^{\lambda_2} + \epsilon \quad (2.15)$$

De acordo com Milon et al. (1984), o modelo geral proposto na equação poderia ser adaptado tanto para o primeiro quanto para o segundo estágio de Rosen. Posteriormente, tal modelagem foi amplamente aplicada e discutida (Bitros & Panas, 1988; Fávero, 2005; Nerlove, 1995).

Em contraponto, Cropper, Deck e McConnell (1988) apontam alguns riscos da seleção da forma funcional levando-se em conta exclusivamente a aderência do modelo aos dados. Os autores mencionam que a teoria econômica impõe poucas restrições à forma funcional dos modelos de preços hedônicos e que, por isso, muitos estudos acabam optando pela função que traz maior aderência aos dados, ao invés de utilizar a forma que melhor estima o preço marginal dos atributos. Além da própria identificação da forma funcional adequada, o estudo também simula o impacto gerado a partir da omissão de atributos, ou mesmo da utilização de *proxies* inadequadas.

Bartik (1987a) aponta a questão da possível endogeneidade das variáveis Z da equação de demanda do segundo estágio de Rosen (2.13) devido a não linearidade da função de preços hedônicos. Como alternativa, Bartik sugere a utilização de variáveis instrumentais exógenas que influenciam a restrição de orçamento do consumidor. Outro tema abordado por Bartik (1987b) é a importância de um modelo hedônico multimercado, em contraposição a alguns estudos do mesmo período (Diamond & Smith, 1985; Ohsfeldt & Smith, 1985; Palmquist, 1984) que apontam a possibilidade de se aplicar o modelo em dados de mercado único.

Brown e Rosen (1982) apontam potenciais fraquezas na estimação do segundo estágio proposto no modelo de Rosen e dão luz ao problema de identificação, um dos aspectos mais significativos a ser considerando na aplicação da modelagem de preços hedônicos. Na mesma linha, Ekeland, Heckman e Nesheim (2002) apontam que o modelo hedônico tradicional apresenta três principais fraquezas: (1) a identificação do modelo pode ser obtida apenas a partir da escolha arbitrária de uma forma funcional; (2) o vetor z é necessariamente correlacionado aos termos de erro das equações de oferta e demanda do segundo estágio de Rosen, o que traz endogeneidade ao modelo; e (3) o modelo requer dados multimercado para estimação das variáveis exógenas aos produtos nas funções de oferta e demanda.

Diewert (2003) traz uma relação de recomendações para a aplicação do modelo de Preços Hedônicos, relativas às principais questões técnicas ainda discutidas acerca do modelo. São abordadas questões como a transformação logarítmica das variáveis dependente e explicativas, bem como técnicas para utilização dos modelos em bases de dados em painel. Especialmente interessante, porém, é a recomendação de ponderação do modelo de regressão em caso de utilização de dados consolidados de vendas. Para isso, o autor cita o modelo de Mínimos Quadrados Ponderados (WLS – *Weighted Least Squares*), onde é possível ponderar cada observação da base de dados a partir de um determinado peso. Diewert sugere que a ponderação do modelo pela quantidade de vendas realizadas (ou mesmo do valor total vendido) de cada produto pode garantir a representatividade do modelo sobre o universo estudado, no caso da utilização de uma base de dados que contenha a informação consolidada de vendas, ao invés de uma base com o nível de transação de venda.

Em consonância com as críticas de Brown e Rosen (1982) e Ekeland, Heckman e Nesheim (2002, 2004) citam que o modelo tradicional de dois estágios proposto por Rosen não explora por completo o conteúdo econômico do modelo hedônico. Endereçando as fraquezas mencionadas, Ekeland, Heckman e Nesheim (2004) propõem dois métodos alternativos: (1) um modelo de transformação não-paramétrica; e (2) um modelo que utiliza variáveis instrumentais para problemas paramétricos não lineares. Os métodos citados seriam, inclusive, robustos a dados para um único mercado.

Bajari e Benkard (2005) propõem uma modelagem que generaliza o modelo de Rosen em alguns aspectos: (1) permite que o modelo seja aplicado em mercados com competição imperfeita; (2) permite a existência de características não observadas dos produtos; (3) generaliza a estimativa do primeiro estágio para não-paramétrica; e (4) aborda o segundo estágio de Rosen de maneira alternativa, permitindo espaços discretos de produtos de características. Ainda buscando alternativas para contextos com variáveis não observadas, Bajari et al. (2012) demonstram abordagem que controla atributos não observados de produtos, robusta a atributos não observados fixos ou variáveis no tempo.

A respeito do mercado de alimentos, Costanigro e Mccluskey (2011) propõem um *framework* resumindo as aplicações do modelo de preço hedônico ao mercado, apontando principais técnicas utilizadas, limitações e desafios observados nas diversas aplicações do modelo.

Segundo Costanigro e Mccluskey (2011), a transformação logarítmica da variável dependente de preço (2.16) é a forma funcional mais utilizada para modelos paramétricos, dada a dispersão normalmente observada para dados da curva de preços. Alternativas de

forma funcional propostas pelos autores são a Box-Cox de parâmetro único (2.17) ou a ainda mais flexível Box-Cox quadrática (2.18).

$$\ln(P_i) = \alpha_0 + \sum_{k=1}^K \alpha_k z_{ki} + \epsilon_i \quad (2.16)$$

$$p_i^\theta = \alpha_0 + \sum_{k=1}^K \alpha_k z_{ki} + \epsilon_i \quad (2.17)$$

$$p_i^\theta = \alpha_0 + \sum_{k=1}^K \alpha_k z_{ki}^\lambda + \frac{1}{2} \sum_{k=1}^K \sum_{j=1}^K \gamma_{kj} z_{ki}^\lambda z_{ji}^\lambda + \epsilon_i \quad (2.18)$$

Além da abordagem paramétrica, Costanigro e Mccluskey (2011) propõem também uma modelagem não paramétrica, conforme equação (2.19). Nesse modelo, não há suposição a respeito da função de densidade de probabilidade do termo de erro ϵ e $g(z)$ não é especificada, o que permite a construção de uma forma funcional completamente aderente aos dados modelados. Em contrapartida, a flexibilidade tende a gerar modelos com ruído e imprecisão, caso os dados disponíveis não apresentem uma grande densidade de observações.

$$p = g(z) + \epsilon \quad (2.19)$$

Quanto à aplicação empírica, diversos trabalhos utilizaram a abordagem de preços hedônicos na análise de diferentes setores e mercados. De acordo com Fávero (2005), a aplicação pode seguir duas abordagens distintas: aplicação de uma única equação hedônica (seja de oferta ou demanda), para avaliação dos preços de bens pelas suas características associadas ou; definição de equações de oferta e demanda, com o intuito de ser possível a determinação do equilíbrio de mercado.

Especificamente ao mercado de alimentos, há diversos exemplos de aplicação nos últimos anos. Huang e Lin (2007) aplicam o primeiro estágio de Rosen a dados em painel agregados semanalmente, para tomates orgânicos no mercado norte-americano, incluindo variáveis socioeconômicas e demográficas.

Satimanon e Weatherspoon (2010) aplicam o primeiro estágio de Rosen ao mercado norte-americano de ovos para estimar o preço hedônico de atributos relativos à sustentabilidade.

Ahmad e Anders (2012) implementam o primeiro estágio de Rosen a dados em painel agregados semanalmente, para alimentos processados no mercado canadense, utilizando transformação log-linear, visando a identificação dos preços hedônicos das diferentes marcas comercializadas.

Szathvary e Trestini (2014) aplicam uma transformação log-linear ao mercado de sucos de fruta italiano, com o intuito de analisar o preço hedônico associado a atributos relativos à nutrição e saudabilidade.

Bonanno (2016) aplica o primeiro estágio de Rosen a dados em painel para iogurtes no mercado norte-americano consolidado (mercado único), utilizando transformação log-linear, identificando preços hedônicos de marcas, atributos saudáveis e atributos não saudáveis.

Waldrop, McCluskey e Mittelhammer (2017) realizam um estudo no mercado norte-americano de vinhos e utilizam dados em painel coletados de catálogos de preço de vinhos, aplicando transformação Box-Cox na variável dependente. O estudo revela que certificados orgânicos e de sustentabilidade obtidos pelos vinhos são valorizados pelo consumidor, mas que essa valorização marginal decresce a cada certificado adicionado ao produto. Um ponto de atenção sobre esse estudo reside no fato de que os preços foram coletados em catálogos, o que não necessariamente garante a efetiva comercialização do produto àquele preço e poderia ferir as premissas do modelo de Rosen.

Roselli et al. (2018) trazem uma aplicação do primeiro estágio de Rosen utilizando transformação log-linear, em dados do mercado online brasileiro de azeite de oliva.

Del Giudice et al. (2018) aplicam o primeiro estágio do modelo de Rosen a dados de vendas de atum na Itália e concluem que produtos com certificados de qualidade relacionados à sustentabilidade possuem correlação positiva com o preço do produto. Esse achado corrobora o constructo proposto por este trabalho.

Edenbrand, Smed e Jansen (2018) aplicam o primeiro estágio de Rosen em base de dados em painel de consumo de produtos processados variados nos mercados da Holanda e Dinamarca, utilizando transformação log-linear, com objetivo de identificar preço hedônico de certificados nutricionais. O estudo aponta, de maneira geral, que os consumidores de diferentes países da Europa valorizam produtos com esse tipo de certificado, em função de sua correlação com a saudabilidade.

Kalaitzandonakes, Lusk e Magnier (2018) aplicam o primeiro estágio Rosen ao mercado de vegetais norte-americano e concluem que alimentos não transgênicos são valorizados pelo consumidor. O estudo faz uso de dados em painel e utiliza variáveis de controle para absorver os movimentos sazonais e tendência dos preços ao longo dos períodos encontrados na base de dados.

Giombi, Muth e Levin (2018) aplicam o primeiro estágio de Rosen a dados de vendas de sopas e utilizam como atributos do vetor Z do modelo as informações nutricionais dos alimentos. O estudo traz achados interessantes que endossam o constructo proposto por este

trabalho, apontando que consumidores dos produtos estudados valorizam aqueles com certificados orgânicos. Além disso, uma parcela significativa dos nutrientes relacionados à saudabilidade utilizados no estudo também apresentou coeficientes positivos.

Bonanno et al. (2019) trazem uma aplicação do primeiro estágio de Rosen ao setor de iogurtes na Itália e avaliam que os certificados e atributos de credibilidade proporcionam um aumento dos preços devido à diferenciação e melhora da reputação dos produtos. No trabalho, são utilizados dados mensais de vendas agregadas e são testadas diferentes formas funcionais, com o intuito de endereçar questões de assimetria de informação entre para o mercado em questão. Para isso, os autores exploram diferentes formulações do termo de erro da regressão realizada. Conforme Equação (2.20), Z é o vetor de variáveis que potencialmente interferem na habilidade do produtor de cobrar preços mais próximos à fronteira de preços do consumidor.

$$\sigma_u^2 = \exp(\lambda' z) \quad (2.20)$$

Jiang et al. (2019) aplicam o primeiro estágio de Rosen para alimentos com etiqueta “green food” para vendas online na China utilizando transformação log-linear.

Ndindeng et al. (2021) trazem aplicação do primeiro estágio de Rosen para atributos de arroz aplicado a três *clusters* de consumidores no mercado de Benin.

Nassiri et al. (2021) realizam aplicação dos dois estágios de Rosen para o mercado de peixes, utilizando transformação log-quadrática em dados multimercado na Polinésia Francesa.

Acciani et al. (2021) aplicam o primeiro estágio de Rosen ao mercado de lentilhas no sul da Itália, utilizando transformação log-linear em seu modelo.

Nerlove (1995) aplica o primeiro estágio de Rosen a dados em painel de vendas de vinhos no mercado sueco. Diferentemente do usual, o autor utiliza a quantidade vendida como variável dependente do modelo e o preço em conjunto com os atributos como variáveis preditoras, assumindo que, para tal mercado específico, o preço e os atributos são exógenos.

Outro estudo que foge do padrão é o realizado por Martínez-Garmendia (2010), que aplica o primeiro estágio de Rosen a dados em painel de vendas de refrigerantes no mercado norte-americano. De acordo com Martínez-Garmendia, alimentos e bebidas processados e embalados trazem um desafio adicional à aplicação de modelo de preços hedônicos: dado que se trata de produtos normalmente com validade estendida, no curto-prazo o preço é definido

exclusivamente pelo vendedor, sem interação do comprador. Sendo assim, a aplicação tradicional de um modelo de preços hedônicos a um mercado com tal característica tende a considerar apenas a estratégia de atribuição de preços do vendedor. Dessa maneira, Martínez-Garmendia (2010) utiliza uma forma funcional duplo-log, porém, diferentemente dos demais autores, utiliza o volume de vendas como variável preditora do preço, junto com demais atributos Z observados nos produtos. A inclusão do volume vendido como variável preditora tem justamente o objetivo de incluir o comportamento do consumidor na equação proposta. O autor reconhece que a inclusão dessa variável potencializa problemas de endogeneidade do modelo e endereça a questão aplicando um modelo de mínimos quadrados em dois estágios (2SLS).

Bimbo et al. (2016) apresentam uma alternativa para considerar os volumes de vendas na ponderação dos resultados do modelo. Os autores aplicam o primeiro estágio de Rosen a dados em painel para a mercado de leite na Itália. O estudo indica que o tratamento UHT no leite é valorizado por parte do consumidor. Esse achado corrobora o constructo deste projeto, dada a relação entre o tratamento UHT e a saudabilidade do leite. Porém, o aspecto mais interessante do estudo é a utilização de um modelo de Mínimos Quadrados Ponderados (WLS) para a regressão. Os autores decidem utilizar dados consolidados de vendas e buscam, com o modelo WLS, dar mais peso na regressão para produtos com vendas mais frequentes, conforme recomendado por Diewert (2003).

A partir das referências trazidas, nota-se que a maior parte dos estudos de modelo hedônico aplicado ao mercado de alimentos limita-se ao primeiro estágio de Rosen – informação corroborada por Costanigro e Mccluskey (2011) e confirmada pelos estudos realizados posteriormente a 2011 e aqui mencionados. Também se destaca o fato de serem observadas diferentes maneiras para lidar com aplicação de preços hedônicos a dados de vendas agrupadas. Como mencionado por Diewert (2003), caso não seja realizado nenhum tipo de ponderação das quantidades ou valores comercializados de um produto, a representatividade do modelo relativa ao universo estudado pode ficar comprometida. Sendo assim, salta aos olhos o fato de que apenas os trabalhos do Nerlove (1995), Martínez-Garmendia (2010) e Bimbo et al. (2016) apresentaram alternativas de modelagem que levem em conta as quantidades comercializadas.

3 PESQUISA

3.1 MÉTODO DE PESQUISA

Este estudo aplica o primeiro estágio de Rosen – descrito na seção anterior – em dados reais de vendas de hambúrgueres congelados comercializados em grandes varejos na região metropolitana de São Paulo. Dessa maneira, a variável dependente do modelo é o preço real de comercialização dos produtos e as variáveis explicativas Z são as características intrínsecas dos produtos.

Como mencionado, o segmento de hambúrgueres congelados foi pioneiro no movimento de diversificação e oferta de alternativas de produtos diferenciados pelo tipo de proteína utilizado e por diversas características em seus respectivos processos produtos. Portanto, foram utilizados atributos relativos a: (1) tipo de proteína utilizada em cada produto, (2) certificados e selos de qualidade que remetam a um processo produtivo mais sustentável e (3) tabela nutricional de cada produto analisado. Essas são características que vão ao encontro das crescentes demandas dos grupos de consumidores que buscam alternativas de alimentação que tragam a percepção de maior cuidado com o ambiente e com questões nutricionais.

O método para aplicação do modelo consiste em uma série de etapas de ajuste dos dados ao constructo teórico proposto e verificações estatísticas que garantam a adequação do modelo⁹. O quadro abaixo traz um resumo dessas etapas, que também serão posteriormente detalhadas neste capítulo.

⁹ Todas as etapas analíticas aqui descritas foram realizadas utilizando o software livre RStudio. O ANEXO A traz o código de programação que executa todos esses passos.

Quadro 1 - Etapas da modelagem de dados proposta

ETAPA	PROPÓSITO	MÉTODO
Definição das variáveis explicativas e dependente	Construir, a partir dos dados coletados, variáveis que representem de maneira adequada a decisão de compra estudada.	- Deflação dos preços no período estudado; - Criação de variáveis explicativas determinantes para a decisão de compra.
Análise fatorial	Identificar multicolinearidade das variáveis explicativas, que possa interferir no cálculo dos coeficientes hedônicos dos atributos.	- Análise de correlação simples; - Estatística Kaiser–Meyer–Olkin (KMO); - Teste de esfericidade de Bartlett.
Determinação do modelo de regressão	Identificar forma funcional mais adequada para o modelo, garantindo consistência dos dados ao constructo teórico.	- Transformação Box-Cox da variável dependente; - Mínimos quadrados ordinários; - Mínimos quadrados ponderados.

Fonte: elaborado pelo Autor.

3.2 COLETA DOS DADOS

O trabalho utilizou uma base de dados privada, obtida através de uma instituição que realiza coletas de dados de vendas em varejo na unidade federativa de São Paulo. Para garantir a confidencialidade dos dados, nenhuma marca será nominalmente mencionada no estudo. Serão, sim, utilizadas variáveis de controle de marcas, dado que tal atributo tem um efeito significativo sobre o preço dos produtos, conforme sugerido pela literatura revisada. Porém, todas variáveis relativas às marcas serão mascaradas.

A base possui informações consolidadas de vendas dos estabelecimentos comerciais aos consumidores finais, com a seguinte granularidade:

Quadro 2 - Granularidade da base de dados

DIMENSÃO	GRÃO	OBSERVAÇÃO
Tempo	Semana	Refere-se ao período de domingo até sábado
Localidade	Zonas do município de São Paulo-SP	Regiões participantes: Centro, Zona Sul, Zona Leste, Zona Oeste, Zona Norte e ABDC
Produto	Produto	

Fonte: elaborado pelo Autor.

Cada ponto de dados da base possui a informação do preço médio ponderado (R\$/unidade) para todas as vendas do referido produto, na referida semana, na respectiva zona do município de São Paulo-SP.

Adicionalmente à base de dados mencionada, também foram coletadas manualmente as informações dos atributos dos produtos estudados, a partir da inspeção visual de suas embalagens no ponto de venda. Trata-se de informações nutricionais, certificados e informações dos ingredientes utilizados no preparo do produto, que serão detalhados mais adiante.

