

**UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO  
CENTRO DE CIÊNCIAS JURÍDICAS E ECONÔMICAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO**

**JULIANO HONORATO CÂNDIDO**

**RECEBER EM CASA OU BUSCAR NO LOCAL? UM ESTUDO DAS  
PREFERÊNCIAS DOS CONSUMIDORES DE E-COMMERCE**

**VITÓRIA**

**2021**

**JULIANO HONORATO CÂNDIDO**

**RECEBER EM CASA OU BUSCAR NO LOCAL? UM ESTUDO DAS  
PREFERÊNCIAS DOS CONSUMIDORES DE E-COMMERCE**

Dissertação apresentada ao Programa de Pós-Graduação em Administração do Centro de Ciências Jurídicas e Econômicas da Universidade do Espírito Santo, como requisito parcial para obtenção do título de Mestre em Administração.

Orientador: Prof. Dr. Adonai José Lacruz

**VITÓRIA**

**2021**

**JULIANO HONORATO CÂNDIDO**

**RECEBER EM CASA OU BUSCAR NO LOCAL: UM ESTUDO DAS  
PREFERÊNCIAS DOS CONSUMIDORES DE E-COMMERCE**

Dissertação apresentada ao Programa de Pós-Graduação em Administração do Centro de Ciências Jurídicas e Econômicas da Universidade do Espírito Santo, como requisito parcial para obtenção do título de Mestre em Administração.

Aprovado em \_\_\_\_\_/\_\_\_\_\_/\_\_\_\_\_

**COMISSÃO EXAMINADORA**

---

Adonai José Lacruz, D. Sc.  
Instituto Federal do Espírito Santo  
PPGAdm – Universidade Federal do  
Espírito Santo  
Orientador

---

Hélio Zanquetto Filho  
Universidade Federal do Espírito Santo  
Coorientador

---

Ricardo Silveira Martins, D. Sc.  
Universidade Federal de Minas Gerais

---

Marcos Paulo V. de Oliveira, D. Sc.  
Universidade Federal do Espírito Santo

*“(...) a essência do meu progresso está em poder aceitar minha decadência. Ou seja, progredir até morrer, porque viver é morrer. E não me arrependo de nada.”*

*Antônio Abujamra*

## AGRADECIMENTOS

O esforço, de certo, possui algum grau de parentesco com o amor. Por isso, não me surpreende a necessidade interna de ser grato a todos aqueles que amo por me permitirem dedicar a eles esse meu esforço.

Agradeço à minha mãe. Meu pilar desde sempre. Desde antes de saber que o que eu sentia era chamado de amor. Desde antes de compreender o que é o esforço. Desde o início dessa vida. E certamente para além do fim da mesma. Devo a ela a vida. Ofereço a ela meu esforço e o meu amor.

Agradeço a minha noiva, minha companheira diária. A essa mulher que tanto me ensina sobre o que é o amor e sobre como ele é feito de paz. Paz e inspiração. Agradeço a ela a paciência de me ver prosperar diariamente. Ofereço a ela meu esforço, meu amor.

Agradeço a meu pai que, no ano em que termino de escrever essa dissertação, deixou pra mim o presente mais belo que podia deixar ao desencarnar. Deixou para mim o amor que não me permiti sentir durante a vida, mas que certamente sentirei durante todo o restante dela. Ofereço a ele meu esforço, meu amor.

Agradeço ao meu orientador e ao meu coorientador, não só pela paciência em me guiar pelo caminho da sabedoria, mas pela insistência em cavar dentro de mim o meu potencial. Acreditem, a cada reunião, cada conversa, cada encontro, cada pancada, eu me senti feliz por fazer parte do esforço de vocês. Por isso, a vocês dois, ofereço meu esforço e o meu amor.

Aos meus irmãos, amigos e familiares, agradeço por compartilharem dessa vida comigo. E faço aqui uma promessa, uma dívida de gratidão: nunca vou parar de lhes oferecer o meu esforço e o meu amor.

## RESUMO

O crescimento dos números do comércio eletrônico varejista mostra que essa modalidade de compra tem grande importância no padrão de consumo global. Entretanto, esse crescimento traz alguns desafios para os gestores nos processos logísticos do *delivery* tradicional, principalmente na última etapa de entrega, quando o produto sai do galpão e chega até a casa do consumidor, conhecido como *last mile* (LM), ou última milha, em português. Problemas como falha na entrega e descumprimento do prazo de entrega, geram insatisfação do consumidor e por isso geram consequências negativas para os varejistas. Nesse sentido, uma das soluções apresentadas pela literatura é a solução do *pick-up point* (PUP), onde o consumidor participa do processo de entrega e busca o produto comprado em um estabelecimento escolhido, ao invés de receber em casa, como no processo de *Home Delivery* (HD). Entretanto, dado o fato de que a solução do PUP é ainda tida como uma solução inovadora, ela gera certo grau de incerteza e ansiedade por parte dos consumidores. Por isso, alguns autores buscam entender quais fatores levam o cliente a optar pelo PUP em detrimento ao HD. Ainda assim, há uma lacuna na literatura sobre a sensibilidade dos consumidores a alterações nos níveis desses fatores. Por isso, o objetivo dessa pesquisa foi o de observar quais são os fatores significantes, a sensibilidade do consumidor a alteração nos níveis desses fatores e qual a ordem de influência que esses fatores exercem sobre a preferência do consumidor pelo PUP em detrimento ao HD. A fim de atingir esse objetivo, foi realizado na pesquisa um desenho experimental, a partir da técnica de Preferência Declarada, para realizar uma *survey* que, após ser aplicada a uma amostra de 405 respondentes, teve seus resultados analisados a partir de um modelo de regressão logística robusta por clusters e pela lente da Teoria dos Custos de Transação. Os resultados mostram que os atributos que influenciam significativamente a escolha do consumidor pelo PUP em detrimento ao HD foram o Preço do Frete e o Prazo de Entrega. Todavia, apesar da literatura apresentar o Preço do Frete como a variável mais influente na maioria dos estudos, os resultados dessa pesquisa apresentam o Prazo de Entrega como o mais influente. Além disso, os resultados mostraram também que o atributo Localização do PUP não foi considerado estatisticamente influente na escolha do consumidor pelo PUP em detrimento ao HD. Por fim, os resultados dessa pesquisa mostram que a variação interna nos níveis do atributo Preço do Frete causam maior impacto na intenção do consumidor em utilizar o PUP do que a variação interna nos níveis do atributo Prazo de Entrega.