3.3 UNIVERSO DE PESQUISA

A base de dados refere-se ao período de 2019-2021 e os dados representam a somatória de vendas em todos os estabelecimentos onde os dados são coletados. Os dados abrangem os Varejos da região metropolitana de São Paulo - SP; sendo assim, ficam de fora da análise canais de vendas como Atacados, Padarias, Mercarias, Revendedores B2B, entre outros.

Dentro deste universo de pontos de vendas, foram analisados 49 produtos distintos de hambúrgueres embalados e congelados, comercializados dentro do período de estudo. Esses produtos pertencem a 16 das maiores marcas do segmento, representando 63,9% das vendas totais do segmento de hambúrgueres embalados e congelados dentro do período de estudo, nas mencionadas localidades. Tal diversidade é importante para garantir um bom nível de heterogeneidade entre as características dos produtos estudados – premissa essa importante para a aplicação do modelo de Preços Hedônicos de Rosen.

Também foram excluídas da base as vendas realizadas por meio de algum tipo de oferta de redução de preço, dado que a base de dados original traz a informação das vendas com e sem desconto. Optou-se por esse filtro, para minimizar a presença de fatores exógenos que possam distorcer o comportamento real dos preços dos produtos analisados e, por conta disso, distorcer os componentes hedônicos avaliados por este estudo.

3.4 VARIÁVEL DEPENDENTE

A variável dependente utilizada no estudo foi o preço unitário por embalagem de hambúrguer comercializada (R\$/unidade). Salienta-se que cada embalagem pode conter diferentes quantidades de hambúrgueres, com diferentes pesos unitários. De qualquer forma, entende-se que o preço por embalagem melhor reflete o fenômeno da decisão de compra por parte do consumidor, pois é esse o preço efetivo informado na gôndola onde o produto é exposto.

Os preços obtidos foram deflacionados a partir do IPC de alimentos, índice que representa a inflação dos alimentos. A base de dados é pública e foi obtida no site da FIPE¹⁰. Entende-se que essa transformação é necessária para garantir que a modelagem possa estimar os componentes hedônicos puros do preço dos produtos, sem impacto de eventuais efeitos inflacionários e comparando de maneira mais adequada as diferentes semanas que compõem a base de dados estudada. O ANEXO A traz a base de IPC-Alimentos considerada. Considera-se que o primeiro mês de estudo (janeiro/2019) é o mês de referência, e os preços de todos os demais meses foi deflacionado, dividindo-os pelo IPC-Alimentos acumulado entre janeiro/2019 e o período do dado.

3.4.1 Transformação BOX-COX

Adicionalmente à correção dos preços pelo índice de IPC-Alimentos, verificou-se a necessidade de realizar uma transformação Box-Cox na variável dependente preço. Dessa maneira, calculou-se o valor de λ que aproxima a distribuição da variável preço à normal, a partir da expressão $preço^\lambda$.

Figura 6 - λ estimado - Variável dependente preço

<pre>Estimated transformation parameter base_treino\$preco 1.026524</pre>

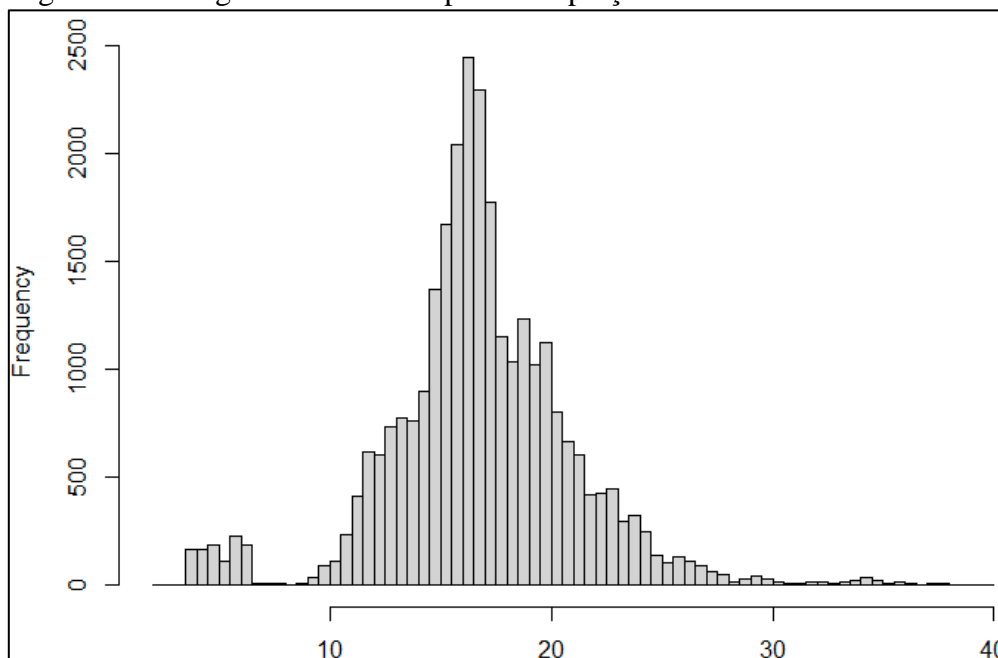
Fonte: elaborado pelo Autor.

Observa-se que o valor calculado de λ (1,0265) é muito próximo de 1, o que indica que a distribuição da variável original é bastante próxima da Distribuição Normal (Figura 7).

¹⁰ <https://www.fipe.org.br/pt-br/indices/ipc/#>

Dessa maneira, optou-se por não realizar a transformação Box-Cox no preço e seguir a modelagem com a variável original.

Figura 7 - Histograma variável dependente preço



Fonte: elaborado pelo Autor.

3.5 VARIÁVEIS EXPLICATIVAS

Nessa sessão, serão apresentados e discutidos os atributos dos produtos utilizados no modelo de regressão. Além disso, também serão descritas as transformações e ajustes realizados nos atributos obtidos, gerando assim as variáveis explicativas finais que foram utilizadas no modelo de regressão.

3.5.1 Características do Produto

Foram coletados quatro tipos de informações na embalagem dos produtos, consideradas potencialmente significativas à tomada de decisão de compra do consumidor: (Os ANEXOS B e C trazem exemplos dessas informações inspecionadas nas embalagens nos pontos de vendas.)

- (1) Tipo de proteína
- (2) Certificados de qualidade do animal
- (3) Certificados de qualidade do processo produtivo

(4) Marca

Tipo de proteína e Certificados de qualidade (do animal e do processo produtivo) são variáveis de interesse do nosso estudo, pois estão relacionadas à questão dos cuidados com a saúde, bem como aos movimentos vegano/vegetariano e ao crescente consumo de proteínas vegetais em substituição às de origem animal. Esses fatores justificam a presença dos referidos atributos neste estudo.

Foram identificados os seguintes certificados nos produtos estudados: *Certificado Angus*, *Certificado Orgânico*, *Certificado Vegano*, *Certificado Wagyu*, *Selo de sustentabilidade*, *Sem antibióticos*, *Sem transgênicos*. Desses, dois pares de certificados foram observados nos mesmos produtos e, portanto, serão analisados de maneira conjunta no modelo de regressão, por meio de uma variável que representa cada par de certificados: (1) *Certificado Wagyu + Selo de sustentabilidade*; (2) *Sem antibióticos + Sem transgênicos*. O Certificado Wagyu refere-se à raça do bovino utilizada para corte e é atrelada à diferenciação e qualidade do produto. Todos os demais certificados relacionam-se às questões de saudabilidade e sustentabilidade.

Quanto ao tipo de proteína, foram identificadas 14 distintas proteínas como ingrediente principal dos produtos: *Berinjela*, *Costela bovina*, *Edamame*, *Filé Mignon*, *Fraldinha*, *Frango*, *Grão de bico*, *Lentilha*, *Misto*, *Mix cogumelos*, *Outra Proteína Vegetal*, *Outro Corte bovino*, *Picanha* e *Soja*. Desses, alguns foram encontrados em apenas um produto, como *Edamame*, *Lentilha*, *Mix de cogumelos*, entre outros. Dessa maneira, buscando reduzir eventuais questões de multicolinearidade, foram criados grupos de tipos de proteína, garantindo que cada grupo possua ao menos dois produtos distintos, medida que atende à premissa do modelo de Preços Hedônicos de Rosen de que os atributos devem ser comuns a diferentes produtos do mercado estudado – garantindo o poder de opção do consumidor entre as ofertas que contenham alguns atributos semelhantes do vetor *Z*.

Além disso, uma parcela significativa dos produtos com proteína vegetal possui algum tipo de certificado de qualidade. Dessa maneira, também para reduzir questões de multicolinearidade entre os certificados e as proteínas, todas as proteínas vegetais foram agrupadas conjuntamente. O Quadro 3 mostra o resultado desses agrupamentos:

Quadro 3 - Grupos por Tipo de proteína

Grupos de Tipo de proteína	Produtos
Costela bovina	4
Costela bovina	4
Fraldinha	2
Fraldinha	2
Frango	6
Frango	6
Outro Corte bovino	2
Filé Mignon	1
Misto	1
Picanha	4
Picanha	4
Proteína Vegetal	7
Proteína Vegetal	7
Berinjela	1
Edamame	1
Grão de bico	1
Lentilha	1
Mix cogumelos	1
Soja	3
Carne bovina	16
Carne bovina	16
Total Geral	49

Fonte: elaborado pelo Autor.

Já as marcas desempenham papel significativo na decisão de compra de alimentos processados, portanto devem ser incluídas no estudo a fim de se evitar omissão de variáveis no modelo final, conforme apontado por Ahmad e Anders (2012). Para este estudo, foram consideradas as quatro maiores marcas (em volume de vendas de hambúrguer no período) e criadas variáveis representando-as.

Todas as características até então descritas neste tópico são variáveis categóricas. Dessa maneira, para inclusão delas no modelo de regressão criaram-se variáveis binárias para cada categoria distinta. Por fim, tais variáveis binárias foram multiplicadas pelo *Peso Total* da embalagem do produto em questão, pois entende-se que, no ato da compra, o consumidor considera, mesmo que de maneira intuitiva, o preço por quilograma de hambúrguer adquirido.

Sendo assim, as variáveis explicativas oriundas das características dos produtos representam o preço marginal por quilograma atrelado àquela característica/atributo do vetor *Z*.

Adicionalmente às variáveis categóricas mencionadas, foi incluída ao modelo uma variável referente à gramatura de cada unidade de hambúrguer contida na embalagem do produto (*peso_unitario*). Com isso, espera-se mensurar o potencial impacto do peso unitário de cada hambúrguer da embalagem no preço final do produto.

3.5.2 Nutrientes

Os produtos do segmento de hambúrguer comercializados em varejos possuem, em suas embalagens, uma tabela nutricional em formato padronizado, informando a quantidade de cada um de seus principais nutrientes por porção. O ANEXO A traz o exemplo da tabela nutricional de alguns dos produtos registrados na coleta de dados.

Em todos os 49 produtos registrados, foram encontradas as informações de *Valor energético (em kcal e kJ)*, *Carboidratos*, *Proteínas*, *Gorduras totais*, *Gorduras saturadas*, *Gorduras trans*, *Fibra alimentar* e *Sódio*. Dessa maneira, todas essas informações foram transformadas em variáveis e incorporadas ao estudo. Algumas informações adicionais foram encontradas em produtos específicos, como a quantidade de *Gorduras monoinsaturadas*, *Gorduras poli-insaturadas*, *Colesterol*, *Ferro*, *Zinco* e *Vitaminas* contidos no produto. Essas, porém, não foram consideradas no estudo, de maneira a mitigar problemas na modelagem de dados por excesso de dados faltantes.

A tabela nutricional dos produtos avaliados traz a quantidade em gramas de cada nutriente em cada porção consumida do produto. Essas porções, porém, não são padronizadas: foram identificados produtos cuja porção informada é de 80g, 100g, 170g, 210g etc. Assim, todas as variáveis nutricionais utilizadas no estudo foram padronizadas, dividindo-se a quantidade (em gramas) do nutriente pela porção (em gramas). A variável resultante é o percentual do nutriente na composição do produto. O Quadro 4 - Variáveis explicativas – Nutrientes traz a relação das variáveis de nutrientes utilizadas no estudo.

Quadro 4 - Variáveis explicativas - Nutrientes

Informação Nutricional	Variável	Descrição da Variável
Valor energético	Energia_kcal	Valor energético (kcal) para cada grama de produto consumido
Carboidratos	Carboidratos	Percentual (%) de carboidratos na composição do produto
Proteínas	Proteinas	Percentual (%) de proteínas na composição do produto
Gorduras totais	Gorduras	Percentual (%) de gorduras totais na composição do produto
Gorduras saturadas	Gorduras_saturadas	Percentual (%) de gorduras saturadas na composição do produto
Gorduras <i>trans</i>	Gorduras_trans	Percentual (%) de gorduras <i>trans</i> na composição do produto
Fibra alimentar	Fibra_alimentar	Percentual (%) de fibra alimentar na composição do produto
Sódio	Sodio	Percentual (%) de sódio na composição do produto

Fonte: elaborado pelo Autor.

Transformadas as variáveis conforme o Quadro 4, temos no Quadro 5, a estatística descritiva das variáveis resultantes:

Quadro 5 - Estatística descritiva - Nutrientes

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
Energia_kcal	28875	226.595	44.521	121	193.75	255	351
Carboidratos	28875	5.378	8.23	0	0	7.875	29.851
Proteinas	28875	15.677	4.214	4	13.846	18.75	24.75
Gorduras	28875	15.378	5.632	3.4	12.308	20	26.25
Gorduras_saturadas	28875	6.21	2.959	0	3.582	8.5	10.7
Gorduras_trans	28875	0.064	0.186	0	0	0	0.875
Sodio	28875	0.382	0.289	0.05	0.115	0.539	1.175
Fibra_alimentar	28875	1.291	1.986	0	0	1.418	7.4

Fonte: elaborado pelo Autor.

3.6 ANÁLISE FATORIAL

Em modelos de Regressão Multivariados, altas correlações entre as variáveis explicativas podem distorcer os coeficientes calculados, fazendo com que os mesmos não representem adequadamente a relação marginal de cada variável explicativa com a variável resposta. As técnicas exploratórias de Análise Fatorial permitem a identificação de multicolinearidade no conjunto de dados, bem como a construção de fatores que representam

de maneira adequada a variância do conjunto de dados original. Busca-se, em suma, representar o comportamento do conjunto de variáveis original, a partir de um número reduzido de novas variáveis (ou fatores).

A primeira etapa da análise fatorial é a confirmação da adequação do conjunto de dados e variáveis à análise. A Estatística Kaiser–Meyer–Olkin (KMO) e o teste de esfericidade de Bartlett são duas reconhecidas ferramentas para se verificar a adequação global da extração de fatores a partir dos dados originais (Fávero & Belfiore, 2017). A estatística KMO fornece a proporção de variância comum às variáveis originais, o que aponta para a existência de um fator comum que represente os dados. Essa estatística varia de 0 a 1 e a tabela abaixo indica a adequação da Análise Fatorial, de acordo com seu resultado.

Quadro 6 - Relação entre a estatística KMO e a adequação global da análise fatorial

Estatística KMO	Adequação Global da Análise Fatorial
Entre 1,00 e 0,90	Muito boa
Entre 0,90 e 0,80	Boa
Entre 0,80 e 0,70	Média
Entre 0,70 e 0,60	Razoável
Entre 0,60 e 0,50	Má
Menor do que 0,50	Inaceitável

Fonte: Fávero e Belfiore (2017).

Já o teste de esfericidade de Bartlett realiza a comparação da matriz de correlações do conjunto de variáveis com a matriz identidade, com um determinado número de graus de liberdade e a um determinado nível de significância, a partir da estatística $\chi^2_{Bartlett}$. Caso o valor calculado da estatística seja maior que o valor crítico, entende-se que a matriz resultante do conjunto de pares de correlações de Pearson entre as variáveis é diferente da matriz identidade e, portanto, o conjunto de dados é adequado para a Análise Fatorial.

Garantida a adequação dos dados, pode-se prosseguir para a Análise Fatorial em si. Neste estudo, foi utilizada a Análise Fatorial por Componentes Principais (PCA), onde são calculados fatores a partir da combinação linear das variáveis originais. Os fatores principais resultantes da análise fatorial são, por definição, ortogonais entre si, o que garante uma tratativa adequada à questão da multicolinearidade. A expressão (3.1) traz o cálculo de cada fator F_j , onde X_i representa a i -ésima variável da base analisada e s_{ij} é conhecido como o *score fatorial*, que equivale ao parâmetro de um modelo linear que relaciona o fator F_j às variáveis originais.

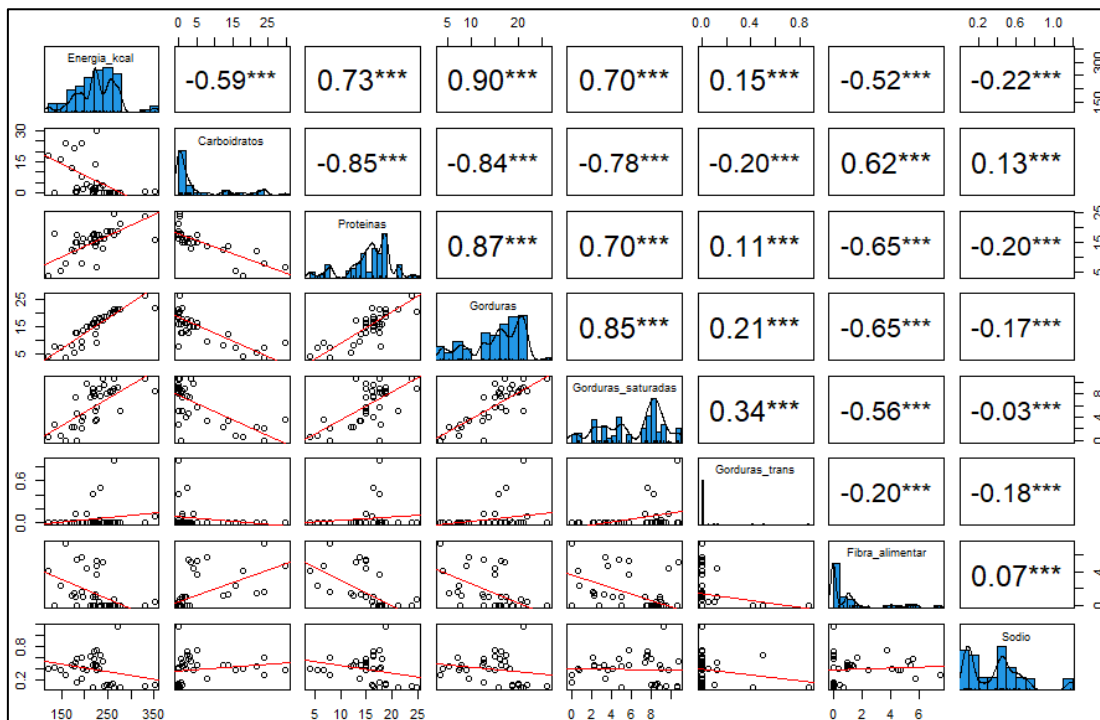
$$F_j = \sum_{i=1}^n s_{ij}X_i \quad (3.1)$$

Cada fator calculado possui um autovalor λ_j , que representa a variância do conjunto de dados explicada pelo respectivo fator. Por definição, o primeiro fator calculado por Componentes Principais é o que apresenta o maior autovalor – que cai sucessivamente para cada fator conseguinte. De acordo com Fávero e Belfiore (2017), o critério mais adotado para escolha da quantidade de fatores necessários para representar o conjunto de dados é o critério da raiz latente (critério de Kaiser), onde são considerados apenas fatores com autovalores correspondentes maiores que 1.

3.6.1 Aplicação da análise fatorial

No problema aqui estudado, observou-se um alto nível de correlação entre as variáveis numéricas de nutrientes que compõem os produtos. O quadro a seguir traz uma visão geral das distribuições dessas variáveis, assim como o cálculo de correlação de Pearson entre as mesmas:

Figura 8 - Distribuição e correlação das variáveis de Nutrientes



Fonte: elaborado pelo Autor.

Fávero e Belfiore (2017) apontam que uma matriz de correlação com muitos valores inferiores a 0,30 representa um indício preliminar de que a análise fatorial não é adequada para um determinado conjunto de dados. Dessa maneira, as variáveis *Gordura_trans* e *Sodio* foram excluídas da análise fatorial e foram mantidas como variáveis explicativas do modelo original. Para as demais variáveis nutricionais (*Energia_kcal*, *Carboidratos*, *Proteinas*, *Gorduras*, *Gorduras_saturadas* e *Fibra_alimentar*) seguiu-se com a análise fatorial, dado que essas apresentam correlações de alta intensidade, tanto positivas quanto negativas – o que aponta para a adequação da realização da análise fatorial e construção de um número reduzido de fatores principais que representem tais variáveis.

A Estatística KMO calculada para as variáveis foi de 0,78, conforme Figura 9. Esse resultado aponta para a adequação da Análise Fatorial. Além disso, a estatística $\chi^2_{Bartlett}$ apresentou um valor de 21.1797,5 - muito acima do valor crítico para os graus de liberdade considerados. O p-valor do teste tende a zero e a hipótese de igualdade entre a matriz identidade e a matriz de correlações das variáveis é rejeitada. Esse resultado também corrobora a realização da Análise Fatorial dos dados.

Figura 9 - Estatística KMO para *Energia_kcal*, *Carboidratos*, *Proteinas*, *Gorduras*, *Gorduras_saturadas* e *Fibra_alimentar*

Kaiser-Meyer-Olkin factor adequacy						
Call: KMO(r = base_predictoras)						
overall MSA = 0.78						
MSA for each item =						
	<i>Energia_kcal</i>	<i>Carboidratos</i>	<i>Proteinas</i>	<i>Gorduras</i>	<i>Gorduras_saturadas</i>	<i>Fibra_alimentar</i>
	0.65	0.71	0.89	0.72	0.92	0.94

Fonte: elaborado pelo Autor.

Dessa forma, foi realizada a Análise Fatorial por Componente Principal para as variáveis *Energia_kcal*, *Carboidratos*, *Proteinas*, *Gorduras*, *Gorduras_saturadas* e *Fibra_alimentar*. A análise gerou seis fatores, que são apresentados em ordem, de acordo com a proporção da variância do conjunto de dados representada no fator. Nota-se que 77,19% da variância do conjunto de dados é representada pelo primeiro componente PC1. O Quadro 7 traz um resumo das estatísticas calculadas para os fatores.

Quadro 7 - Fatores PCA para Energia_kcal, Carboidratos, Proteínas, Gorduras, Gorduras_saturadas e Fibra_alimentar

Fator	Desvio Padrão	Proporção da variância	Autovalor λ
PC1	2,1521	0,7719	4,6315
PC2	0,7345	0,0899	0,5395
PC3	0,6418	0,0687	0,4119
PC4	0,5483	0,0501	0,3006
PC5	0,3066	0,0157	0,0940
PC6	0,1497	0,0037	0,0224

Fonte: elaborado pelo Autor.

Observa-se que o fator PC1 é o único fator com autovalor acima de 1. Assim, seguindo o critério da raiz latente (critério de Kaiser), seguiu-se com esse único fator para as etapas seguintes da modelagem, em substituição às variáveis originais *Energia_kcal*, *Carboidratos*, *Proteínas*, *Gorduras*, *Gorduras_saturadas* e *Fibra_alimentar*.

Por fim, o Quadro 8 traz a relação do *score fatorial* de cada variável para o componente PC1. Como mencionado, o *score fatorial* equivale ao parâmetro de um modelo linear envolvendo o fator com as variáveis originais. Sendo assim, a partir desses valores, nota-se que o componente PC1 é positivamente relacionado com as variáveis *Carboidratos* e *Fibra_alimentar*, enquanto é negativamente relacionado com as variáveis *Energia_kcal*, *Proteínas*, *Gorduras* e *Gorduras_saturadas*. Esses valores serão posteriormente importantes no entendimento dos resultados das regressões realizadas e na validação dos constructos propostos, pois permitem relacionar de maneira direcional as variáveis representadas pelo componente PC1 ao seu componente hedônico a ser estimado.

Quadro 8 - Score fatorial para Energia_kcal, Carboidratos, Proteínas, Gorduras, Gorduras_saturadas e Fibra_alimentar

	PC1
Energia_kcal	-0.3924933
Carboidratos	0.4152209
Proteínas	-0.4259329
Gorduras	-0.4527395
Gorduras_saturadas	-0.4071321
Fibra_alimentar	0.3484137

Fonte: elaborado pelo Autor.

3.7 MODELO DE REGRESSÃO

Diferentemente da maior parte dos trabalhos com aplicação de preços hedônicos citados, esta dissertação utiliza dados de vendas agregadas (por semana, região e produto) para a aplicação do primeiro estágio do modelo de preços hedônicos de Rosen. Em seu trabalho sobre aplicação de preços hedônicos a produtos embalados, Martínez-Garmendia (2010) menciona os desafios de se utilizarem dados consolidados de vendas para a aplicação do modelo de preços hedônicos. O autor argumenta que, com a utilização de dados agregados de vendas, a preferência de consumo não é medida no nível individual, o que pode impedir que a análise capture de maneira adequada o comportamento de consumo individual para cada ato de compra. Martínez-Garmendia propõe, em seu trabalho, a utilização da variável de quantidade de vendas como variável preditora do modelo. Nesse contexto, o autor utiliza um modelo de regressão em dois estágios (2SLS) de maneira a endereçar questões de endogeneidade causadas pela simultaneidade entre quantidade de vendas e preço.

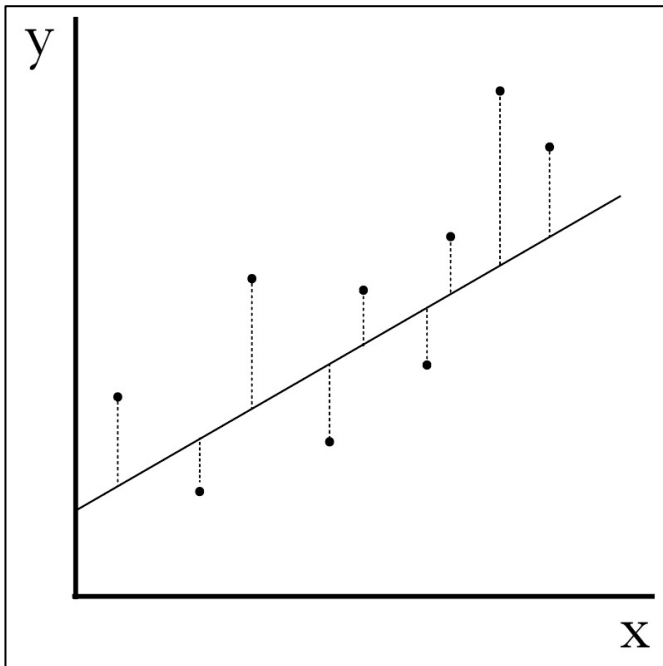
Nesta pesquisa, buscou-se uma alternativa para lidar com os desafios de utilizar dados consolidados de vendas. Adicionalmente à regressão dos dados via Método dos Mínimos Quadrados Ordinários (OLS), encontrada na maior parte das aplicações de preços hedônicos, também foi realizada a regressão via Método dos Mínimos Quadrados Ponderados (WLS). É uma solução similar à observada no estudo de Bimbo et al. (2016). Nas próximas seções, far-se-á um breve resumo desses diferentes métodos.

3.7.1 Método dos mínimos quadrados ordinários (OLS)

No Método dos Mínimos Quadrados Ponderados (OLS), busca-se a estimação dos parâmetros de um modelo linear de regressão a partir da minimização do termo de erro obtido entre os valores reais da variável dependente e seus valores projetados através do modelo. No caso mais simples, tem-se a seguinte equação, onde se deseja estimar A e B de maneira a traçar a reta que minimiza sua distância com relação aos pares de dados, conforme Figura 10:

$$Y = AX + B \tag{3.2}$$

Figura 10 - Resíduos da regressão OLS



Fonte: elaborado pelo Autor.

De acordo com De Levie (1986), a minimização da somatória dos resíduos (3.3) pode ser obtida a partir dos seguintes valores de A (3.4) e B (3.5):

$$\min \sum_i (Y_i - (AX_i + B))^2 \quad (3.3)$$

$$A = \frac{N \sum_i X_i Y_i - \sum_i X_i \sum_i Y_i}{N \sum_i X_i^2 - (\sum_i X_i)^2} \quad (3.4)$$

$$B = \frac{\sum_i X_i^2 \sum_i Y_i - \sum_i X_i \sum_i X_i Y_i}{N \sum_i X_i^2 - (\sum_i X_i)^2} \quad (3.5)$$

3.7.2 Método dos mínimos quadrados ponderados – *Weighted Least Squares* (WLS)

O Método dos Mínimos Quadrados Ponderados é um caso particular do Método dos Mínimos Quadrados Generalizados e é uma alternativa de modelagem para casos nos quais há presença de termos de erro heterocedásticos (Fávoro & Belfiore, 2017). Nele, podem-se atribuir diferentes pesos aos diferentes pontos de dados no cálculo dos coeficientes do modelo (De Levie, 1986).

Partindo da equação original (3.2), inclui-se a constante W_i à equação do termo de erro e, conseqüentemente, às equações dos parâmetros A e B :

$$\min \sum_i W_i (Y_i - (AX_i + B))^2 \quad (3.6)$$

$$A = \frac{\sum_i W_i \sum_i W_i X_i Y_i - \sum_i W_i X_i \sum_i W_i Y_i}{\sum_i W_i \sum_i W_i X_i^2 - (\sum_i W_i X_i)^2} \quad (3.7)$$

$$B = \frac{\sum_i W_i X_i^2 \sum_i W_i Y_i - \sum_i W_i X_i \sum_i W_i X_i Y_i}{\sum_i W_i \sum_i W_i X_i^2 - (\sum_i W_i X_i)^2} \quad (3.8)$$

Observa-se que, na situação particular em que todo $W_i = 1$, as equações do WLS se igualam às do OLS. A ideia por trás desse método é atribuir diferentes pesos aos pontos de dados para o problema de minimização dos resíduos.

Dessa maneira, diferentemente da utilização do WLS (que é tradicionalmente uma alternativa de modelagem em casos de heterocedasticidade), neste trabalho o método utilizado como uma tentativa de adaptação do modelo de dados ao constructo teórico de preços hedônicos, que prevê a utilização de dados não agregados das transações realizadas.

Em termos práticos, utilizou-se da quantidade vendida (em unidades) como constante W_i para cada um dos pontos da base de dados. Metodologia similar também foi encontrada no estudo de Bimbo et al. (2016). Porém, na ocasião, os autores utilizaram a raiz quadrada da quantidade de vendas para ponderar a regressão, enquanto aqui será utilizada a quantidade vendida sem nenhuma transformação. Com isso, espera-se que o modelo de regressão WLS dê mais peso à minimização dos erros para produtos / semanas com mais unidades vendidas. Em outras palavras: caso a base de dados utilizada tivesse a granularidade por transação de venda, consequentemente produtos com mais unidades vendidas teriam mais pontos de dados e, dessa maneira, maior influência na estimação dos parâmetros do modelo. Com o WLS busca-se simular este comportamento, a partir da atribuição do peso proporcional à quantidade vendida.

A tabela ilustra a diferença dos pesos por produto, considerando o critério de número de pontos de dados N e o critério de unidades vendidas W_i . Nota-se que, pelo método OLS, os pesos ficam distribuídos de maneira mais igualitária entre os produtos ($\%N$), enquanto no WLS há uma discrepância maior no peso ($\%W$) atribuído a cada produto.

Tabela 1 - Proporções de W_i e N

Produto	%N	%W			
1	3,25%	32,80%	25	2,68%	0,54%
2	3,25%	17,33%	26	2,59%	0,35%
3	3,25%	7,33%	27	0,75%	0,34%
4	2,62%	4,00%	28	1,95%	0,28%
5	3,14%	3,75%	29	3,09%	0,26%
6	3,25%	2,99%	30	2,87%	0,26%
7	3,19%	2,92%	31	1,96%	0,25%
8	3,13%	2,63%	32	2,97%	0,23%
9	3,24%	2,53%	33	0,37%	0,23%
10	3,25%	2,26%	34	1,01%	0,19%
11	3,25%	1,99%	35	1,07%	0,17%
12	3,24%	1,87%	36	1,11%	0,13%
13	2,45%	1,80%	37	1,22%	0,11%
14	3,24%	1,64%	38	1,00%	0,09%
15	2,70%	1,49%	39	0,82%	0,07%
16	3,24%	1,43%	40	0,76%	0,06%
17	1,00%	1,35%	41	0,85%	0,05%
18	2,60%	1,34%	42	1,41%	0,04%
19	2,51%	1,20%	43	0,38%	0,03%
20	3,19%	0,86%	44	0,34%	0,02%
21	2,47%	0,80%	45	0,43%	0,02%
22	3,22%	0,78%	46	0,38%	0,01%
23	2,95%	0,62%	47	0,16%	0,01%
24	1,80%	0,56%	48	0,23%	0,01%
			49	0,21%	0,01%

Fonte: elaborado pelo Autor.

3.7.3 Pressupostos dos modelos de regressão

O Quadro 9 traz alguns pressupostos que devem ser verificados para modelos de regressão, bem como as consequências em caso de violação dos mesmos:

Quadro 9 - Pressupostos do modelo de regressão

Pressuposto	Violação	Consequência
Os resíduos apresentam distribuição normal	Valor-P dos testes t e do teste F não são válidos	Potencial problema de especificação do modelo
Não existem correlações elevadas entre as variáveis explicativas e existem mais observações do que variáveis explicativas.	Multicolinearidade	Impacto na estimativa dos parâmetros da regressão
Os resíduos não apresentam correlação com qualquer variável X .	Heterocedasticidade	Erros-padrão viesados dos parâmetros da regressão
Os resíduos são aleatórios e independentes	Autocorrelação dos resíduos para modelos temporais	Potencial problema de especificação do modelo

Fonte: Fávero e Belfiore (2017).

Considerando que a modelagem deste trabalho não tem fins preditivos, os pressupostos de normalidade e autocorrelação dos resíduos não serão verificados. Entende-se que um eventual problema de especificação do modelo poderia reduzir sua capacidade preditiva, porém não influencia os parâmetros estimados.

A multicolinearidade dos modelos será testada através das estatísticas *VIF* e *Tolerance*, que são dadas pelas seguintes expressões, onde R_k^2 se refere à estatística R^2 para a regressão da k-ésima variável explicativa do modelo pelas demais variáveis.

$$Tolerance = 1 - R_k^2 \quad (3.9)$$

$$VIF = \frac{1}{Tolerance} \quad (3.10)$$

Dessa maneira, de acordo com Fávero e Belfiore (2017), uma estatística *Tolerance* de 0,20 (ou um *VIF* de 5) significa que a variável explicativa compartilha 80% de variância com as demais – o que é um indício importante de colinearidade da variável em questão.

Com relação à heterocedasticidade, Fávero e Belfiore (2017) citam o teste de Breuch-Pagan/Cook-Weisberg, que apresenta, como hipótese nula, o fato de a variância dos termos de erro ser constante (erros homocedásticos) e, como hipótese alternativa, o fato de a variância dos termos de erro não ser constante, ou seja, os termos de erro serem uma função de uma ou mais variáveis explicativas (erros heterocedásticos). Fávero e Belfiore (2017) também citam o método de Huber-White para erros-padrão robustos, método que calcula os erros-padrão dos parâmetros do modelo robustos a heterocedasticidade. Dessa maneira, caso seja verificada a heterocedasticidade dos modelos, será utilizado o método de Huber-White, garantindo, assim, a robustez dos resultados.

4 APRESENTAÇÃO DOS RESULTADOS

Nesta seção, serão apresentados os resultados dos modelos estimados para os dados, tanto via Método dos Mínimos Quadrados Ordinários (OLS), quanto via Método dos Mínimos Quadrados Ponderados (WLS).

Em ambos os modelos foi realizada a regressão da variável de interesse preço, considerando as variáveis explicativas expostas na seção anterior. O Quadro 10 traz todas as variáveis explicativas, oriundas dos atributos levantados, que foram incluídas ao modelo:

Quadro 10 - Variáveis explicativas do modelo

Tipo de característica	Variável	Descrição da Variável
Certificados	Certificado_vegano	Interação do Valor binário da categoria * Peso da embalagem
	Certificado_wagyu_e_selo_sustentabilidade	
	Certificado_angus	
	Sem_antibioticos_e_transgenicos	
	Certificado_organico	
Marca	MARCA1	
	MARCA2	
	MARCA3	
	MARCA4	
Nutrientes	Gorduras_trans	% de Gorduras <i>trans</i> na composição do produto
	Sodio	% de Sódio na composição do produto
	fator_nutrientes_1	Componente extraído da análise fatorial dos nutrientes
Tipo de proteína	Costela.bovina	Interação do Valor binário da categoria * Peso da embalagem
	Proteina.Vegetal_19	Interação do Valor binário da categoria no ano 2019 * Peso da embalagem
	Proteina.Vegetal_20	Interação do Valor binário da categoria no ano 2020 * Peso da embalagem
	Frango	Interação do Valor binário da categoria * Peso da embalagem
	Outro. Corte.bovino	
	Picanha	
	Fraldinha	
Outros	peso_unitario	Gramatura de cada unidade de hambúrguer contida na embalagem

Fonte: elaborado pelo Autor.