**Palavras-chave:** Comércio eletrônico. Last Mile. Preferência Declarada. Pick-up Point. Delivery. Regressão Logística.

## ABSTRACT

The growth of retail e-commerce numbers show that this purchase modal is very important in the global consumption pattern. However, this growth brings some challenges for managers in the logistical processes of traditional delivery, especially at the last delivery process, known as the Last Mile (LM), when the product leaves the warehouse and arrives at the consumers' home. Problems such as deliveries failure and failure to attend the delivery deadline cause consumer dissatisfaction and therefore generate negative consequences for retailers. In this sense, one of the solutions presented by literature is the Pick-up Point (PUP) Solution, where the consumer participates in the delivery process and pick-up the purchased product at the chosen establishment, instead of receiving at home, as in the traditional Home Delivery (HD) process. However, given the fact that the PUP Solution is still regarded as an innovative solution, it generates some degree of uncertainty and anxiety on part of the consumers. Therefore, some authors seek to understand which factors lead the consumers to choose PUP instead of HD. Still there is a gap in the literature about consumers sensitivity to changes at the level of these factors. So the goal of this dissertation was to observe which are the significant factors, the consumers' sensitivity to changes at the level of these factors and the order of influence that these factors exert on the consumer's preference for PUP over HD. In order to achieve this goal, the dissertation performed an design of experiment, using Stated Preference technique, to apply a survey that, after being applied to a sample of 405 respondents, had its results analyzed using a robust logistic regression model by clusters and by Transactional Costs Theory' point of view. The results shows that the attributes that significantly influence consumer choice for PUP over HD were the Shipping Fee and Delivery Time. On the other hand, although the literature presents the Shipping Price as the most influential factor in most studies, the results here discussed show that Delivery Time was the most influential one. In addition, the results also showed that the PUP Location attribute wasn't considered statistically influential in consumer choice for PUP over HD. Finally, the results of this dissertation show that the internal variation in levels of Shipping Price attribute has a greater impact on the consumers' intention to use the PUP than the internal variation in the levels of the Delivery Time attribute.

**Keywords:** E-commerce. Last Mile. Stated preference. Pick-up point. Delivery. Logistic regression.

## LISTA DE FIGURAS

Figura 1 – Vendas do comércio eletrônico no mundo.....	18
Figura 2 – Produção anual de trabalhos.....	19
Figura 3 – Esquematização do LM.....	22
Figura 4 – Procedimentos de Preferência Declarada.....	36
Figura 5 – Fluxograma de processo da survey .....	46
Figura 6 – Caracterização da amostra inovação .....	52
Figura 7 – Caracterização da amostra experiências anteriores.....	53
Figura 8 – Caracterização da amostra por idade.....	54
Figura 9 – Caracterização da amostra – utilização do PUP.....	54
Figura 10 – Caracterização da amostra – tamanho da cidade.....	55
Figura 11 – Curva ROC.....	65

## LISTA DE QUADROS

Quadro 1 – Quadro sinótico de atributos e estudos anteriores .....	37
Quadro 2 – Atributos utilizados no instrumento de Preferência Declarada .....	38
Quadro 3 – Níveis de atributos utilizados .....	40
Quadro 4 – Relação de cenário versus número .....	44
Quadro 5 – Variável dummy Preço .....	48
Quadro 6 – Variável dummy Distância .....	48
Quadro 7 – Variável dummy Prazo .....	49

## LISTA DE TABELAS

Tabela 1- Formas de resposta de uma survey de PD.....	34
Tabela 2 - Estatística descritiva .....	56
Tabela 3 - Distribuição de frequência.....	56
Tabela 4 - Saídas da regressão logística .....	58
Tabela 5 - Regressão logística stepwise forward.....	58
Tabela 6 - Percentual de influência das variáveis dummy .....	60
Tabela 7 - Regressão Logística das Razões de Probabilidade.....	61
Tabela 8 - AIC e BIC modelo final .....	62
Tabela 9 - Teste de Hosmer-Lemeshow .....	63
Tabela 10 – Matriz de classificação .....	64
Tabela 11 – Saídas do erro de especificação .....	66
Tabela 12 – Saídas jackknife .....	67

## **LISTA DE SIGLAS**

AIC – Critério de informação de Akaike

BIC – Critério de informação bayesiano

DCE – Discrete Choice Experiment

HD – Home Delivery

IBGE – Instituto Brasileiro de Geografia e Estatística

LM – Last Mile

PD – Preferência Declarada

PNAD – Pesquisa Nacional por Amostra de Domicílios Contínua

PUP – Pick-up Point

TCT – Teoria dos Custos de Transação

## SUMÁRIO

<b>1. INTRODUÇÃO.....</b>	<b>14</b>
1.1 OBJETIVOS.....	17
<b>1.1.1 Objetivo geral.....</b>	<b>17</b>
<b>1.1.2 Objetivos específicos.....</b>	<b>17</b>
1.2 JUSTIFICATIVA.....	18
<b>1.2.1 Justificativa do tema.....</b>	<b>18</b>
<b>1.2.2 Justificativa do método.....</b>	<b>20</b>
<b>2. REVISÃO DA LITERATURA.....</b>	<b>21</b>
2.1 O LAST MILE E SUAS QUE STÕES NO CONTEXTO DO COMÉRCIO ELETRÔNICO.....	21
2.2 TEORIA DOS CUSTOS DE TRANSAÇÃO (TCT).....	25
<b>2.2.1 Oportunismo e racionalidade limitada.....</b>	<b>26</b>
<b>2.2.2 Categorias dos custos de transação.....</b>	<b>28</b>
2.3 MODELO DE UTILIDADE ALEATÓRIA.....	30
<b>3. PROCEDIMENTO METODOLÓGICO.....</b>	<b>33</b>
3.2 PREFERÊNCIA DECLARADA.....	33
3.3 DESENHO EXPERIMENTAL.....	35
<b>3.2.1 Problema de pesquisa.....</b>	<b>36</b>
<b>3.2.2 Atributos das alternativas.....</b>	<b>37</b>
<b>3.2.3 Níveis dos atributos.....</b>	<b>39</b>
<b>3.2.4 Desenho Estatístico.....</b>	<b>40</b>
<b>3.2.5 Aplicação da survey teste.....</b>	<b>42</b>
<b>3.2.6 Aplicação da survey.....</b>	<b>43</b>
3.3 MODELO DE ESCOLHA DISCRETA E REGRESSÃO LOGÍSTICA.....	47
<b>3.3.1 Modelo de Regressão Logística.....</b>	<b>47</b>

<b>3.3.2 Variáveis independentes e coeficientes logísticos.....</b>	<b>48</b>
<b>3.3.3 Interpretação dos coeficientes de regressão .....</b>	<b>49</b>
<b>3.3.4 Tamanho da amostra .....</b>	<b>50</b>
<b>4. ANÁLISE DOS RESULTADOS .....</b>	<b>51</b>
4.1 CARACTERIZAÇÃO DA AMOSTRA .....	51
4.2 ANÁLISE DESCRITIVA DA SURVEY .....	56
4.3 ANÁLISE DA REGRESSÃO LOGÍSTICA.....	57
<b>5. DISCUSSÃO TEÓRICA DOS RESULTADOS E IMPLICAÇÕES .....</b>	<b>68</b>
5.1 CONTRIBUIÇÕES TEÓRICAS.....	70
5.2 CONTRIBUIÇÕES PRÁTICAS.....	71
<b>6. CONCLUSÃO.....</b>	<b>73</b>
<b>REFERENCIAS .....</b>	<b>75</b>
<b>APÊNDICE A – SITUAÇÃO HIPOTÉTICA E CENÁRIOS DA SURVEY DE PD.....</b>	<b>82</b>
<b>APÊNDICE B – QUESTÕES SOCIODEMOGRÁFICAS.....</b>	<b>87</b>
<b>APÊNDICE C – SAÍDAS EM DFBETA.....</b>	<b>89</b>
<b>APÊNDICE D – SCRIPT REGRESSÃO LOGÍSTICA .....</b>	<b>90</b>
<b>APÊNDICE E – SAÍDAS DA REGRESSÃO LOGÍSTICA.....</b>	<b>91</b>

## 1. INTRODUÇÃO

As compras por meio eletrônico são uma nova alternativa dos clientes adquirirem o que precisam, e que foi viabilizado pela internet e pela digitalização dos negócios. Essa modalidade de compra vem crescendo de forma consistente desde os anos 2000. Em dados de um estudo realizado pela eMarketer (2018), por exemplo, a participação do comércio eletrônico nas vendas no varejo ao redor do mundo era de 7,4% em 2015, com projeção de alcançar 17,5% em 2021. Ao analisar os dados do estudo de 2019, observa-se que a previsão se mantém na casa dos 17% para 2021 e a tendência de estar em crescente se comprova ao se observar que, se em 2015 a participação era de 7,4%, em 2019 se confirmou uma participação de 14,1% (EMARKETER, 2019).

No ambiente brasileiro, dados da Associação Brasileira de Comércio Eletrônico (ABCOMM, 2019) e do Ministério da Ciência, Tecnologia, Inovações e Comunicações (MCTIC, 2018) apontam crescimento acumulado de mais de 570% no volume de vendas online entre 2003 e 2018. Esses dados mostram como o comércio eletrônico já é uma realidade no padrão de consumo global e como o mercado brasileiro acompanha este padrão, sendo o Brasil o maior mercado varejista online da América Latina, de acordo com o International Post Corporation (2017).

Alguns argumentos explicam o crescimento desses números do comércio eletrônico no mundo. Do ponto de vista da organização, quando comparado ao modo de compra em locais físicos, o canal de vendas pela internet permite um maior alcance de consumidores e possibilita ultrapassar seus limites geográficos, físicos e de tempo, além de acelerar a comunicação entre cliente e vendedor e reduzir os custos de transação, de propaganda e promoção, otimizando assim seus recursos (NEJADIRANI *et al.*, 2011; SAVRUL *et al.*, 2014; FRANCKE; VISSER, 2015; VAKULENKO *et al.*, 2019).

Do ponto de vista do consumidor, a compra pela internet apresenta benefícios para quem opta por essa modalidade pois, diferentemente da compra presencial, a compra online aumenta a comodidade e conveniência do cliente, uma vez que não é necessário se deslocar até nenhum local físico para comparar preços, ofertas e produtos, além de possibilitar a compra a qualquer hora do dia, confirmando assim a maior comodidade (CHEN; DUBINSKY, 2003; FENG; HUANG, 2005; HSIAO, 2009; VAKULENKO *et al.*, 2019).

Por fim, o número de pessoas conectadas à internet é também um fator concomitante ao crescimento do comércio eletrônico. Em dados da Pesquisa Nacional por Amostra de

Domicílios Contínua (PNAD), realizada em 2017, 74,9% dos domicílios brasileiros utilizam a internet (PNAD, 2017), e esse grande percentual de pessoas são potenciais compradores online.

Apesar de todas essas vantagens observadas na literatura, quando comparado ao método de compra físico, a compra online oferece alguns desafios de gestão para as organizações. Maes *et al.* (1999) define o processo de compra online como um conjunto de seis estágios realizados em sequência, sendo eles: a Identificação da necessidade, Escolha do produto, Escolha do vendedor, Negociação, Compra e Entrega e avaliação. Cada uma dessas seis etapas apresenta diferentes dificuldades e entraves durante a sua execução, o que afeta diretamente na satisfação do consumidor com relação ao processo de compra (ZEITHAML, 2000; BOYER; HULT, 2005; RAO *et al.*, 2011; VAKULENKO *et al.*, 2019). Lidar com estes problemas que ocorrem durante o processo de compra online é, portanto, essencial para o varejista satisfazer a expectativa do cliente e, conseqüentemente, fidelizá-lo.

Devido à complexidade do tema, uma vez que ele aborda tópicos de comportamento do consumidor, *marketing*, qualidade do serviço e operações logísticas, optou-se nessa pesquisa por realizar um recorte de estudo, focando na etapa menos eficiente do processo de compra online, o *Last Mile* (LM) (GEVAERS *et al.*, 2014; ZHANG *et al.*, 2018; SILVA *et al.*, 2019).

Por definição, LM é última parte da cadeia de suprimentos, onde o transporte do produto é feito do centro de distribuição até o cliente final (GEVAERS *et al.*, 2009). O problema desse processo é decorrente de questões logísticas, como, por exemplo, atraso na entrega e entregas perdidas por ausência de alguém para o recebimento, gerando assim problemas do ponto de vista financeiro dado os altos custos para realizar a entrega do produto (EUROPEAN COMMISSION, 2013). Nesse contexto, Van Duin *et al.* (2016) comentam que cerca de 25% dos problemas de LM estão relacionados a entregas não realizadas, o que é um problema pois, Rao *et al.* (2011) apontam que consumidores que experienciaram algum tipo de falha na entrega, tendem a reduzir a frequência e o tamanho, em termos de valores, de pedidos futuros com o vendedor em questão.