A variável *Carne.bovina*, referente aos produtos cuja embalagem informa o ingrediente “Carne Bovina”, foi excluída do modelo, dado que seria o resultado da combinação linear dos outros tipos de ingredientes utilizados. Dessa maneira, os demais parâmetros calculados para as demais proteínas já trarão o comparativo entre a respectiva proteína e a carne bovina padrão, que é a referência de mercado.

Outro ajuste importante realizado nas variáveis explicativas foi a quebra da *Proteína.Vegetal* em três variáveis – uma para cada ano contido na base de dados. Com isso, espera-se mensurar a evolução do preço hedônico da proteína vegetal, uma vez que esse tipo de proteína ainda faz parte de um mercado em crescimento e que, conseqüentemente, pode ter oscilações importantes em seus preços.

4.1 CARACTERIZAÇÃO DA AMOSTRA

A amostra de dados utilizada para os modelos e análises possui 28.843 pontos, distribuídos entre 49 produtos, 6 regiões de vendas e 156 semanas. O ANEXO 0 traz uma série de gráficos e tabelas para caracterização da base de dados e variáveis. Os gráficos e tabelas da Figura 18 até a Figura 26 trazem as frequências dos dados, frequências ponderadas por W (utilizadas no WLS) e gráficos Box Plot do preço (já deflacionado) quebrados nas variáveis de interesse do estudo.

Observa-se, pela Figura 18, que Carne Bovina possui frequências N e W acima de 30%, enquanto a Proteína Vegetal tem uma frequência N de 25,13%, porém apenas 9,15% de frequência W, o que indica uma diferença importante para a comparação dos modelos OLS e WLS. Com relação ao gráfico Box Plot de preços, a Proteína Vegetal foi a que apresentou maior preço médio (R\$ 18,71), enquanto Fraldinha apresentou o menor (R\$ 13,85). O maior preço identificado para Proteína Vegetal foi de R\$ 33,33, enquanto o mínimo foi de R\$ 7,41.

A evolução anual dos preços, observada na Figura 21, revela estabilidade nos preços deflacionados, com média de R\$ 16,73, R\$ 16,42 e R\$ 17,48 para 2019, 2020 e 2021 respectivamente. Por outro lado, há um aumento importante na variabilidade dos preços em 2021, com aumento da distância entre o primeiro e terceiro quartil da distribuição, e preço máximo de R\$ 39,83.

De todos os certificados e selos (Figura 22 a Figura 26), apenas produtos com selo Sem Antibióticos e Transgênicos não apresentam preço médio superior em comparação com o restante da amostra. A maior diferença no preço médio ocorre para o Certificado Wagyu e Selo de Sustentabilidade, com um preço médio de R\$ 34,28, contra R\$ 16,78 do restante da

amostra. Adicionalmente, todos os certificados e selos são pouco frequentes na base, sendo mais encontrados o Certificado Angus (Frequência N = 10,82% e Frequência W = 4,24%) e o Certificado Vegano (Frequência N = 11,23% e Frequência W = 1,55%).

Por fim, a Figura 27 traz um gráfico de dispersão do preço médio já deflacionado (R\$/unidade) dos 49 produtos pelo Peso Total da embalagem (Kg). O preço médio mais alto observado é de R\$34,28, enquanto o maior Peso Total em uma embalagem é de 0,9Kg. No caso dos produtos à base de Proteína Vegetal, o produto com maior preço médio (R\$ 24,84) é também o com maior Peso Total (0,46 Kg).

4.2 APLICAÇÃO DO MÉTODO DOS MÍNIMOS QUADRADOS ORDINÁRIOS (OLS)

Foi estimado o modelo OLS para as variáveis explicativas e dependente mencionadas. Tal modelo estimou um parâmetro não significativo para a variável *Picanha* e, então, utilizou-se do algoritmo *Stepwise* de seleção de variáveis para expurgar do modelo eventuais variáveis residuais e não-significativas estatisticamente. A estatística $R^2_{Ajustado}$ para o modelo foi de 0,5525, indicando que 55,25% da variabilidade da variável dependente preço é representada pela regressão realizada.

Para o modelo estimado, realizou-se o teste de Breuch-Pagan/Cook-Weisberg, que confirmou a presença de heterocedasticidade:

Figura 11 - Teste de Breuch-Pagan/Cook-Weisberg - OLS

```
studentized Breusch-Pagan test
data: step_lm1
BP = 4828.3, df = 20, p-value < 2.2e-16
```

Fonte: elaborado pelo Autor.

Sendo assim, utilizou-se do método de Huber-White para erros-padrão robustos, gerando os parâmetros finais da regressão OLS:

Figura 12 - Outputs do modelo OLS

t test of coefficients:					
	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	14.59946	0.12100	120.6588	< 2.2e-16	***
Proteina.Vegetal_19	16.09294	0.43193	37.2584	< 2.2e-16	***
Proteina.Vegetal_20	14.12265	0.41874	33.7262	< 2.2e-16	***
Proteina.Vegetal_21	12.74604	0.41720	30.5514	< 2.2e-16	***
Costela.bovina	17.31867	0.24877	69.6162	< 2.2e-16	***
Frango	3.43254	0.17898	19.1785	< 2.2e-16	***
Outro.Corte.bovino	-2.21093	0.12030	-18.3777	< 2.2e-16	***
Fraldinha	-4.02612	0.23618	-17.0471	< 2.2e-16	***
MARCA_2	8.39422	0.13158	63.7959	< 2.2e-16	***
MARCA_1	-1.67147	0.20575	-8.1238	4.697e-16	***
MARCA_3	6.16893	0.18274	33.7574	< 2.2e-16	***
MARCA_4	13.89152	0.25657	54.1440	< 2.2e-16	***
Gorduras_trans	0.61368	0.11753	5.2216	1.786e-07	***
Sodio	-5.85919	0.10795	-54.2768	< 2.2e-16	***
Certificado_vegano	2.02090	0.32933	6.1363	8.554e-10	***
Certificado_wagyu_e_selo_sustentabilidade	52.87377	0.70817	74.6629	< 2.2e-16	***
Certificado_angus	8.56908	0.17384	49.2941	< 2.2e-16	***
Sem_antibioticos_e_transgenicos	3.99627	0.63719	6.2717	3.623e-10	***
Certificado_organico	21.85510	0.40602	53.8279	< 2.2e-16	***
peso_unitario	-4.97186	0.63610	-7.8162	5.628e-15	***
fator_nutrientes_1	0.57159	0.02352	24.3022	< 2.2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

Fonte: elaborado pelo Autor.

Adicionalmente, calculou-se a estatística *VIF* para as variáveis explicativas, conforme Figura 13.

Figura 13 - Estatística VIF - OLS

Variável	VIF
Proteina.Vegetal_19	2.85
Proteina.Vegetal_20	3.3
Proteina.Vegetal_21	3.53
Costela.bovina	2.12
Frango	2.89
Outro.Corte.bovino	1.7
Fraldinha	1.24
MARCA_2	2.71
MARCA_1	2.15
MARCA_3	3.5
MARCA_4	3.75
Gorduras_trans	2.17
Sodio	2.05
Certificado_vegano	4.34
Certificado_wagyu_e_selo_sustentabilidade	1.09
Certificado_angus	2.17
Sem_antibioticos_e_transgenicos	1.4
Certificado_organico	1.7
peso_unitario	3.83
fator_nutrientes_1	6.89

Fonte: elaborado pelo Autor.

Nota-se que a variável *fator_nutrientes_1*, resultante da análise fatorial, apresenta uma estatística *VIF* moderadamente alta de 6,89. Esse resultado é, de certa forma, esperado, pois diferentes ingredientes possuem diferentes composições nutricionais. Sendo assim, a substituição do tipo de proteína como ingrediente do hambúrguer acarreta em uma diferente composição nutricional.

4.3 APLICAÇÃO DO MÉTODO DOS MÍNIMOS QUADRADOS PONDERADOS (WLS)

Foi estimado o modelo WLS para as variáveis explicativas e dependente mencionadas. A estatística $R^2_{Ajustado}$ para o modelo foi de 0,6445, indicando que 64,45% da variabilidade da variável dependente preço é representada pela regressão realizada.

Para o modelo estimado, realizou-se o teste de Breuch-Pagan/Cook-Weisberg, que confirmou a presença de heterocedasticidade:

Figura 14 - Teste de Breuch-Pagan/Cook-Weisberg - WLS

```

studentized Breusch-Pagan test

data: lm2
BP = 6511905, df = 21, p-value < 2.2e-16

```

Fonte: elaborado pelo Autor.

Sendo assim, utilizou-se do método de Huber-White para erros-padrão robustos, gerando os parâmetros finais da regressão OLS:

Figura 15 - Outputs do modelo WLS

```

t test of coefficients:

              Estimate Std. Error  t value  Pr(>|t|)
(Intercept)    16.585523   0.134583  123.2366 < 2.2e-16 ***
Proteina.vegetal_19  11.051058   0.432530   25.5498 < 2.2e-16 ***
Proteina.vegetal_20   9.948352   0.419395   23.7207 < 2.2e-16 ***
Proteina.vegetal_21   5.934489   0.417602   14.2109 < 2.2e-16 ***
Costela.bovina    18.778219   0.244856   76.6907 < 2.2e-16 ***
Frango           -2.193124   0.179199  -12.2385 < 2.2e-16 ***
Outro.Corte.bovino -5.315749   0.143605  -37.0163 < 2.2e-16 ***
Picanha          -1.496754   0.248669   -6.0191 1.775e-09 ***
Fraldinha        -8.213202   0.295181  -27.8243 < 2.2e-16 ***
MARCA_2           8.198636   0.132157   62.0371 < 2.2e-16 ***
MARCA_1           2.580307   0.205740   12.5416 < 2.2e-16 ***
MARCA_3           8.795988   0.191419   45.9514 < 2.2e-16 ***
MARCA_4          19.763819   0.335611   58.8891 < 2.2e-16 ***
Gorduras_trans    -1.072757   0.130058   -8.2483 < 2.2e-16 ***
Sodio            -6.860314   0.123792  -55.4180 < 2.2e-16 ***
Certificado_vegano  2.435259   0.333511    7.3019 2.911e-13 ***
Certificado_wagyu_e_selo_sustentabilidade 52.913019   0.708675   74.6647 < 2.2e-16 ***
Certificado_angus   8.001020   0.173865   46.0186 < 2.2e-16 ***
Sem_antibioticos_e_transgenicos -5.460692   0.640584   -8.5246 < 2.2e-16 ***
Certificado_organico 28.394003   0.410318   69.2001 < 2.2e-16 ***
peso_unitario     -20.063393   0.704904  -28.4626 < 2.2e-16 ***
fator_nutrientes_1  1.032615   0.023503   43.9359 < 2.2e-16 ***

```

Fonte: elaborado pelo Autor.

Adicionalmente, calculou-se a estatística *VIF* para as variáveis explicativas, conforme Figura 16:

Figura 16 - Estatística VIF - WLS

Variável	VIF
Proteina.Vegetal_19	2.61
Proteina.Vegetal_20	3.33
Proteina.Vegetal_21	2.7
Costela.bovina	2.07
Frango	2.73
Outro.Corte.bovino	15.83
Picanha	3.33
Fraldinha	2.54
MARCA_2	5.66
MARCA_1	2.69
MARCA_3	19.03
MARCA_4	10.79
Gorduras_trans	3.28
Sodio	5.3
Certificado_vegano	2.79
Certificado_wagyu_e_selo_sustentabilidade	1.01
Certificado_angus	2.41
Sem_antibioticos_e_transgenicos	2.91
Certificado_organico	1.22
peso_unitario	11.27
fator_nutrientes_1	5.85

Fonte: elaborado pelo Autor.

Observa-se que a estatística *VIF* aqui calculada difere bastante da calculada na OLS, por conta da alteração do critério de ponderação dos erros-padrão dos modelos auxiliares calculados. Algumas variáveis como *MARCA_3*, *MARCA_4* e *peso_unitario* apresentaram *VIF* acima de 10, o que pode prejudicar a estimação de seus parâmetros estimados pelo modelo WLS. De qualquer forma, nossas principais variáveis de interesse (tipo de proteína, certificados e nutrientes) permaneceram com *VIF* moderadamente baixo, com exceção para o *fator_nutrientes_1*, que manteve a estatística acima de 5, assim como na OLS.

4.4 DISCUSSÃO

Nesta seção serão discutidos os resultados obtidos a partir dos modelos OLS e WLS estimados.

Inicialmente, observa-se que ambos os modelos apresentaram valores moderados para a estatística $R^2_{Ajustado}$ (0,55 no OLS e 0,64 no WLS), o que indica que uma parcela significativa da variabilidade da variável dependente não é explicada pelas regressões realizadas. Dado que o estudo não tem fins preditivos, as análises de multicolinearidade e heterocedasticidade são suficientes para garantir a validade dos parâmetros calculados pelos modelos, porém os valores moderados de $R^2_{Ajustado}$ indicam a possível omissão de variáveis nos modelos, sejam elas intrínsecas ou extrínsecas aos produtos.

Com relação aos parâmetros calculados, o Quadro 11 traz, de maneira resumida, os coeficientes e estatística VIF calculados para ambos os modelos, já sendo excluído o coeficiente de Picanha para o OLS, que não é estatisticamente significativo:

Quadro 11 - Coeficientes e VIF OLS/WLS

Tipo variável	Variável	Coeficiente		VIF	
		OLS	WLS	OLS	WLS
Intercepto	(Intercept)	14,60	16,59		
Proteínas	Proteína.Vegetal_19	16,09	11,05	2,85	2,61
	Proteína.Vegetal_20	14,12	9,95	3,30	3,33
	Proteína.Vegetal_21	12,75	5,93	3,53	2,70
	Costela.bovina	17,32	18,78	2,12	2,07
	Frango	3,43	-2,19	2,89	2,73
	Outro.Corte.bovino	-2,21	-5,32	1,70	15,83
	Picanha		-1,50		3,33
	Fraldinha	-4,03	-8,21	1,24	2,54
Nutrientes	Gorduras_trans	0,61	-1,07	2,17	3,28
	Sodio	-5,86	-6,86	2,05	5,30
	fator_nutrientes_1	0,57	1,03	6,89	5,85
Certificados	Certificado_vegano	2,02	2,44	4,34	2,79
	Certificado_wagyu_e_selo_sustentabilidade	52,87	52,91	1,09	1,01
	Certificado_angus	8,57	8,00	2,17	2,41
	Sem_antibioticos_e_transgenicos	4,00	-5,46	1,40	2,91
	Certificado_organico	21,86	28,39	1,70	1,22
Marcas	MARCA_2	8,39	8,20	2,71	5,66
	MARCA_1	-1,67	2,58	2,15	2,69
	MARCA_3	6,17	8,80	3,50	19,03
	MARCA_4	13,89	19,76	3,75	10,79
Outros	peso_unitario	-4,97	-20,06	3,83	11,27

Fonte: elaborado pelo Autor.

O primeiro achado significativo a partir do quadro é o fato de que as três variáveis de Proteína Vegetal (*Proteína.Vegetal_19*, *Proteína.Vegetal_20* e *Proteína.Vegetal_21*) são estatisticamente significativas e possuem parâmetros positivos em ambos os modelos. Isso corrobora a hipótese inicial de que o consumidor está disposto a pagar um valor adicional por quilograma por produtos que utilizam proteína vegetal, comparativamente à proteína bovina utilizada como referência nos modelos. A proteína vegetal também apresenta um preço hedônico superior às outras proteínas estudadas (Frango, Picanha, Fraldinha e Outros Cortes bovinos), excetuando-se a Costela bovina, cujo preço hedônico é maior que o da Proteína Vegetal, tanto no OLS quanto no WLS.

Em ambos os modelos, também se observa uma diminuição ano a ano do preço hedônico por quilograma de proteína vegetal no produto. Isso pode acontecer por um amadurecimento do mercado para esse tipo de proteína, seja pelo aumento de competição a partir da entrada de novos produtos concorrentes, seja pela popularização do segmento e adaptação dos preços para atingir diferentes camadas socioeconômicas de consumidores.

Com relação aos nutrientes, Sódio e o *fator_nutrientes_1* apresentam coeficientes concordantes em ambos os modelos, positivos e negativos respectivamente. Já as Gorduras *trans* apresentam coeficiente positivo para o modelo OLS e negativo para o WLS. Os coeficientes negativos para Sódio corroboram as hipóteses iniciais do estudo, uma vez que o excesso desse nutriente é negativamente atrelado à saudabilidade. No caso das Gorduras *trans*, esperavam-se também coeficientes negativos, porém esse resultado foi observado somente no WLS.

O *fator_nutrientes_1* é o resultado da combinação dos seguintes nutrientes e seus respectivos pesos: *Energia_kcal* (-0,39); *Carboidratos* (0,41); *Proteínas* (-0,42); *Gorduras* (-0,45); *Gorduras_saturadas* (-0,41) e; *Fibra_alimentar* (0,34). Logo, como o coeficiente da variável *fator_nutrientes_1* é positivo em ambos os modelos, entende-se que a redução das calorias, proteínas, gorduras totais e gorduras saturadas contribui para o aumento do preço do produto – mesmo efeito observado para o aumento dos carboidratos e das fibras alimentares. De qualquer forma, sabe-se que as referidas variáveis possuem alto nível de correlação entre si, o que dificulta sua análise individualizada.

Com exceção da variável *Sem_antibioticos_e_transgenicos*, todos os certificados apresentaram coeficientes positivos e estatisticamente significativos. O *Certificado_organico* apresentou parâmetros com valor superior a 20 em ambos os modelos, o que reforça o constructo do estudo de que um nicho de consumidores valoriza e está disposto a pagar mais de R\$20/Kg em produtos que apresentem tal certificado – que é relacionado às questões de

sustentabilidade e saudabilidade. Já a variável *Certificado_wagyu_e_selo_sustentabilidade* apresenta coeficientes próximos a 53 em ambos os modelos, sendo a variável com maior coeficiente calculado. O resultado era esperado em função de o certificado estar relacionado com a qualidade da carne, porém a utilização de uma variável conjunta entre Certificado Wagyu e Selo Sustentabilidade impede o cálculo separado de cada um dos coeficientes, o que limita as conclusões a respeito.

Salta aos olhos a diferença de magnitude do coeficiente calculado para a variável *peso_unitario* no modelo OLS (-4,97) e no modelo WLS (-20,06), o que pode ser explicado pela estatística VIF do modelo WLS para a variável em questão (11,27), indicando alta colinearidade com outras variáveis do modelo e potencial distorção no cálculo de seu coeficiente.

No geral, os modelos WLS e OLS apresentam resultados concordantes em termos de direção, porém com algumas diferenças importantes relativas à magnitude dos coeficientes. Nota-se o fato de a estatística VIF possuir valores muito mais elevados no WLS para variáveis como *Outro.Corte.bovino* (15,83), *Marca_3* (19,03), *Marca_4* (10,79) e *peso_unitario* (11,27). Como observado na seção 0, a base de dados apresenta diferenças importantes nas frequências observadas para N e W, o que influencia de maneira significativa o cálculo de alguns parâmetros do modelo e, também, na estimativa das estatísticas VIF, que levam em conta os pesos do WLS para estimativa dos modelos auxiliares utilizados em seu cálculo.

As diferenças entre os parâmetros calculados para ambos os modelos podem ser explicadas pela diferente ponderação utilizada para minimização do erro-padrão de cada modelo. Utilizando a proteína vegetal como exemplo, a Figura 17 traz a média do preço dos 15 produtos à base dessa proteína, bem como as respectivas frequências para cada produto. Nota-se que o produto 26, que tem 43,78% de frequência W e apenas 10,43% de frequência N, possui um preço médio de R\$ 15,27 – abaixo da média global para produtos com proteínas vegetal. Não coincidentemente, a média simples dos preços para produtos com proteína vegetal (R\$ 18,71) é maior do que a média ponderada por W (R\$ 16,79). Isso explica os preços hedônicos inferiores estimados para as três variáveis de proteína vegetal.

Figura 17 - Frequências e preços médios – Produtos Proteína Vegetal

num_produto	Frequência N	Frequência W	Preço
5	12,89%	15,61%	16,96
16	12,83%	8,48%	18,85
2	12,28%	2,87%	21,73
18	11,84%	2,50%	21,25
26	10,43%	43,78%	15,27
27	10,35%	14,63%	15,81
32	7,74%	3,09%	22,10
36	4,43%	1,45%	18,35
41	4,26%	1,85%	18,46
39	4,00%	2,11%	14,20
38	3,39%	0,57%	21,09
48	1,72%	0,18%	19,25
40	1,52%	0,33%	23,82
43	1,48%	2,49%	13,35
35	0,84%	0,06%	24,84
Total	100,00%	100,00%	18,71

Fonte: elaborado pelo Autor.

Assumindo-se que a frequência W representa de maneira fidedigna a população de consumidores e atos de compra representados pela base de dados, entende-se que o modelo WLS cumpre o papel de estimar preços hedônicos que melhor representa esse universo. Por outro lado, a ponderação simples dos dados a partir do número de observações do modelo OLS aparenta não representar de maneira adequada tal universo, porém permite outros tipos de achados: por um lado, o preço hedônico da variável *Proteina.Vegetal_19* do WLS indica que os consumidores representados pela base aceitaram pagar R\$11,05/Kg a mais por esses produtos; por outro lado, o preço hedônico do OLS de R\$16,09/Kg indica que, dentro desse universo, alguns segmentos de consumidores estão dispostos a pagar um valor ainda mais elevado por quilograma, o que justificaria o parâmetro mais elevado calculado para o OLS.

5 CONCLUSÕES E RECOMENDAÇÕES

Como observado no Referencial Teórico apresentado, o campo de pesquisa de Preços Hedônicos é bastante rico, tanto em aplicações, quanto em estudos que buscam revisar e aprimorar questões metodológicas acerca do tema. Porém, entende-se que este trabalho avança em áreas pouco exploradas dentro do campo: (1) estimativa de preços hedônicos utilizando bases de dados agrupadas; (2) aplicação ao segmento de alimentos processados vendidos em varejos.

O trabalho aplica os modelos de regressão de Mínimos Quadrados Ordinários (OLS) e Mínimos Quadrados Ponderados (WLS) a uma base de dados em painel com vendas semanais de hambúrgueres embalados comercializados em varejos em São Paulo - SP entre os anos de 2019 e 2021. Com o modelo WLS, busca-se propor uma alternativa para aplicação do modelo de preços hedônicos utilizando bases de dados agrupadas, a partir da consideração da quantidade de unidades vendidas na ponderação do modelo.

A utilização de base de dados com agrupamento de vendas já havia sido anteriormente observada em estudos como os de Huang e Lin (2007) e Ahmad e Anders (2012), porém apenas em Bimbo et al. (2016) foi observada a utilização do modelo WLS como técnica para lidar com base de dados com tal característica. A aplicação do referido método é passível de futuras discussões, para validação e refinamento do modelo, porém entende-se que se trata de uma alternativa promissora e flexível, capaz de avaliar o comportamento dos dados de compras agrupadas de uma forma que modelos como o OLS não permitem. Outros tipos de abordagem, como Modelagem Multinível, também são adequados para o estudo de dados com características semelhantes aos aqui utilizados. A comparação e aprofundamento nas semelhanças e diferenças desses tipos de abordagens são promissores para o campo de pesquisa de aplicação de preços hedônicos.

Apesar das diferenças observadas nos modelos OLS e WLS do presente trabalho, ambos convergem para a validação do constructo proposto, de que o tipo de proteína utilizada e selos de qualidade relativos à sustentabilidade e saudabilidade já são percebidos e valorizados pelo mercado consumidor de hambúrgueres congelados da região de São Paulo - SP. Tais resultados são concordantes com estudos anteriores, como os de Satimanon e Weatherspoon (2010), Szathvary e Trestini (2014), Bonanno (2016), Edenbrand, Smed e Jansen (2018), Jiang et al. (2019), que também apontam para a valorização por parte do consumidor de esses tipos de atributos.

Complementarmente, recomenda-se a aplicação do segundo estágio de Rosen incluindo variáveis exógenas de oferta e demanda ao modelo, de maneira a entender as características socioeconômicas que influenciam o preço hedônico dos atributos pesquisados, bem como eventuais características das empresas produtoras que exercem influência sobre ele.

Por fim, por se tratar de um mercado em desenvolvimento, vê-se com bons olhos a replicação do estudo para períodos mais recentes, visando ao acompanhamento da evolução dos preços hedônicos desses atributos com o amadurecimento do mercado. Também é oportuna a aplicação do método a outros segmentos e regiões do país, para identificação de convergências e divergências com o estudo atual.

REFERÊNCIAS

- Acciani, C., De Boni, A., Bozzo, F., & Roma, R. (2021). Pulses for healthy and sustainable food systems: The effect of origin on market price. *Sustainability (Switzerland)*, *13*(1), 1–15. <https://doi.org/10.3390/su13010185>
- Aertsens, J., Verbeke, W., Mondelaers, K., & Van Huylenbroeck, G. (2009). Personal determinants of organic food consumption: a review. *British Food Journal*, *111*(10), 1140–1167. <https://doi.org/10.1108/00070700910992961>
- Ahmad, W., & Anders, S. (2012). The Value of Brand and Convenience Attributes in Highly Processed Food Products. *Canadian Journal of Agricultural Economics*, *60*(1), 113–133. <https://doi.org/10.1111/j.1744-7976.2011.01234.x>
- Bajari, P., & Benkard, C. L. (2005). Demand Estimation with Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach. *Journal of Political Economy*, *113*(6), 1239–1276. <https://doi.org/10.1086/498586>
- Bajari, P., Fruehwirth, J. C., Kim, K. II, & Timmins, C. (2012). A Rational Expectations Approach to Hedonic Price Regressions with Time-Varying Unobserved Product Attributes: The Price of Pollution. *American Economic Review*, *102*(5), 1898–1926. <https://doi.org/10.1257/aer.102.5.1898>
- Baker, S., Thompson, K. E., Engelken, J., & Huntley, K. (2004). Mapping the values driving organic food choice. *European Journal of Marketing*, *38*(8), 995–1012. <https://doi.org/10.1108/03090560410539131>
- Bartik, T. J. (1987a). The Estimation of Demand Parameters in Hedonic Price Models. *Journal of Political Economy*, *95*(1), 81–88. <https://doi.org/https://www.jstor.org/stable/1831300>
- Bartik, T. J. (1987b). Estimating Hedonic Demand Parameters with Single Market Data: The Problems Caused by Unobserved Tastes. *The Review of Economics and Statistics*, *69*(1), 3. <https://doi.org/10.2307/1937920>
- Bimbo, F., Bonanno, A., Liu, X., & Viscecchia, R. (2016). Hedonic analysis of the price of UHT-treated milk in Italy. *Journal of Dairy Science*, *99*(2), 1095–1102. <https://doi.org/10.3168/jds.2015-10018>
- Bitros, G. C., & Panas, E. E. (1988). Measuring Product Prices Under Conditions of Quality Change: The Case of Passenger Cars in Greece. *The Journal of Industrial Economics*, *37*(2), 167–186. <https://www.jstor.org/stable/2098563>
- Bonanno, A. (2016). A Hedonic Valuation of Health and Nonhealth Attributes in the U.S. Yogurt Market. *Agribusiness*, *32*(3), 299–313. <https://doi.org/10.1002/agr.21448>
- Bonanno, A., Bimbo, F., Costanigro, M., Lansink, A. O., & Viscecchia, R. (2019). Credence attributes and the quest for a higher price – A hedonic stochastic frontier approach. *European Review of Agricultural Economics*, *46*(2), 163–192. <https://doi.org/10.1093/erae/jby024>

- Brown, J. N., & Rosen, H. S. (1982). On the estimation of structural hedonic price models. *The Econometric Society*, 50(3), 765–768. <https://doi.org/10.2307/1912614>
- Costanigro, M., & McCluskey, J. J. (2011). Hedonic Price Analysis in Food Markets. In J. L. Lusk, J. Roosen, & J. F. Shogren (Eds.), *The Oxford Handbook of the Economics of Food Consumption and Policy* (Issue February 2016, p. 20). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199569441.013.0007>
- Court, A. T. (1939). Hedonic price indexes with automotive examples. *The Dynamics of Automobile Demand*, 99–119. http://www.econ.wayne.edu/agoodman/research/pubs/court_hedonic.pdf
- Cropper, M. L., Deck, L. B., & McConnell, K. E. (1988). On the Choice of Functional Form for Hedonic Price Functions. *The Review of Economics and Statistics*, 70(4), 668–675. <https://www.jstor.org/stable/1935831>
- De Levie, R. (1986). When, why, and how to use weighted least squares. *Journal of Chemical Education*, 63(1), 10–15. <https://doi.org/10.1021/ed063p10>
- Del Giudice, T., Stranieri, S., Caracciolo, F., Ricci, E. C., Cembalo, L., Banterle, A., & Cicia, G. (2018). Corporate Social Responsibility certifications influence consumer preferences and seafood market price. *Journal of Cleaner Production*, 178, 526–533. <https://doi.org/10.1016/j.jclepro.2017.12.276>
- Diamond, D. B., & Smith, B. A. (1985). Simultaneity in the market for housing characteristics. *Journal of Urban Economics*, 17(3), 280–292. [https://doi.org/10.1016/0094-1190\(85\)90051-8](https://doi.org/10.1016/0094-1190(85)90051-8)
- Diewert, E. (2003). Hedonic Regressions: A Review of Some Unresolved Issues. *News.Ge.*
- Edenbrandt, A. K., Smed, S., & Jansen, L. (2018). A hedonic analysis of nutrition labels across product types and countries. *European Review of Agricultural Economics*, 45(1), 101–120. <https://doi.org/10.1093/erae/jbx025>
- Ekeland, I., Heckman, J. J., & Nesheim, L. (2002). Identifying Hedonic Models. *American Economic Review*, 92(2), 304–309. <https://doi.org/10.1257/000282802320189447>
- Ekeland, I., Heckman, J. J., & Nesheim, L. (2004). Identification and Estimation of Hedonic Models. *Journal of Political Economy*, 112(S1), S60–S109. <https://doi.org/10.1086/379947>
- Fávero, L. P. L. (2005). *O mercado imobiliário residencial da região metropolitana de São Paulo: uma aplicação de modelos de comercialização hedônica de regressão e correlação canônica* [Universidade de São Paulo]. <http://www.teses.usp.br/teses/disponiveis/12/12139/tde-05122005-151150/>
- Fávero, L. P. L., & Belfiore, P. (2017). *Manual de Análise de Dados* (1st ed.). LTC.
- Fouto, N. M. M. D. (2004). *Determinação de uma função de preços hedônicos para computadores pessoais no Brasil*. Universidade de São Paulo.
- Giombi, K. C., Muth, M. K., & Levin, D. (2018). A comparative analysis of hedonic models

- of nutrition information and health claims on food products: An application to soup products. *Journal of Food Products Marketing*, 24(7), 906–926.
<https://doi.org/10.1080/10454446.2018.1428259>
- Griliches, Z. (1961). Hedonic Price Indexes for Automobiles: An Econometric of Quality Change. *National Bureau of Economic Research Volume, The Price*(class VIII), 173–196. <http://www.nber.org/chapters/c6492>
- Griliches, Z. (1971). *Price Indexes and Quality Change* (Z. Griliches (ed.)). Harvard University Press. <https://doi.org/10.4159/harvard.9780674592582>
- Hoppe, A., Barcellos, M. D. de, Vieira, L. M., & Matos, C. A. de. (2012). Comportamento do consumidor de produtos orgânicos: uma aplicação da teoria do comportamento planejado. *BASE - Revista de Administração e Contabilidade Da Unisinos*, 9(2).
<https://doi.org/10.4013/base.2012.92.06>
- Huang, C. L., & Lin, B.-H. (2007). A Hedonic Analysis of Fresh Tomato Prices among Regional Markets. *Review of Agricultural Economics*, 29(4), 783–800.
<https://doi.org/10.1111/j.1467-9353.2007.00387.x>
- Jiang, Y., Wang, H. H., Jin, S., & Delgado, M. S. (2019). The promising effect of a green food label in the new online market. *Sustainability (Switzerland)*, 11(3), 1–14.
<https://doi.org/10.3390/su11030796>
- Kalaitzandonakes, N., Lusk, J., & Magnier, A. (2018). The price of non-genetically modified (non-GM) food. *Food Policy*, 78(March), 38–50.
<https://doi.org/10.1016/j.foodpol.2018.02.005>
- Lancaster, K. J. (1966). A New Approach to Consumer Theory. *The University of Chicago Press*, 74(2), 132–157. <http://www.jstor.org/stable/182883>
- Martínez-Garmendia, J. (2010). Application of hedonic price modeling to consumer packaged goods using store scanner data. *Journal of Business Research*, 63(7), 690–696.
<https://doi.org/10.1016/j.jbusres.2009.05.002>
- Milon, J. W., Gressel, J., & Mulkey, D. (1984). Hedonic Amenity Valuation and Functional Form Specification. *Land Economics*, 60(4), 378. <https://doi.org/10.2307/3145714>
- Moraes, C. H. de C., Alvarenga, M. dos S., Moraes, J. M. M., & Cyrillo, D. C. (2021). Exploring Psychosocial Determinants of Eating Behavior: Fruit and Vegetable Intake Among Brazilian Adolescents. *Frontiers in Nutrition*, 8(December), 1–11.
<https://doi.org/10.3389/fnut.2021.796894>
- Nassiri, A., Thébaud, O., Holbrook, S. J., Lauer, M., Rassweiler, A., Schmitt, R. J., & Claudet, J. (2021). Hedonic evaluation of coral reef fish prices on a direct sale market. *Marine Policy*, 129(March), 1–12. <https://doi.org/10.1016/j.marpol.2021.104525>
- Ndindeng, S. A., Twine, E. E., Mujawamariya, G., Fiamohe, R., & Futakuchi, K. (2021). Hedonic Pricing of Rice Attributes, Market Sorting, and Gains from Quality Improvement in the Beninese Market. *Agricultural and Resource Economics Review*, 50(1), 170–186. <https://doi.org/10.1017/age.2020.24>

- Nerlove, M. (1995). Hedonic price functions and the measurement of preferences: The case of Swedish wine consumers. *European Economic Review*, 39(9), 1697–1716. [https://doi.org/10.1016/0014-2921\(95\)00013-5](https://doi.org/10.1016/0014-2921(95)00013-5)
- Ohsfeldt, R. L., & Smith, B. A. (1985). Estimating the Demand for Heterogeneous Goods. *The Review of Economics and Statistics*, 67(1), 165. <https://doi.org/10.2307/1928450>
- Palmquist, R. B. (1984). Estimating the Demand for the Characteristics of Housing. *The Review of Economics and Statistics*, 66(3), 11. <https://doi.org/10.2307/1924995>
- Qi, X., & Ploeger, A. (2019). Explaining consumers' intentions towards purchasing green food in Qingdao, China: The amendment and extension of the theory of planned behavior. *Appetite*, 133(November 2018), 414–422. <https://doi.org/10.1016/j.appet.2018.12.004>
- Qi, X., & Ploeger, A. (2021). An integrated framework to explain consumers' purchase intentions toward green food in the Chinese context. *Food Quality and Preference*, 92(March), 104229. <https://doi.org/10.1016/j.foodqual.2021.104229>
- Roselli, L., Carlucci, D., Rover, O. J., & De Gennaro, B. (2018). The Effects of Extrinsic Cues on Olive Oil Price in Brazil. *Journal of International Food and Agribusiness Marketing*, 30(1), 70–87. <https://doi.org/10.1080/08974438.2017.1387883>
- Rosen, S. (1974). Hedonic Prices and Implicit Markets : Product Differentiation in Pure Competition. *Journal of Political Economy*, 82(1), 34–55. <https://www.jstor.org/stable/1830899>
- Ruby, M. B., Heine, S. J., Kamble, S., Cheng, T. K., & Waddar, M. (2013). Compassion and contamination. Cultural differences in vegetarianism. *Appetite*, 71, 340–3348. <https://doi.org/10.1016/j.appet.2013.09.004>
- Satimanon, T., & Weatherspoon, D. D. (2010). Hedonic analysis of sustainable food products. *International Food and Agribusiness Management Review*, 13(4), 57–74. <https://doi.org/10.22004/ag.econ.96336>
- Szathvay, S., & Trestini, S. (2014). A Hedonic Analysis of Nutrition and Health Claims on Fruit Beverage Products. *Journal of Agricultural Economics*, 65(2), 505–517. <https://doi.org/10.1111/1477-9552.12056>
- Waldrop, M. E., McCluskey, J. J., & Mittelhammer, R. C. (2017). Products with multiple certifications: insights from the US wine market. *European Review of Agricultural Economics*, 44(4), 658–682. <https://doi.org/10.1093/erae/jbx010>
- Waugh, F. V. (1928). Quality Factors Influencing Vegetable Prices. *Journal of Farm Economics*, 10(2), 185–196. <https://www.jstor.org/stable/1230278>