Diante do exposto, Al Nawayseh *et al.* (2013) sugerem o pick-up point (PUP) como a melhor proposta de solução para os problemas apresentados pelo LM no *Home Delivery* (HD) – sistema onde o consumidor faz a compra online e o produto é entregue em sua residência –. No sistema de PUP, o consumidor faz a compra online nos mesmos moldes tradicionais – define sua necessidade, define o produto, define o vendedor, negocia, compra, aguarda a entrega e avalia – porém, no subprocesso de entrega há uma alteração no LM, de modo que o produto não é mais entregue ao cliente diretamente, mas sim deixado em estabelecimentos parceiros do

varejista, localizados estrategicamente em locais de grande circulação de pessoas. Os PUP recebem a mercadoria e as armazenam, por um determinado tempo, até que o cliente vá ao local retirá-la (FRANCKE; VISSER, 2015).

As vantagens do PUP podem ser observadas tanto do ponto de vista do cliente como do ponto de vista do varejista. Por parte do varejista, o PUP reduz os custos logísticos e aumenta a eficiência no LM, a partir do contingenciamento de carga (FRANCKE; VISSER, 2015; MUÑOZ-VILLAMIZAR *et al.*, 2015; LE; UKKUSURI, 2019; SILVA *et al.*, 2019) e otimização dos custos de transporte (FRANCKE; VISSER, 2015; LE; UKKUSURI, 2019; SIVA *et al.*, 2019), além da redução do impacto ambiental (EDWARDS *et al.*, 2010; MUÑOZ-VILLAMIZAR *et al.*, 2015; LIU *et al.*, 2017; LE; UKKUSURI, 2019; SILVA *et al.*, 2019). Do ponto de vista do cliente, o PUP oferece vantagens como a eliminação do problema de não recebimento de mercadoria devido à ausência de alguém no local para recebe-la, além da redução do tempo que o consumidor precisa aguardar no local de entrega, uma vez que os horários de entrega podem ser agendados para serem realizados dentro de uma janela média de 2 horas, com um custo adicional de apenas 25% do custo operacional (AGATZ *et al.*, 2011), além de possibilitar custos de transporte mais baixos e prazos de entrega menores (CHEN *et al.*, 2018; LE; UKKUSURI, 2019).

Apesar das vantagens apresentadas de adotar como método de entrega o PUP, um estudo de Vakulenko *et al.* (2019) mostrou que, quando são adotadas inovações no LM, tais como o próprio PUP, há um aumento no grau de preocupação e ansiedade por parte dos clientes, o que pode ser um problema para as organizações, uma vez que para o cliente a jornada da compra online só é finalizada após o recebimento do produto. Desse modo, alguns estudos buscaram levantar e compreender qual a influência dos fatores envolvidos no *delivery* na percepção do consumidor ao optar por um serviço de entrega diferente do HD (XU; HONG, 2013; MORGANTI *et al.*, 2014; FAI *et al.*, 2018; VAKULENKO *et al.*, 2018; YUEN *et al.*, 2019). Entretanto, apenas os estudos de Oliveira *et al.* (2017) e Silva *et al.* (2019) buscaram realizar uma análise desse tipo no contexto brasileiro, o que mostra que a discussão ainda foi pouco desenvolvida, sendo, portanto, um problema, dada a situação de tendência de manutenção do crescimento dos números do e-commerce no Brasil.

Analisando o panorama da literatura global sobre soluções de *delivery* no LM, é possível observar que os estudos aqui citados abordam os fatores que influenciam a percepção do consumidor ao optar por um serviço de entrega diferente do HD. Entretanto, há um *gap* no que diz respeito às especificidades desses fatores. Ou seja, dentro da literatura, o que se observa são

estudos que verificam quais fatores influenciam, mas não verificam as especificidades desses fatores ou a sensibilidade dos consumidores a alterações nos níveis desses fatores.

Nesse sentido, uma vez apresentado este *gap* existente na literatura, os problemas das falhas que ocorrem no LM, em comparação ao HD, e suas consequências na satisfação do cliente, bem como o crescimento do mercado de compras online no decorrer dos anos – e sua previsão de crescimento para o futuro –, apresentam-se duas questões de pesquisa: **Dentre os fatores determinantes na escolha do consumidor entre *Pick-up Point* e *Home Delivery*, apresentados pela literatura, qual a sensibilidade do consumidor a alterações internas nos níveis desses fatores e qual a ordem de peso de influência que esses fatores exercem na opção do consumidor pelo *Pick-up Point* em detrimento ao *Home Delivery*, nas entregas de compras online?**

Aqui, entende-se por fatores determinantes as características envolvidas em cenários de escolha que influenciam na preferência do consumidor, neste caso, de receber o produto em casa ou em ir busca-lo em um PUP. Tais fatores são levantados na literatura de logística de entrega dos e-commerces e serão observados pelas lentes da teoria do custo de transação (TCT), de modo a dar suporte ao modelo de utilidade aleatória, como base teórica, para observar as preferências dos consumidores, que foram exploradas por meio da aplicação do método de Preferência Declarada (PD), utilizando a técnica de *discrete choice experiment* (DCE) e regressão logística para desenvolver os resultados e possíveis proposições ao final.

## 1.1 OBJETIVOS

### 1.1.1 Objetivo geral

Com intuito de responder às três questões de pesquisa, o objetivo geral dessa pesquisa é o de classificar os atributos envolvidos nos cenários em termos da influência deles sobre a preferência dos consumidores ao optarem pelo modo de entrega do tipo PUP em detrimento ao HD e analisar a existência da sensibilidade do consumidor a alterações nos níveis dos atributos.

### 1.1.2 Objetivos específicos

- Identificar quais fatores influenciam mais na escolha pelo PUP em detrimento ao HD;
- Identificar a sensibilidade dos consumidores à alterações nos níveis dos atributos envolvidos;