ANEXOS

A. Código de programação R

```

rm(list = ls())
setwd('C:/Users/Rafael/Documents/Acadêmico/Mestrado/Dissertação/bases')

#####
#      INSTALAÇÃO E CARREGAMENTO DE PACOTES NECESSÁRIOS      #
#####
#Pacotes utilizados
pacotes <-
c(
  "plotly",
  "tidyverse",
  "reshape2",
  "knitr",
  "kableExtra",
  "rgl",
  "car",
  "nlme",
  "lmtest",
  "fastDummies",
  "msm",
  "lmeInfo",
  "jtools",
  "psych",
  "party",
  "ggpubr",
  "dplyr",
  "data.table",
  "MASS",
  "tidyverse",
  "plm",
  "stargazer",
  "magrittr",
  "haven",
  "readxl",
  "nortest",
  "MASS",
  "car",
  "rpart",
  "rpart.plot",
  "Hmisc",
  "gridExtra",
  "grid",
  "PerformanceAnalytics",
  "ggplot",
  "vtable"
)

if (sum(as.numeric(!pacotes %in% installed.packages())) != 0) {
  instalador <- pacotes[!pacotes %in% installed.packages()]
}

```

```

for (i in 1:length(instalador)) {
  install.packages(instalador, dependencies = T)
  break()
}
sapply(pacotes, require, character = T)
} else {
  sapply(pacotes, require, character = T)
}

flattenCorrMatrix <- function(cormat, pmat) {
  ut <- upper.tri(cormat)
  data.frame(
    row = rownames(cormat)[row(cormat)[ut]],
    column = rownames(cormat)[col(cormat)[ut]],
    cor = (cormat)[ut],
    p = pmat[ut]
  )
}

#####
#           FUNÇÕES AUXILIARES PARA A MODELAGEM           #
#####
# Function to add histograms
panel.hist <- function(x, ...) {
  usr <- par("usr")
  on.exit(par(usr))
  par(usr = c(usr[1:2], 0, 1.5))
  his <- hist(x, plot = FALSE)
  breaks <- his$breaks
  nB <- length(breaks)
  y <- his$counts
  y <- y/max(y)
  rect(breaks[-nB], 0, breaks[-1], y, col = rgb(0, 1, 1, alpha = 0.5), ...)
  # lines(density(x), col = 2, lwd = 2) # Uncomment to add density lines
}

# Shapiro-Francia normality test até 30.000 pontos
sf.test2 <- function (x)
{
  DNAME <- deparse(substitute(x))
  x <- sort(x[complete.cases(x)])
  n <- length(x)
  if ((n < 5 || n > 30000))
    stop("sample size must be between 5 and 5000")
  y <- qnorm(ppoints(n, a = 3/8))
  W <- cor(x, y)^2
  u <- log(n)
  v <- log(u)
  mu <- -1.2725 + 1.0521 * (v - u)
  sig <- 1.0308 - 0.26758 * (v + 2/u)
  z <- (log(1 - W) - mu)/sig
  pval <- pnorm(z, lower.tail = FALSE)
  RVAL <- list(statistic = c(W = W), p.value = pval, method = "Shapiro-Francia normality test",
    data.name = DNAME)
  class(RVAL) <- "htest"
}

```

```

return(RVAL)
}

#####
#          IMPORTAÇÃO E MANIPULAÇÃO DOS DADOS          #
#####
base <- read.csv2('scan_sp.csv')
names(base) <- make.names(names(base))
produtos <- read_excel("atributos_produtos.xlsx")
ipc_alimentos <- read_excel("ipc_alimentos_mensal.xlsx")
ipc_alimentos$Valor <- NULL
names(produtos) <- make.names(names(produtos))

produtos <- produtos[is.na(produtos$Ingrediente) == F,]
produtos$peso_unitario <- produtos$Peso.total / produtos$unidades
atributos <-
c(
  'Energia_kcal',
  'Carboidratos',
  'Proteinas',
  'Gorduras',
  'Gorduras_saturadas',
  'Gorduras_trans',
  'Fibra_alimentar',
  'Sodio'
)
produtos[, atributos] <-
produtos[, atributos] / produtos$porção * 100
atributos <-
c(
  atributos,
  'Ingrediente2',
  'Peso.total',
  'Certificado_vegano',
  'Certificado_wagyu_e_selo_sustentabilidade',
  'Certificado_angus',
  'Sem_antibioticos_e_transgenicos',
  'Certificado_organico',
  'peso_unitario',
  'MARCA2'
)
produtos_f <- produtos[, c('LONG', 'EAN', atributos)]
produtos_f[is.na(produtos_f)] <- 0

#Filtros
base_ii <- merge.data.frame(base, produtos_f, by = 'LONG')
base_ii <- base_ii[base_ii$VISAO == 'SPLIT',]
base_ii <-
base_ii[base_ii$CANAL %in% c('AS CENTRO',
  'AS ZONA NORTE',
  'AS ZONA SUL',
  'AS ABCD',
  'AS ZONA LESTE',
  'AS ZONA OESTE'),]

```

```

categoricas <- c('Ingrediente2', 'MARCA2')

#Ajustes campos
base_ii$SEMANA <- as.Date(base_ii$SEMANA)
base_ii$VENDAS.VALOR..in.000. <-
  as.numeric(base_ii$VENDAS.VALOR..in.000.)
base_ii$VENDAS..in.000.KG. <- as.numeric(base_ii$VENDAS..in.000.KG.)
base_ii$VENDAS.UNID.NAO.CONV..in.000. <-
  as.numeric(base_ii$VENDAS.UNID.NAO.CONV..in.000.)
base_ii$VENDAS.UNID.NAO.CONV..tpr.only...in.000. <-
  as.numeric(base_ii$VENDAS.UNID.NAO.CONV..tpr.only...in.000.)
base_ii$VENDAS.UNID.NAO.CONV..tpr.only...in.000.[is.na(base_ii$VENDAS.UNID.NAO.CONV.
.tpr.only...in.000.)] <- 0
base_ii$VENDAS.VALOR..tpr.only...in.000. <-
  as.numeric(base_ii$VENDAS.VALOR..tpr.only...in.000.)
base_ii$VENDAS.VALOR..tpr.only...in.000.[is.na(base_ii$VENDAS.VALOR..tpr.only...in.000.)] <-
  0

#Construindo variável de preço, desconsiderando quantidade e valor vendidoscom desconto
base_ii$preco <-
  (base_ii$VENDAS.VALOR..in.000. - base_ii$VENDAS.VALOR..tpr.only...in.000.) /
  (base_ii$VENDAS.UNID.NAO.CONV..in.000.-
  base_ii$VENDAS.UNID.NAO.CONV..tpr.only...in.000.)

base_ii$t <- as.numeric(difftime(base_ii$SEMANA,min(base_ii$SEMANA), units = "weeks") + 1)

base_ii$ano <- as.numeric(format(base_ii$SEMANA-4, format = "% Y"))
base_ii$mes <- as.numeric(format(base_ii$SEMANA-4, format = "% m"))

base_ii <- base_ii[!is.na(base_ii$preco),]
base_ii <- base_ii[base_ii$preco > 0,]
base_ii <- base_ii[base_ii$preco != Inf,]
colSums(is.na(base_ii))

#Loop de criação das variáveis binárias
for (i in 1:length(categoricas))
{
  coluna_i <- unique(base_ii[, categoricas[i]])
  for (j in 1:(length(coluna_i)))
  {
    categoria_ij <- coluna_i[j]
    base_ii$coluna_j <-
      ifelse(base_ii[, categoricas[i]] == categoria_ij, base_ii$Peso.total, 0)
    colnames(base_ii)[which(names(base_ii) == "coluna_j")] <-
      categoria_ij
    if (i == 1 & j == 1)
    {
      binarias <- c(categoria_ij)
    } else
    {
      binarias <- c(binarias, categoria_ij)
    }
  }
}
}

```



```

base_ii <- base_ii[!duplicated(as.list(base_ii))]

#Ajustando nomes das colunas para aplicação dos modelos
names(base_ii) <- make.names(names(base_ii))
binarias <- make.names(binarias)

base_ii$id <- paste(base_ii$PROD_TAG, base_ii$MKT_TAG, sep = '|')
base_ii <- base_ii[base_ii$ano != 2022, ]

binarias <- c(binarias)
numericas <- c(atributos[!atributos %in% c('Ingrediente2', 'MARCA2')])
preditoras <- c(binarias, numericas)

base_ii <- merge.data.frame(base_ii, ipc_alimentos, by = c('ano', 'mes'), all.x=T)
base_ii$preco <- base_ii$preco/base_ii$Acum
resposta <- c('preco')

base_ii$unidades <- (base_ii$VENDAS.UNID.NAO.CONV..in.000. -
base_ii$VENDAS.UNID.NAO.CONV..tpr.only...in.000.)*1000
base_ii <- base_ii[base_ii$unidades>=1,]

base_ii$Proteina.Vegetal_19 <- 0
base_ii$Proteina.Vegetal_20 <- 0
base_ii$Proteina.Vegetal_21 <- 0

base_ii$Proteina.Vegetal_19[base_ii$ano == 2019] <- base_ii$Proteina.Vegetal[base_ii$ano == 2019]
base_ii$Proteina.Vegetal_20[base_ii$ano == 2020] <- base_ii$Proteina.Vegetal[base_ii$ano == 2020]
base_ii$Proteina.Vegetal_21[base_ii$ano == 2021] <- base_ii$Proteina.Vegetal[base_ii$ano == 2021]

#####
# ANÁLISE DE COMPONENTE PRINCIPAL #
#####

preditoras <-
preditoras[!preditoras %in% c('Outras.Marcas', 'Carne.bovina', 'Peso.total', 'Proteina.Vegetal')]
base_fim <- base_ii[, c(resposta, 'Proteina.Vegetal_19', 'Proteina.Vegetal_20', 'Proteina.Vegetal_21',
preditoras, 'unidades')]

#Estatística descritiva nutrientes
st(base_ii[, c(
'Energia_kcal',
'Carboidratos',
'Proteinas',
'Gorduras',
'Gorduras_saturadas',
'Gorduras_trans',
'Sodio',
'Fibra_alimentar'
)])

#var_pca <- preditoras[!preditoras %in% c('AS.ABCD', 'AS.ZONA.SUL', 'AS.ZONA.NORTE',
'AS.CENTRO', 'AS.ZONA.OESTE')]
var_pca <-

```

```

c(
  'Energia_kcal',
  'Carboidratos',
  'Proteinas',
  'Gorduras',
  'Gorduras_saturadas',
  'Fibra_alimentar'
)
min_MSAi <- 0
var_excluida_f <- NULL
while (min_MSAi < 0.6)
{
  base_preditoras <- scale(base_ii[, c(var_pca)])
  analise_KMO <- KMO(base_preditoras)
  min_MSAi <- min(analise_KMO$MSAi)

  if (min_MSAi < 0.6)
  {
    var_excluida <- names(analise_KMO$MSAi)[analise_KMO$MSAi == min_MSAi]
    var_excluida_f <- c(var_excluida_f, var_excluida)
    var_pca <- var_pca[var_pca != var_excluida]
  }
}

cortest.bartlett(base_preditoras)

preditoras.pca <-
  prcomp(base_preditoras, center = TRUE, scale. = TRUE)
summary(preditoras.pca)
fatores.pca <- preditoras.pca$x

#####
#                MODELAGENS OLS E WLS                #
#####

base_modelo <- base_fim[!names(base_fim) %in% var_pca]
base_modelo <- cbind(base_modelo, fatores.pca[,c('PC1')])
names(base_modelo)[names(base_modelo)=='fatores.pca[, c(\"PC1\")'] <- 'fator_nutrientes_1'

zeradas <- colSums(base_modelo)
zeradas <- zeradas[zeradas == 0]
base_modelo <- base_modelo[, !names(base_modelo) %in% names(zeradas)]

res2 <- rcorr(as.matrix(base_modelo))

diag(res2$r) = NA
correlacoes <- flattenCorrMatrix(res2$r, res2$p)

#Histograma variável dependente preço
hist(base_modelo$preco, breaks = 100)

#cálculo do lambda Box-Cox da variável dependente
lambda_BC <- powerTransform(base_modelo$preco)
lambda_BC

```

```

#Modelo OLS
lm1 <-
  lm(formula(paste(resposta, '~ .-unidades', sep = "")), data = base_modelo)
summary(lm1)
step_lm1 <- step(lm1, k = 3.841459)
summary(step_lm1)
vif_lm1 <- as.data.table(cbind(round(vif(step_lm1),2), names(vif(step_lm1)) ))
names(vif_lm1) <- c('VIF','Variável')
vif_lm1 <- vif_lm1[,c('Variável', 'VIF')]
grid.table(vif_lm1)
coef_lm1 <- as.data.table(cbind(round(step_lm1$coefficients,2), names(step_lm1$coefficients) ))
names(coef_lm1) <- c('coef','Variável')
coef_lm1 <- coef_lm1[,c('Variável', 'coef')]

sf.test2(step_lm1$residuals)
hist(step_lm1$residuals, breaks = 100)

#Diagnóstico de heterocedasticidade: teste de Breuch-Pagan/Cook-Weisberg
bptest(step_lm1)
coeftest(step_lm1, vcov = vcovHC(step_lm1, type = "HC0"))

#Modelo WLS
lm2 <-
  lm(formula(paste(resposta, '~ .-unidades', sep = "")), data = base_modelo, weights = unidades)
summary(lm2)
vif_lm2 <- as.data.table(cbind(round(vif(lm2),2), names(vif(lm2)) ))
names(vif_lm2) <- c('VIF','Variável')
vif_lm2 <- vif_lm2[,c('Variável', 'VIF')]
grid.table(vif_lm2)
coef_lm2 <- as.data.table(cbind(round(lm2$coefficients,2), names(lm2$coefficients) ))
names(coef_lm2) <- c('coef','Variável')
coef_lm2 <- coef_lm2[,c('Variável', 'coef')]

sf.test2(lm2$residuals)
hist(lm2$residuals, breaks = 100)

#Diagnóstico de heterocedasticidade: teste de Breuch-Pagan/Cook-Weisberg
bptest(lm2)
coeftest(lm2, vcov = vcovHC(lm1, type = "HC0"))

#####
#                EXPORTAÇÃO DOS DADOS                #
#####

base_out <- cbind(base_ii, exp(lm1$fitted), lm1$residuals)
write.csv2(base_out, 'output_analise.csv')
write.csv2(base_ii, 'base_ajustada.csv')

```

B. Exemplos de tipos de proteína verificados



C. Exemplos de certificados de produtos verificados



D. Exemplos de tabelas nutricionais verificadas

INFORMAÇÃO NUTRICIONAL		
Porção de 80g (1/2 hambúrguer)		
Quantidade por porção		%VD*
Valor energético	155 kcal = 651 kJ	8
Carboidratos	6,3 g	2
Proteínas	11 g	15
Gorduras totais	7,5 g	13
Gorduras saturadas	2,7 g	12
Gorduras trans	0 g	**
Fibra alimentar	4,6 g	18
Sódio	178 mg	7

* % Valores Diários de referência com base em uma dieta de 2000 kcal ou 8400 kJ. Seus valores diários podem ser maiores ou menores dependendo de suas necessidades energéticas.
** Valor Diário não estabelecido.

INFORMAÇÃO NUTRICIONAL		
Porção de 80g (1/2 hambúrguer)		
Quantidade por porção		%VD*
Valor energético	172 kcal = 722 kJ	9
Carboidratos	9,2g	3
Proteínas	11 g	15
Gorduras totais	7,5 g	14
Gorduras saturadas	2,7 g	12
Gorduras trans	0 g	**
Fibra alimentar	4,6 g	18
Sódio	125 mg	5

* % Valores Diários de referência com base em uma dieta de 2000 kcal ou 8400 kJ. Seus valores diários podem ser maiores ou menores dependendo de suas necessidades energéticas.
** Valor Diário não estabelecido.

INGREDIENTES: CARNE BOVINA (CORTES BOVINOS E 10% DE COSTELA), NÃO CONTÉM GLÚTEN, NÃO CONTÉM CONSERVANTES.

INFORMAÇÃO NUTRICIONAL		
Porção de 80g (1/2 unidade)		
Quantidade por porção		%VD*
Valor energético	214kcal - 903kJ	11%
Carboidratos	0g	0%
Proteínas	15g	20%
Gorduras totais	17g	31%
Gorduras saturadas	7,4g	34%
Gorduras trans	0g	**
Fibra alimentar	0g	0%
Sódio	94mg	4%

*% Valores Diários com base em uma dieta de 2.000kcal ou 8.400kJ. Seus valores diários podem ser maiores ou menores dependendo de suas necessidades energéticas.

INFORMAÇÃO NUTRICIONAL		
PORÇÃO DE 80 g (1 E 1/2 UNIDADE)		
QUANTIDADE POR PORÇÃO		% VD (*)
VALOR ENERGÉTICO	176 kcal = 739 kJ	9%
CARBOIDRATOS	1,8 g	1%
PROTEÍNAS	13 g	17%
GORDURAS TOTAIS	13 g	24%
GORDURAS SATURADAS	6,1 g	28%
GORDURAS TRANS	0 g	**
FIBRA ALIMENTAR	0 g	0%
SÓDIO	570 mg	24%

INFORMAÇÃO NUTRICIONAL		
Porção de 80g (1/2 unidade)		
Quantidade por porção		VD*
Valor energético	232 kcal - 968 kJ	12%
Carboidratos	0g	0%
Proteínas	15g	20%
Gorduras totais	18g	34%
Gorduras saturadas	7,9g	36%
Gorduras trans	0g	**
Fibra alimentar	0g	0%
Sódio	284 mg	12%

*Valores Diários com base em uma dieta de 2.000kcal ou 8.400kJ. Seus valores diários podem ser maiores ou menores dependendo de suas necessidades.

The New Meat Burger - Porção de 80g		
QUANTIDADE POR PORÇÃO		%VD*
Valor Energético	178 kcal = 748 kJ	9
Carboidratos	11 g	4
Proteínas	11 g	15
Gorduras Totais	10 g	18
Gorduras Saturadas	4,4 g	20
Gorduras trans	0 g	**
Fibra Alimentar	4,4 g	18
Sódio	320 mg	13
Vitamina B1 (Tiamina)	1,2 mg	100
Vitamina B2 (Riboflavina)	1,3 mg	100
Vitamina B3 (Niacina)	16 mg	100
Vitamina B6 (Piridoxina)	1,3 mg	100
Vitamina B9 (Ácido Fólico)	240 µg	100
Vitamina B12 (Cobalamina)	2,4 µg	100
Vitamina B7 (Biotina)	30 µg	100
Ferro	14 mg	100

* % Valores Diários com base em uma dieta de 2.000 kcal ou 8400 kJ. Seus valores diários podem ser maiores ou menores dependendo de suas necessidades energéticas. ** Valor diário não estabelecido.