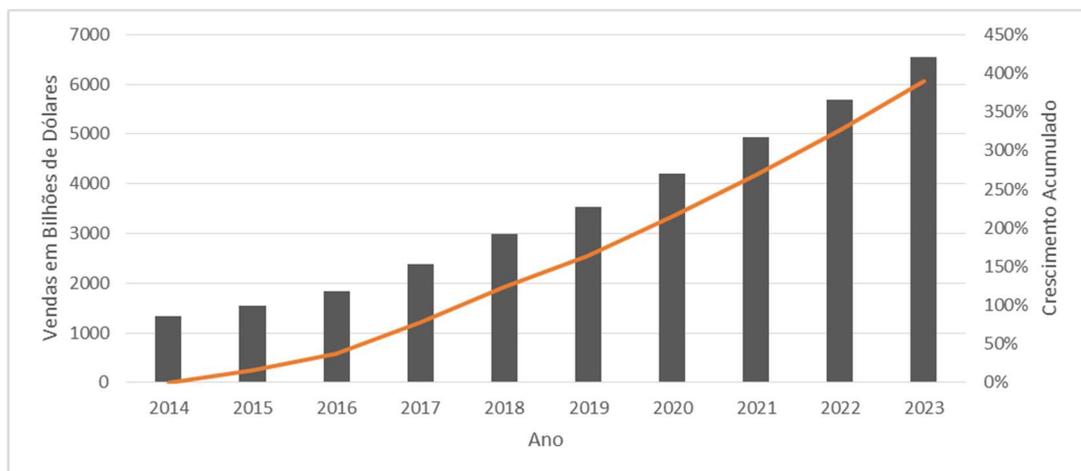
- Gerar proposições que expliquem o motivo de determinado fator ter maior peso de influência na preferência dos respondentes pelo PUP.
- Gerar proposições que expliquem a sensibilidade dos consumidores a alterações nos níveis dos atributos envolvidos.

## 1.2 JUSTIFICATIVA

### 1.2.1 Justificativa do tema

Os números do mercado de comércio eletrônico continuam em franca ascensão. No gráfico da Figura 1 é possível observar um estudo realizado pela Statista (2019), em que fica claro como a movimentação em vendas (em bilhões de dólares) está em crescente desde 2014, com previsão de crescimento acumulado de aproximadamente 400% até 2023.

Figura 1 – Vendas do comércio eletrônico no mundo



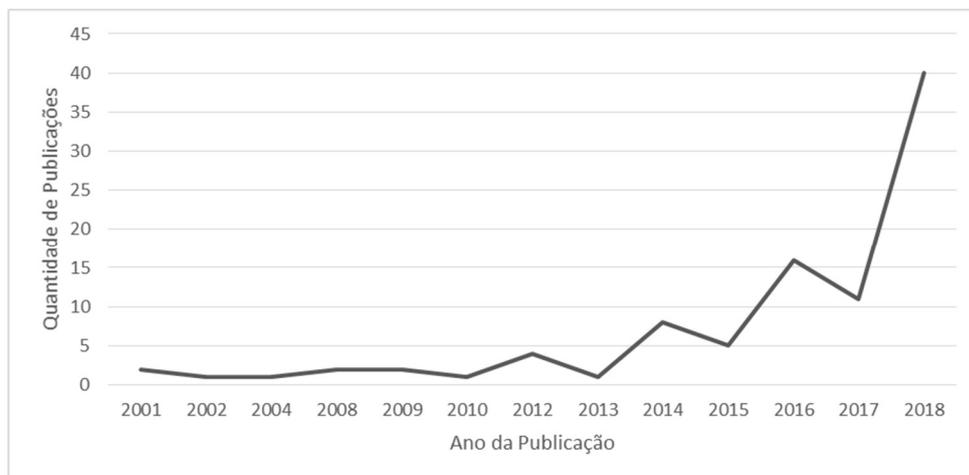
Fonte: Adaptado de Statista (2019).

Com o crescimento deste mercado, cresceram também as preocupações decorrentes de problemas que ocorrem nos processos da cadeia de suprimentos de vendas online. Problemas como os de entrega no LM fazem com que organizações busquem soluções, visando atender as expectativas dos consumidores, uma vez que não bastam as vantagens de se realizar a compra online – tais como comodidade, por exemplo –, pois os clientes desejam também o mesmo nível de vantagens no processo de entrega do produto (VAKULENKO *et al.*, 2019). Dada esta demanda das organizações, a academia passou a se interessar pelo tema e por suas discussões teóricas e empíricas (e.g. RAO *et al.*, 2011; OLIVEIRA *et al.*, 2017; NGUYEN *et al.*, 2018; SILVA, 2018; CHEN; WEI, 2018; VAKULENKO *et al.*, 2019; NGUYEN *et al.*, 2019) aumentando assim a produção científica relacionada ao assunto, como se observa no gráfico da

Figura 2, desenvolvido a partir de um estudo bibliométrico dos termos LM e “*delivery*”, utilizando como base de dados a *SCOPUS* e a *Web of Science*.

Como é possível observar, os temas envolvidos nesta pesquisa passaram a ter maior relevância a partir de 2010, com crescimento significativo a partir de 2014, sendo portanto um tema recente e relevante para a comunidade acadêmica, uma vez que essa tendência de crescimento deve se manter ativa, dado que o crescimento dos números do e-commerce não chegou a um platô de estagnação, como posto pelo estudo do Statista (2019), expresso na Figura 2.

Figura 2 – Produção anual de trabalhos



Fonte: elaborado pelo autor.

Apesar do crescimento significativo das publicações envolvendo os termos LM e *delivery*, no Brasil apenas os estudos de Oliveira *et al.* (2017) e Silva *et al.* (2019) analisam os *trade-offs* feitos pelos clientes ao fazer uma escolha entre o HD e o PUP.

Nesse sentido, essa pesquisa pode contribuir com a questão levantada por Le e Ukkusuri (2019), que explicam que o setor de entrega compartilhada – do qual o PUP faz parte – ainda sofre com uma espécie de “Dilema do ovo e da galinha”, uma vez que as organizações não sabem se investem na utilização de estratégias do tipo PUP no LM para gerar a demanda ou se esperam a demanda ser gerada para investir nesse tipo de estratégia. Por isso, o estudo contribui para a literatura de marketing, estratégias e gestão de operações logísticas por meio de evidências geradas acerca dos fatores que influenciam positivamente a percepção de utilidade do consumidor com relação ao PUP em detrimento ao HD.

Como contribuições práticas, essa pesquisa contribui para as organizações de varejo online na elaboração de suas estratégias de logística a fim de atender as expectativas dos consumidores, o que possibilita o aumento da sua satisfação e sua posterior intenção de recompra, fidelizando

o cliente e aumentando as chances da organização se manter competitiva no mercado (ZEITHAML, 2000; BOYER; HULT, 2005; RAO *et al.*, 2011; VAKULENKO *et al.*, 2019).

### **1.2.2 Justificativa do método**

Levando em consideração o fato de que não há, até o ano dessa pesquisa, nenhum PUP no território do Espírito Santo, Brasil, que é o local de recorte da pesquisa, fez-se necessário a escolha de um método que pudesse trabalhar com essa restrição de exploração. Nesse sentido, a escolha pela PD se justifica por atender a essa restrição, uma vez que trata-se de uma família de técnicas que estima a estrutura de preferências individuais de respondentes, a partir da percepção de utilidade que esses respondentes têm com relação a alternativas de escolha em determinadas situações hipotéticas (KROES; SHELDON, 1988; LOUVIERE *et al.*, 2000).