INFORMAÇÃO NUTRICIONAL		
PORÇÃO DE 80 g (1 UNIDADE)		
QUANTIDADE POR PORÇÃO		% VD (*)
Valor Energético	181 kcal = 760 kJ	9%
Carboidratos	4,0 g	1%
Proteínas	12 g	16%
Gorduras Totais	13 g	24%
Gorduras Saturadas	4,6 g	21%
Gorduras Poli-insaturadas	4,8 g	**
Gorduras Monoinsaturadas	3,3 g	**
Gorduras Trans	0 g	**
Colesterol	0 mg	**
Fibra Alimentar	4,3 g	17%
Ferro	2,7 mg	19%
Vitamina B12	0,72 mcg	30%
Sódio	376 mg	16%

(*) % Valores Diários de referência com base em uma dieta de 2.000 kcal ou 8.400 kJ. Seus valores diários podem ser maiores ou menores, dependendo de suas necessidades energéticas. ** VD não estabelecido.

INFORMAÇÃO NUTRICIONAL		
Porção de 134g (2 unidades)		
QUANTIDADE POR PORÇÃO		%VD (*)
VALOR ENERGÉTICO	237kcal/992kJ	12%
CARBOIDRATOS	29g	10%
PROTEÍNAS	16g	21%
GORDURAS TOTAIS	6,9g	13%
GORDURAS SATURADAS	3,1g	14%
GORDURAS TRANS	0g	**
FIBRA ALIMENTAR	1,9g	8%
SÓDIO	604mg	25%

(*) % VALORES DIÁRIOS DE REFERÊNCIA COM BASE EM UMA DIETA DE 2.000 kcal OU 8.400 kJ. SEUS VALORES DIÁRIOS PODEREM SER MAIORES OU MENORES DEPENDENDO DE SUAS NECESSIDADES ENERGÉTICAS. (**) VD NÃO ESTABELECIDO.

INFORMAÇÃO NUTRICIONAL		
QUANTIDADE POR PORÇÃO		
Porção de 80 g (1/2 unidade)		%VD (*)
Valor energético	211 kcal = 882 kJ	10
Carboidratos	0,2 g	0
Proteínas	19,8 g	25
Gorduras totais	16,5 g	29
Gorduras saturadas	7,1 g	32
Gordura trans	0 g	**
Fibra alimentar	0,6 g	1
Sódio	56 mg	2

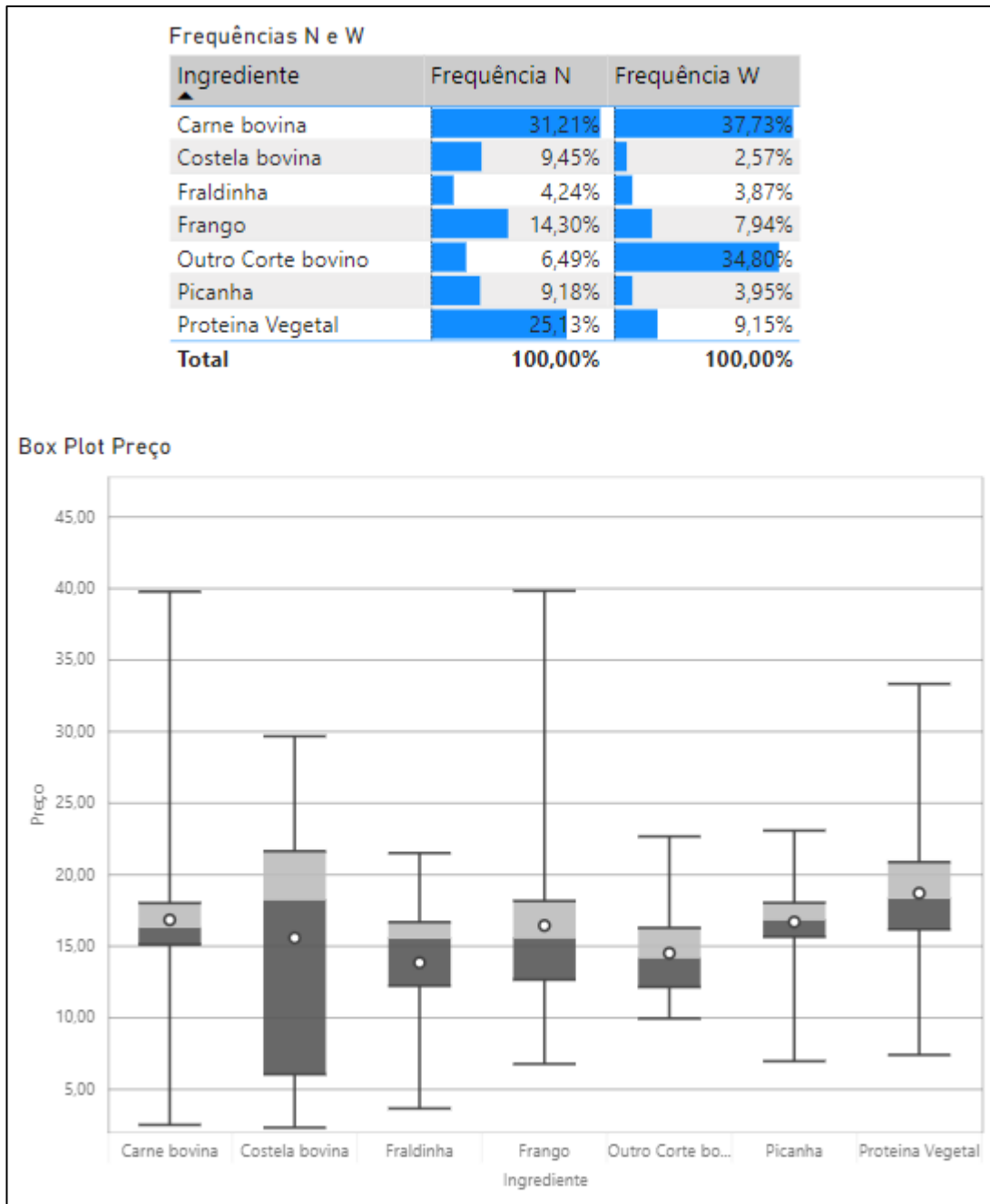
(*) Valores Diários com base em uma dieta de 2.000 kcal ou 8.400 kJ. Seus valores diários podem ser maiores ou menores, dependendo de suas necessidades energéticas.
(**) VD não estabelecido.

E. IPCA de alimentos do período de estudo

ano	mes	Valor	Acum
2019	1	1,13%	1,000
2019	2	1,64%	1,016
2019	3	1,75%	1,034
2019	4	0,23%	1,037
2019	5	-1,05%	1,026
2019	6	-0,51%	1,020
2019	7	-0,11%	1,019
2019	8	0,17%	1,021
2019	9	-0,99%	1,011
2019	10	-0,09%	1,010
2019	11	1,62%	1,026
2019	12	2,96%	1,057
2020	1	0,60%	1,063
2020	2	0,04%	1,064
2020	3	0,93%	1,073
2020	4	1,26%	1,087
2020	5	0,66%	1,094
2020	6	0,69%	1,102
2020	7	0,63%	1,109
2020	8	1,27%	1,123
2020	9	2,15%	1,147
2020	10	2,51%	1,176
2020	11	2,24%	1,202
2020	12	2,08%	1,227
2021	1	1,50%	1,245
2021	2	-0,31%	1,242
2021	3	0,27%	1,245
2021	4	0,83%	1,255
2021	5	0,04%	1,256
2021	6	0,09%	1,257
2021	7	1,52%	1,276
2021	8	2,29%	1,305
2021	9	1,38%	1,323
2021	10	0,90%	1,335
2021	11	-0,15%	1,333
2021	12	1,23%	1,349

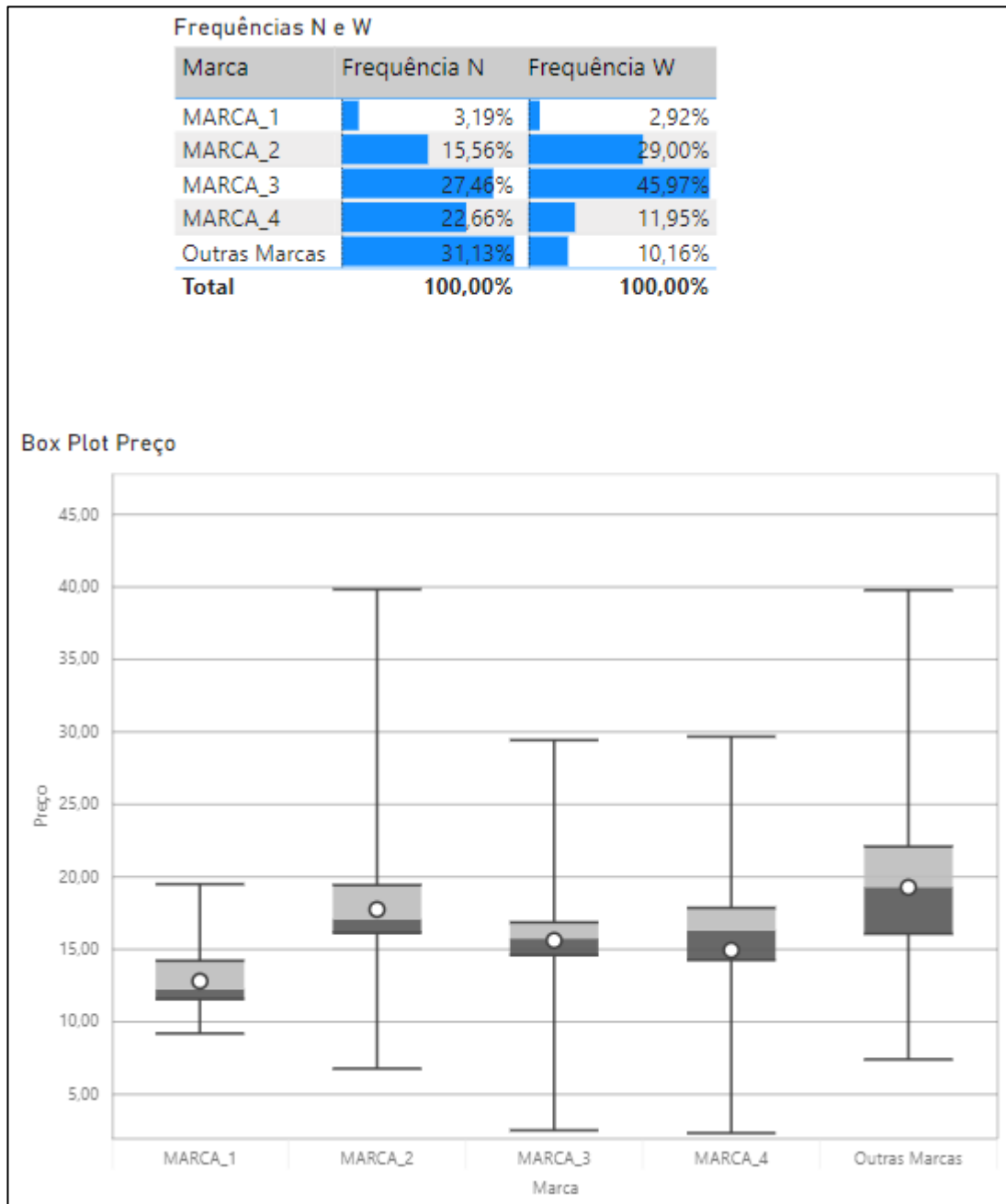
F. Gráficos descritivos da amostra de dados

Figura 18 - Frequências e Box Plot de preço por Ingrediente



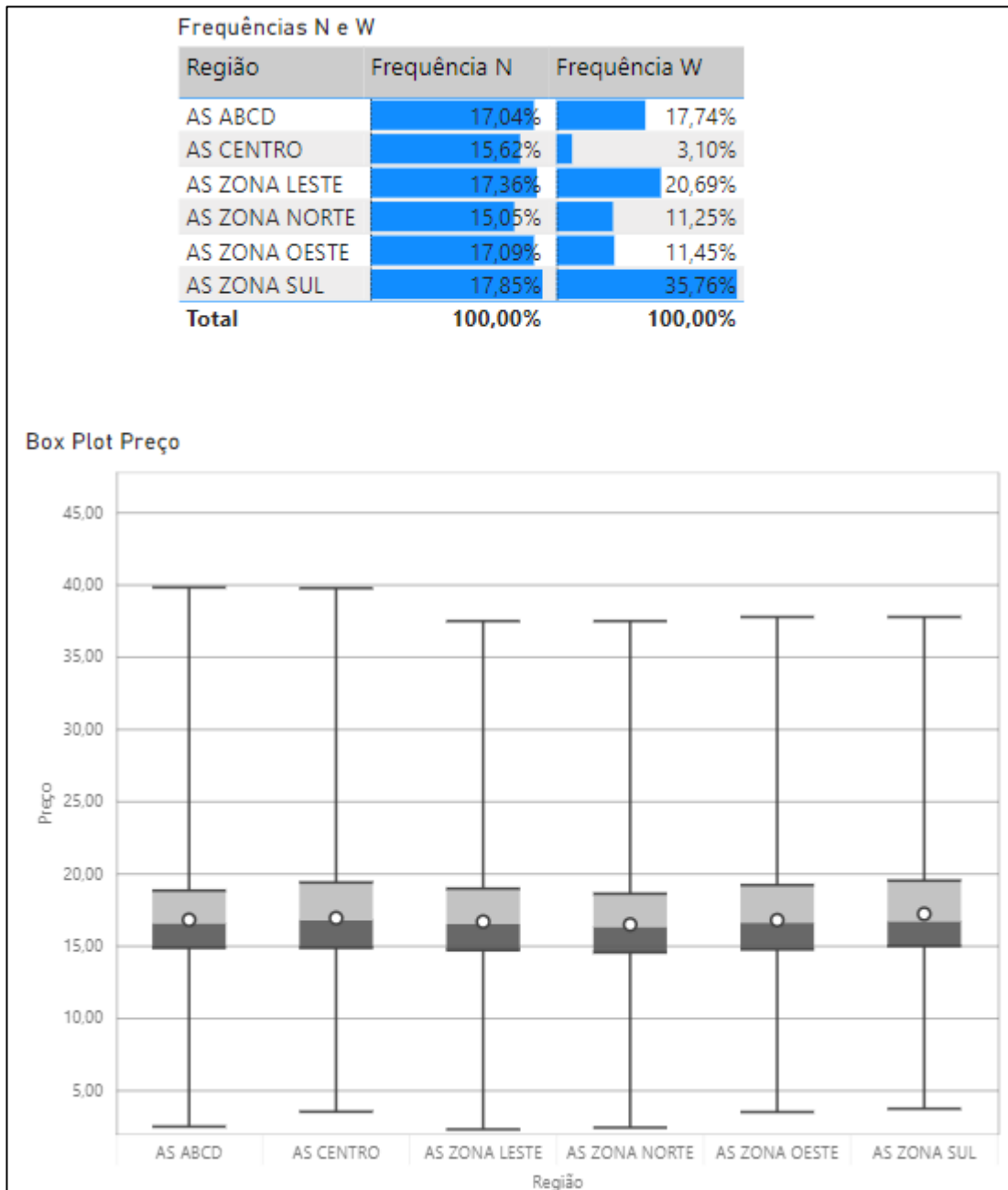
Fonte: elaborado pelo Autor.