## 2. REVISÃO DA LITERATURA

Neste capítulo são apresentados e explanados os tópicos teóricos que embasam o desenvolvimento desta pesquisa e o problema de pesquisa, bem como a base teórica por trás do método de PD.

Levando em consideração os principais tópicos que envolvem o contexto do problema de pesquisa, este capítulo está organizado na seguinte ordem: primeiramente o conceito de LM e os problemas relacionados à esta etapa da logística de entrega de produtos, no contexto do e-commerce, bem como o PUP como solução a esses problemas; a TCT e como, a partir do olhar dela, o PUP se justifica como opção em detrimento ao HD; e, por fim, o Modelo de Utilidade Aleatória, que embasa a técnica de PD, que aqui é utilizada como base para compreender as preferências dos consumidores como uma escolha racional.

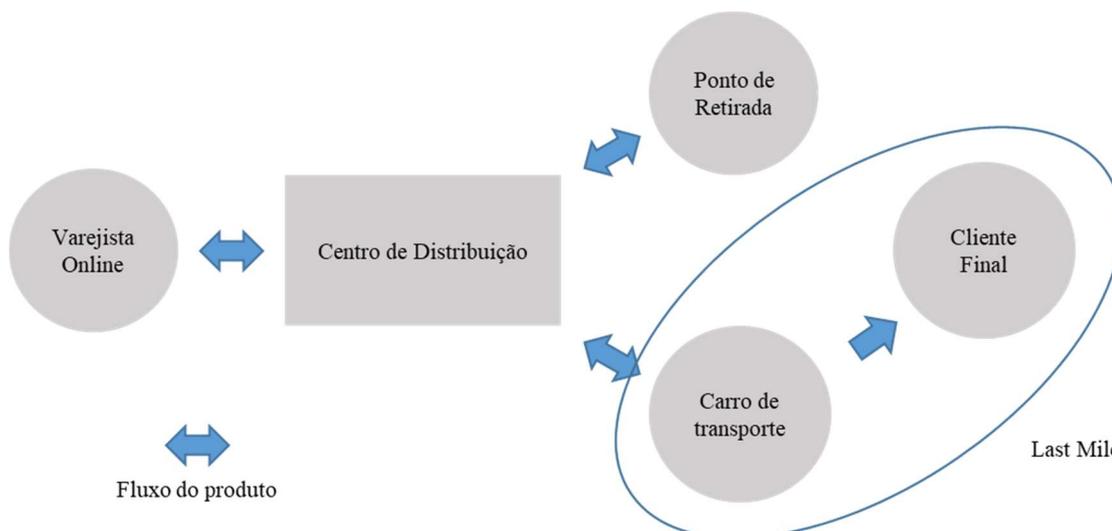
### 2.1 O LAST MILE E SUAS QUESTÕES NO CONTEXTO DO COMÉRCIO ELETRÔNICO

O processo de compra online é composto por um conjunto de seis estágios que são realizados em sequência, sendo eles: Identificação da Necessidade, Escolha do Produto, Escolha do Fornecedor, Negociação, Compra e Entrega e Avaliação. Na etapa de Compra e Entrega o cliente escolhe a forma como vai efetuar o pagamento e decide qual será a opção de entrega (MAES *et al.*, 1999). Após definido o modo de pagamento e de entrega, inicia-se o processo operacional logístico para levar o produto do fornecedor até o cliente. Dentro deste processo, a última etapa dele, quando o produto sai do galpão de distribuição do varejista ou do provedor logístico para ser entregue ao cliente, ou seja, a perna final da cadeia logística de entrega, é denominada de LM. Essa entrega pode ser realizada tanto diretamente ao cliente como também em algum ponto específico para que o cliente a busque (GEVAERS *et al.*, 2011), como se observa no esquema apresentado na Figura 3.

Esta etapa da cadeia de suprimentos dos varejistas online é a etapa menos eficiente devido principalmente a dificuldades de desempenho, que geram um alto custo nesse processo logístico (GEVAERS *et al.*, 2014; ZHANG *et al.*, 2018; SILVA *et al.*, 2019). Muitos são os problemas que a literatura aborda e que explicam esse alto custo do LM. Boyer e Prud'homme (2009), Agatz *et al.* (2011) e Gevaers *et al.* (2011), por exemplo, explicam que um dos grandes problemas está relacionado a falhas nas entregas devido à dificuldade em agendar para o cliente uma janela de entrega específica, devido aos altos custos desse tipo de operação. Como consequência, dado a falta de informação com relação ao momento exato da entrega, muitas

vezes o cliente não está disponível para receber a mercadoria, gerando a falha de entrega, problemas de processos e aumento dos custos logísticos.

Figura 3 – Esquematização do Last Mile



Fonte: Adaptado de Gevaers et al., 2011.

Além do problema de não encontrar o cliente no local de entrega, outro problema no LM decorre do atraso na entrega. De acordo com o trabalho de Rao *et al.* (2011), o atraso nas entregas acarreta em aumento da ansiedade do cliente nos próximos pedidos, além de reduzir o tamanho e o valor das futuras compras feitas por esse cliente com o mesmo varejista. Esse atraso ocorre pelo fato de que geralmente as transportadoras, buscando realizar consolidação de carga e otimizar a logística de entrega, retardam a entrega, de modo a otimizar os custos logísticos (CAMPBELL; SAVELSBERGH, 2005).

Para lidar com essas contingências do processo logístico de entrega, Lee e Whang (2001) apontaram cinco estratégias que as empresas podem adotar na formulação da estratégia de operação, sendo elas: Adiantamento Logístico, Desmaterialização, Troca de Recursos, Remessas Niveladas e *Clicks-and-Mortar*. Dentro dessas cinco estratégias, destaca-se para esta pesquisa a estratégia de *Clicks-and-Mortar*, que, segundo os autores, está relacionada com a ideia de envolver os consumidores em uma espécie de cooperação no LM. Nesse sentido, as encomendas seriam deixadas em postos de auto coleta onde os clientes poderiam ir até este local e fazer a retirada do produto, o que eliminaria riscos como os da encomenda não ser entregue devido à ausência de alguém no local. Além de reduzir os custos logísticos de entrega, uma vez que possibilita ao varejista realizar a consolidação de carga.

A estratégia de *clicks-and-mortar* é apresentada de várias maneiras na literatura, tais como Estações de Encomendas Automatizadas (MORGANTI *et al.*, 2014a), *Pick-up Points*

(McKINNON; TALLAM, 2003; AL NAWAYSEH *et al.*, 2013; MORGANTI *et al.*, 2014b; ZHOU *et al.*, 2016; VYT *et al.*, 2017) e *Parcel Locker* (VAN DUIN *et al.*, 2016; VAKULENKO *et al.*, 2018; DEUSTCH; GOLANY, 2018; ; TSAI; TIWASING, 2021). Todavia, o estudo realizado por Al Nawayseh *et al.* (2013) mostra evidências de que, de acordo com o nível de investimento que estão dispostos a realizar e a preferência dos consumidores, o PUP é a melhor solução estratégica logística.

Atualmente, levando em consideração que essa pesquisa foi desenvolvida entre os anos de 2019 e 2021, a literatura do LM no e-commerce vem abordando algumas situações específicas, principalmente a partir de 2020, quando se deu início a pandemia de coronavírus no mundo.

Alguns autores vêm tentando analisar soluções ainda mais vantajosas do ponto de vista da eficiência do processo, como por exemplo o estudo de Chen *et al.* (2021) que abordou a possibilidade de utilização de robôs no LM, dada a necessidade do distanciamento social gerado pela pandemia. Nesse estudo, os autores mostram que a utilização de tecnologias como robôs pode reduzir o custo no LM, principalmente ao atender clientes próximos uns dos outros, reforçando o conceito do contingenciamento de cargas como vantagem competitiva.

Outro estudo, realizado por Srivatsa e Marathe (2021), propõe a utilização de um caminhão itinerário que baseia sua localização geográfica de acordo com a densidade de demanda dos locais, reduzindo assim a soma do deslocamento do total de usuários. Além deles, estudos como os de Perbolli *et al.* (2021), que abordam o compartilhamento de centros de distribuição no LM, de Lemardelè *et al.* (2021), que trazem a utilização de drones no LM para atender a demanda de localizações complexas e, por fim, de Bergmann, Wagner e Winkenbach (2021), que abordam a logística compartilhada como solução no LM, todos esses estudos mostram que há uma corrente atual de pesquisa que visa discutir soluções de eficiência no LM, seja por meio de novas tecnologias – como robôs e drones – ou por meio de estratégias de compartilhamento e contingenciamento de cargas.

Outra corrente de estudos atualmente do LM e e-commerce está atrelado a estratégias de sustentabilidade. O estudo de Comi e Savchenko (2021), por exemplo, aborda uma metodologia onde, a partir do tamanho do pacote a ser entregue, avalia-se a utilização de métodos de entrega utilizando bicicleta ou a pé. Já o estudo de Rai, Verlinde e Macharis (2021), por exemplo, aborda uma outra visão, analisando o tipo de consumidor que estaria interessado em inovações sustentáveis no LM. Aproveitando da discussão do consumidor e da sustentabilidade, Ignat e Chankov (2020) discutem e sugerem que a quantidade de informações relativas à sustentabilidade ofertadas ao consumidor afeta a escolha do modo de entrega do *delivery*. Além

deles, autores como Siragusa *et al.* (2020), os quais trazem a discussão do carro elétrico como solução sustentável para o LM, Villa e Monzón (2020), que abordam a possibilidade de utilização de metrô como alternativas sustentáveis para as logísticas urbanas, e Patella *et al.* (2021), que apresentam uma revisão sistemática de veículos “verdes” no LM, destacam como esse tema é relevante nas discussões atuais de LM.

Além das estratégias de aumento da eficiência e de adoção de sustentabilidade, destaca-se também a discussão sobre a preferência do consumidor com relação ao LM, que é o espaço onde essa pesquisa se encaixa. Nesse sentido, observa-se trabalhos como os de Tsai e Tiwasing (2021), que abordaram a intenção dos consumidores em utilizar os chamados *smart lockers* a partir de uma visão baseada nas teorias de recursos integrados, da difusão da inovação e do comportamento planejado; o de Caspersen, Navrud e Bengtsson (2021), que observaram a intenção ou não das consumidoras em pagar por soluções de LM que sejam amigáveis ecologicamente – aqui fazendo um link com outro tema bem difundido que é o das estratégias de sustentabilidade –; e trabalhos como de Hagen e Scheel-Kopeinig (2021), que trazem a discussão sobre a intenção ou não dos consumidores em utilizar um conceito de estação compartilhada de retirada como forma de aumentar a eficiência de custos e sustentável do LM.

Todos esses temas são discutidos amplamente pela literatura de LM no contexto do comércio eletrônico. Todavia, é preciso levar em consideração que nos anos de 2020 e 2021, o mundo sofreu com a pandemia do covid-19, que acabou impactando a forma de vida das pessoas até então e, conseqüentemente, gerando algumas discussões na academia com relação ao efeito da pandemia no LM.

Na grande maioria dos locais do mundo, houve a adoção do isolamento social como forma de prevenção ao vírus e isso afetou o modo como os consumidores passaram a perceber valor nas transações e no LM isso não foi diferente. Nesse sentido, destacam-se estudos como os de Suguna *et al.* (2021), que identificaram os fatores mais relevantes do LM durante a crise do covid-19 como sendo o tipo de mercadoria adquirida, a eficiência da rota e o cumprimento dos cronogramas; como o de Settey *et al.* (2021), que abordaram uma metodologia para localização de galpões logísticos, levando em consideração a necessidade de redução de deslocamentos causado pelo covid-19; o estudo de Jiang *et al.* (2021), que aborda um algoritmo eficiente para a designação de pontos de coleta para consumidores que moram distantes dos grandes centros, dado a dificuldade de locomoção durante o covid-19; além de soluções que reduzem o contato e a interação humana no LM, como no estudo de Kapser, Abdelrahman e Bernecker (2021) que verificaram a aceitação de homens e mulheres à veículos autônomos nas entregas, no estudo de

Patchou, Sliwa e Wietfeld (2021) que simularam a utilização de drones e frota terrestre para reduzir as interações humanizadas entre os atores no LM e também o já citado estudo de Chen et al. (2021), que abordou a utilização de robôs no LM.

Como visto, a discussão atual sobre o LM tem suas bases na otimização da etapa de LM, na sustentabilidade dessa etapa e na percepção do consumidor nessa fase da logística de entrega. Todavia, a influência do covid-19 tem sido percebida pela academia e seus efeitos estudados, principalmente na mitigação da interação humana dentro dessa etapa. Na próxima seção é abordado a Teoria dos Custos de Transação e como ela pode influenciar a escolha do consumidor pelo método de entrega no LM.