Figura 19 - Frequências e Box Plot de preço por Marca



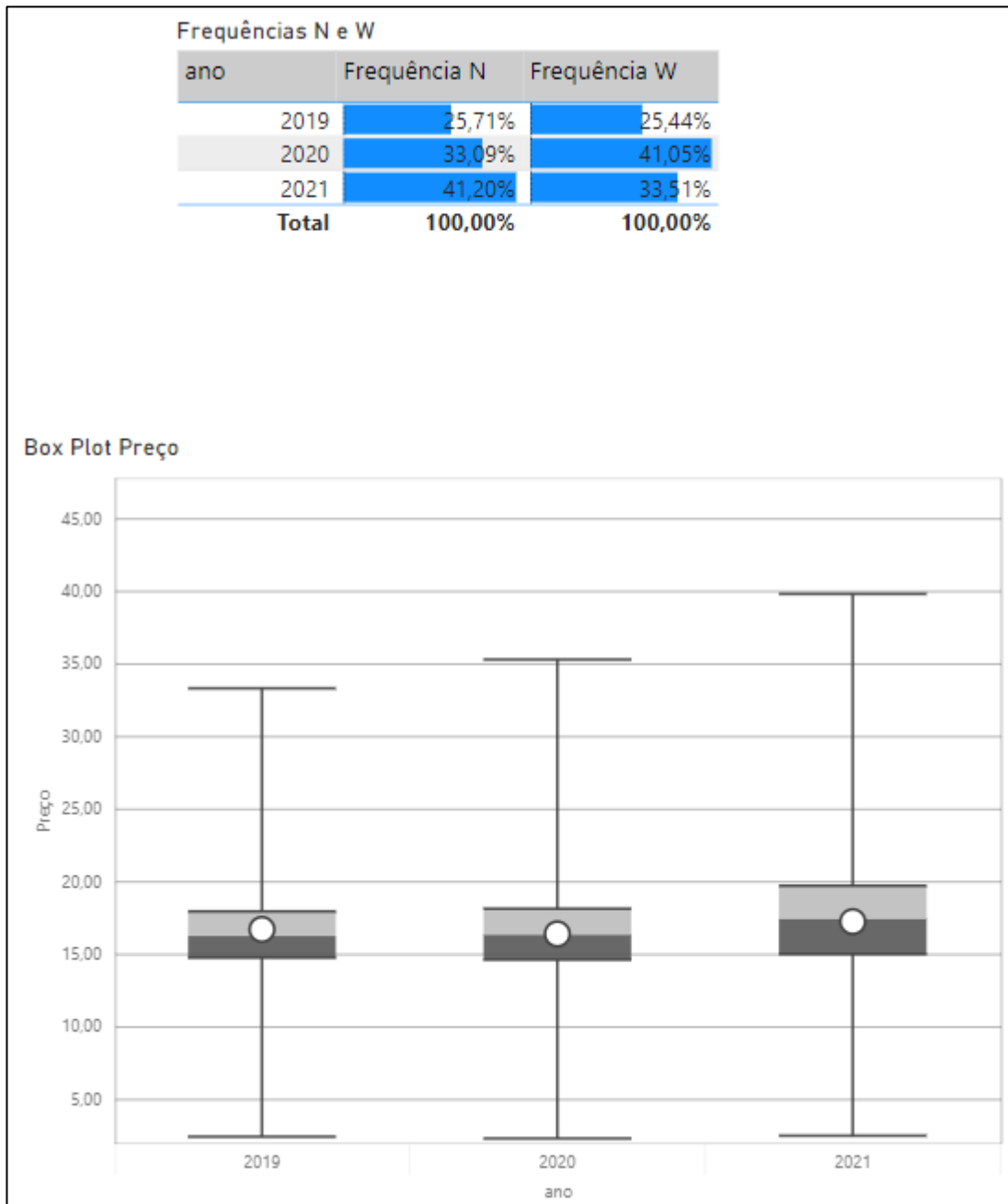
Fonte: elaborado pelo Autor

Figura 20 - Frequências e Box Plot de preço por Região



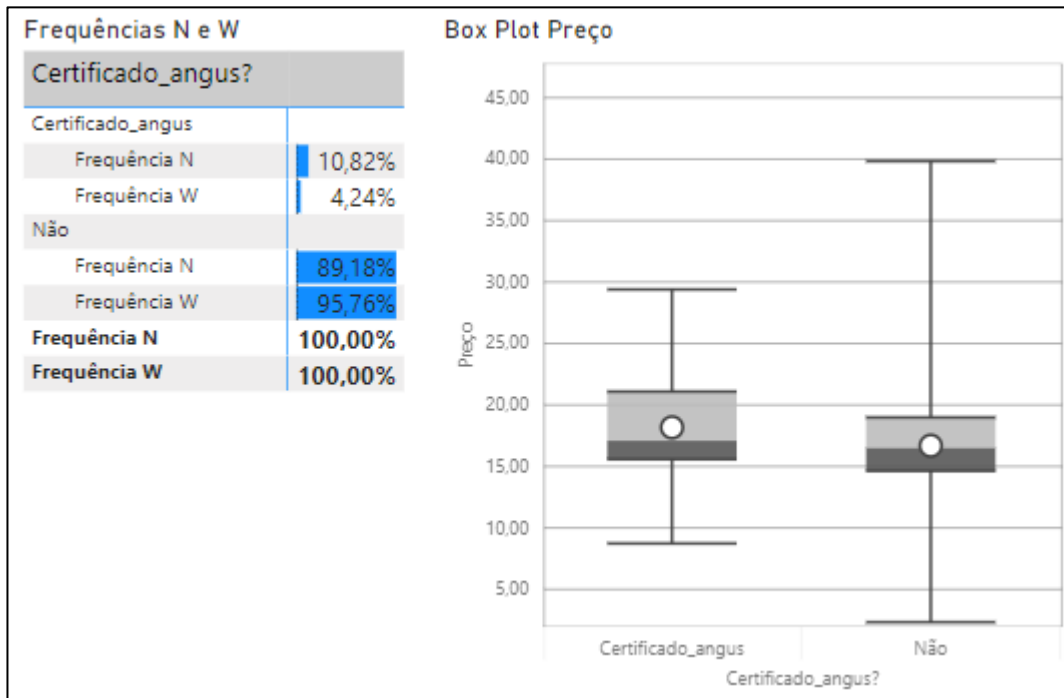
Fonte: elaborado pelo Autor.

Figura 21 - Frequências e Box Plot de preço por ano



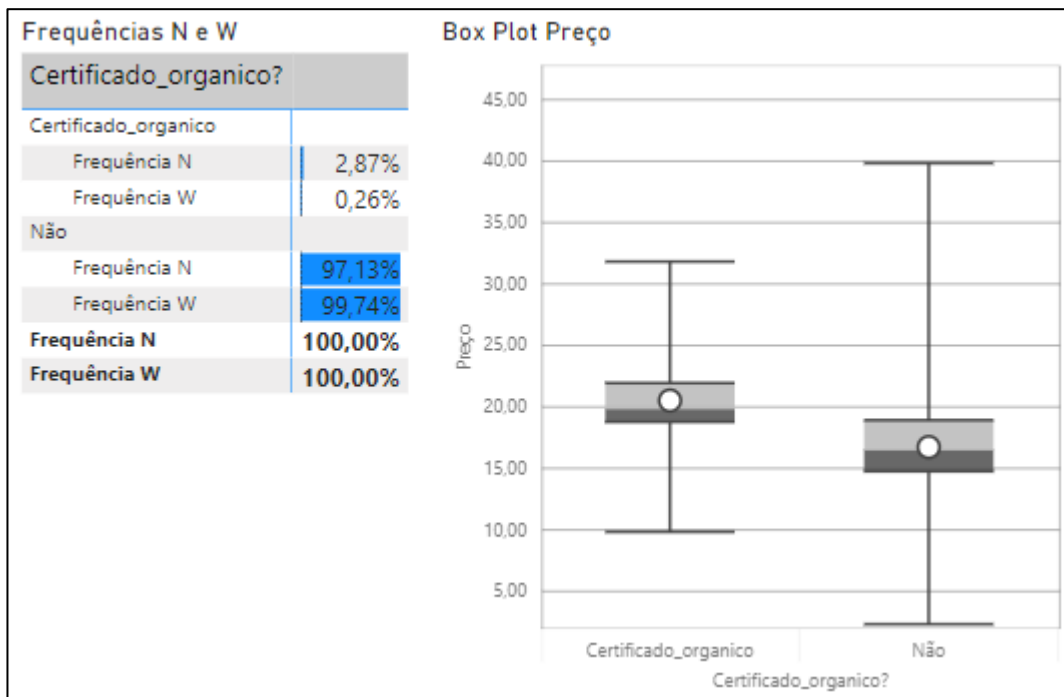
Fonte: elaborado pelo Autor

Figura 22 - Frequências e Box Plot de preço – Certificado Angus



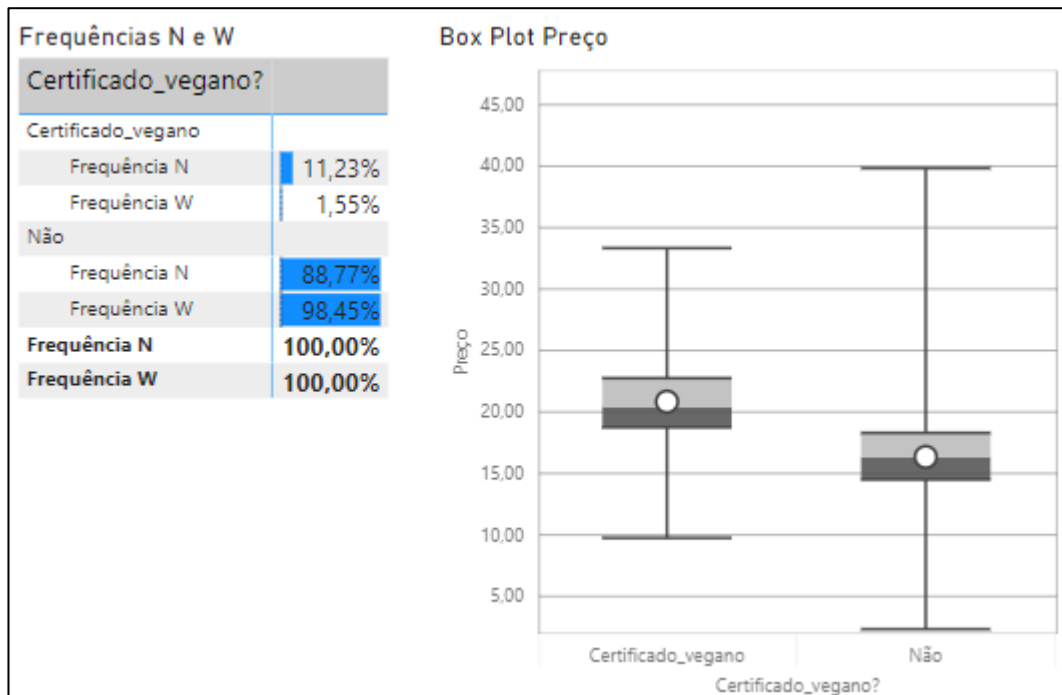
Fonte: elaborado pelo Autor.

Figura 23 - Frequências e Box Plot de preço – Certificado Orgânico



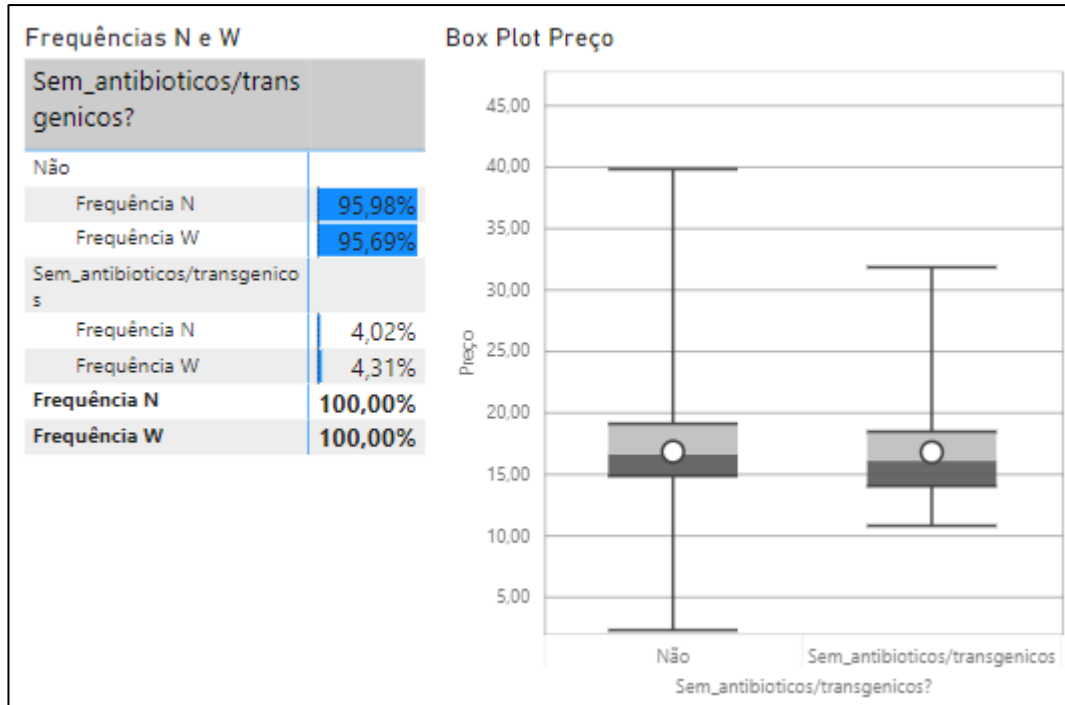
Fonte: elaborado pelo Autor.

Figura 24 - Frequências e Box Plot de preço – Certificado Vegano



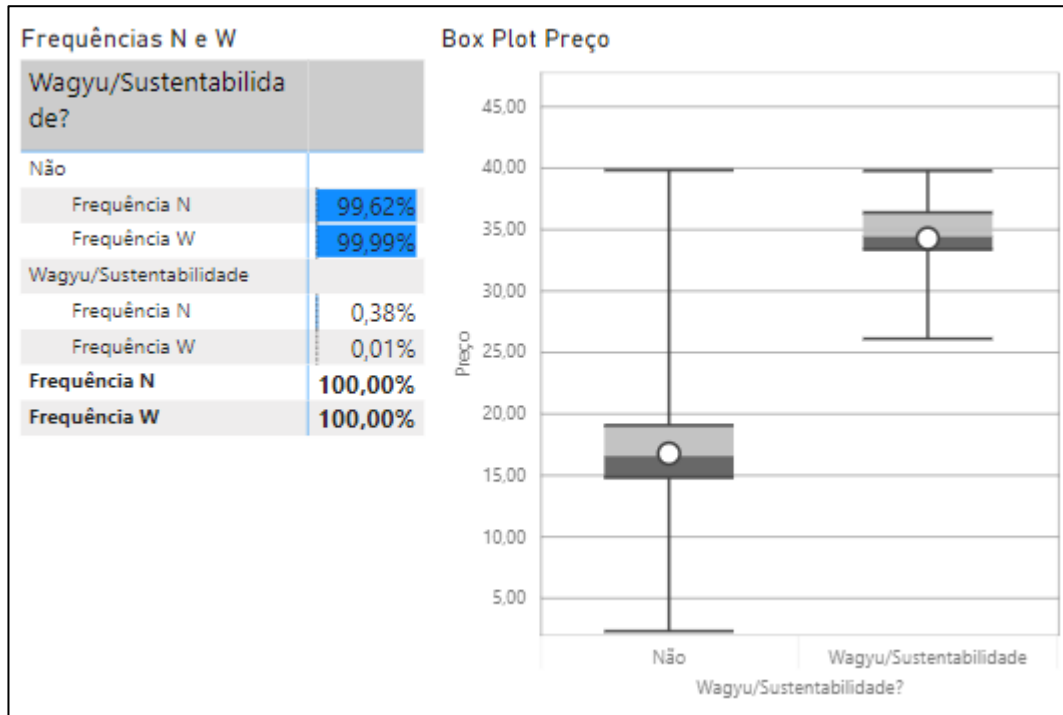
Fonte: elaborado pelo Autor.

Figura 25 - Frequências e Box Plot de preço – Sem Antibióticos / Transgênicos



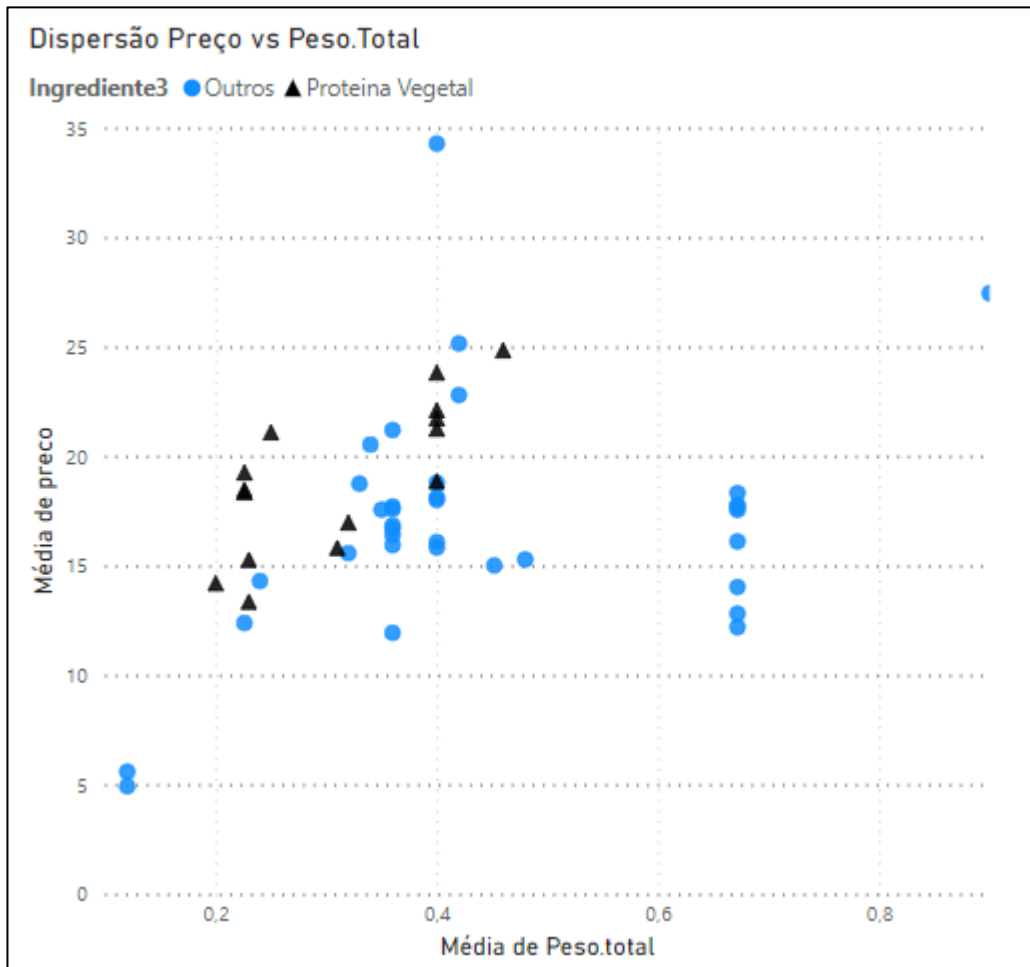
Fonte: elaborado pelo Autor.

Figura 26 - Frequências e Box Plot de preço – Certificado Wagyu / Selo Sustentabilidade



Fonte: elaborado pelo Autor.

Figura 27 - Dispersão Preço vs Peso da Embalagem



Fonte: elaborado pelo Autor.

G. OLS sem seleção de variável

Figura 28 - OLS sem seleção de variável

```

Residuals:
    Min       1Q   Median       3Q      Max
-11.9508  -1.6350  -0.1619   1.6289  17.0870

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    14.64403    0.10790  135.722 < 2e-16 ***
Proteina.Vegetal_19  16.10831    0.36226   44.467 < 2e-16 ***
Proteina.Vegetal_20  14.13611    0.32925   42.934 < 2e-16 ***
Proteina.Vegetal_21  12.75583    0.30533   41.778 < 2e-16 ***
Costela.bovina    17.29284    0.25170   68.705 < 2e-16 ***
Frango           3.43140    0.14161   24.231 < 2e-16 ***
Outro.Corte.bovino -2.29883    0.18493  -12.431 < 2e-16 ***
Picanha          -0.34984    0.31543   -1.109    0.267
Fraldinha        -4.27323    0.36835  -11.601 < 2e-16 ***
MARCA_2           8.38449    0.12320   68.054 < 2e-16 ***
MARCA_1          -1.68700    0.21207   -7.955 1.86e-15 ***
MARCA_3           6.21444    0.16219   38.316 < 2e-16 ***
MARCA_4          14.13502    0.32460   43.546 < 2e-16 ***
Gorduras_trans    0.56856    0.14111    4.029 5.61e-05 ***
Sodio            -5.90687    0.09486  -62.266 < 2e-16 ***
Certificado_vegano  1.99053    0.28276    7.040 1.97e-12 ***
Certificado_wagyu_e_selo_sustentabilidade 52.89115    0.72627   72.826 < 2e-16 ***
Certificado_angus  8.57562    0.19643   43.658 < 2e-16 ***
Sem_antibioticos_e_transgenicos  3.96371    0.37185   10.659 < 2e-16 ***
Certificado_organico 21.89630    0.39384   55.597 < 2e-16 ***
peso_unitario     -5.19056    0.64443   -8.054 8.29e-16 ***
fator_nutrientes_1  0.57140    0.02081   27.464 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.897 on 28821 degrees of freedom
Multiple R-squared:  0.5528,    Adjusted R-squared:  0.5525
F-statistic: 1697 on 21 and 28821 DF,  p-value: < 2.2e-16

```