## 2.2 TEORIA DOS CUSTOS DE TRANSAÇÃO (TCT)

Quando se aborda os custos envolvidos em uma transação comercial, é preciso compreender que não está embutido ali apenas os custos da produção de um determinado bem ou da oferta de um determinado serviço. Coase (1937) discorre sobre o argumento de que quando, por exemplo, um agente econômico produtivo vai ao mercado, por necessidade, buscar um fornecedor de insumos ou serviços, ele se depara com outros três tipos de custos, os chamados custos relacionados à transação, sendo eles: os custos de busca de informação, custos de negociação e barganha e, por fim, os custos para garantir o cumprimento de um acordo ou contrato. A transação é definida por Williamson (1985, p. 552) como um evento que ocorre quando “um bem ou serviço é transferido através de uma interface tecnologicamente separável”, sendo, portanto, objeto de estudo enquanto relação contratual entre os indivíduos envolvidos.

Os custos de transação são advindos da ideia de que as transações são como contratos formados entre esses agentes econômicos. Esses contratos são formados por expectativas com relação à entrega e a conduta desses agentes envolvidos, o que cria um ambiente de incertezas e riscos que geram os custos. Tais custos são distinguidos como *ex ante* e *ex post*, sendo os *ex ante* os custos para se estabelecer segurança no acordo a ser estabelecido. Nesse sentido, encaixam-se aqui custos da negociação, da elaboração e da segurança contratual, de modo a evitar os riscos envolvidos dada as incertezas do ambiente em que se dá a transação. Por sua vez, os custos *ex post* são aqueles relacionados ao monitoramento, adaptação as novas realidades e renegociação de um acordo, quando as situações expõem essa necessidade (WILLIAMSON, 1985).

O autor Williamson (1985) argumenta ainda que, por trás da incerteza e do risco envolvidos nessas transações, que levam aos custos transacionais, estão dois possíveis comportamentos desempenhados pelos agentes envolvidos, sendo eles: o comportamento oportunista e a racionalidade limitada.

### 2.2.1 Oportunismo e racionalidade limitada

Como explicitado, as relações econômicas, a luz da TCT, são estudadas com base nos contratos firmados entre as partes que executam a troca mercadológica. Essas trocas comerciais suscitam custos que, como mostrado na seção 2.2, não estão relacionados apenas à produção dos bens ou serviços, mas também relacionados à transação desses bens. Por trás desses custos, estão dois fatores: o comportamento oportunista e a racionalidade limitada.

O comportamento oportunista é definido por Williamson (1985) como a busca dos agentes envolvidos na transação em obter mais vantagens em prol do interesse próprio, de maneira consciente, gerando assim custos relacionados a desconfiança inerente a uma transação. De acordo com o autor, o comportamento oportunista ocorre tanto no *ex ante* como no *ex post*, de modo que, se esse comportamento não existisse, bastaria uma cláusula elaborada nos contratos *ex ante* garantindo que todas as relações e informações da transação ocorreriam de maneira clara, sincera e honesta, e não haveriam mais problemas *ex post*. Entretanto, essa é uma situação ideal e relativamente utópica, de modo que são necessários custos na elaboração e na salvaguarda de contratos *ex ante* para garantir que comportamentos oportunistas não influenciem nas trocas comerciais.

Um exemplo de comportamento oportunista pode ser observado a partir do estudo de Campbell e Savelsbergh (2005), onde os autores desenvolvem um método para maximizar os lucros em um processo de entrega, usando, por exemplo, a consolidação de cargas para reduzir os custos de entrega. Em contrapartida, Agatz *et al.* (2011) explicam que quanto menor a janela de entrega, maior a satisfação do consumidor. Todavia, há uma incompatibilidade de se operar com janelas de entrega que satisfaçam o consumidor e que respeitem o contingenciamento de carga, o que levaria a um aumento dos custos logísticos da transportadora.

Dentro desse contexto, deve-se levar também em consideração o estudo de Rao *et al.* (2011) que comenta que, diferentemente do processo de compra online, o processo de entrega do produto é feito após o pagamento e de maneira pouco visível, fazendo com que o cliente fique

particularmente preocupado com um possível comportamento oportunista do varejista, uma vez que o pagamento já foi feito, mas o produto ainda não foi recebido.

Esse comportamento oportunista pode acontecer com o varejista prometendo realizar a entrega em uma data tal que atenda a expectativa do consumidor, mas entregando de fato em outra data, por conta da possibilidade de contingenciar a carga. Nesse sentido, na perspectiva do vendedor, o PUP reduz o custo de transação, uma vez que uma das vantagens do PUP para o varejista é justamente o contingenciamento de carga (FRANCKE; VISSER, 2015; MUÑOZ-VILLAMIZAR *et al.*, 2015; LE; UKKUSURI, 2019), fazendo com que o risco de um comportamento oportunista por parte da transportadora seja reduzido, já que ela não mais precisa fazer o *trade-off* entre janela de entrega, citado por Agatz *et al.* (2011), e a satisfação do cliente.

Já a racionalidade limitada baseia-se na ideia levantada por Simon (1955) de que, devido a limitação natural do indivíduo em obter, processar e carregar todas as informações existentes em uma determinada situação, o agente fará sua transação ou tomará sua decisão baseado nas informações que estiverem disponíveis, buscando assim não o resultado perfeito, mas sim o melhor possível dentro dessas limitações, o que gera custos também, devido a impossibilidade de se conhecer e de analisar todos os riscos envolvidos. Williamson (1985) explica que os agentes, dada a limitação racional *ex ante*, acabam por incluir cláusulas de segurança nos acordos contratuais para tentar reduzir ao máximo os riscos de problemas *ex post*, de modo a gerar custos devido justamente para reduzir o impacto da racionalidade limitada.

Um exemplo da influência da racionalidade limitada, aplicada ao contexto de relacionamento entre clientes e varejistas no LM, pode ser observado nos estudos de Rao *et al.* (2014) e Yuen *et al.* (2019). Os autores explicam que, devido à ausência de informação exata sobre a data e horário de entrega dos produtos, o consumidor precisa estar disponível para receber a encomenda em uma janela de tempo relativamente grande, aumentando os custos dessa transação. Nesse sentido, o PUP reduz o custo de transação, por exemplo, ao diminuir a incerteza que a limitação da informação do horário de entrega cria, uma vez que o consumidor pode ir até o local para retirar o produto na hora que preferir.

Ambos os comportamentos, oportunista e de racionalidade limitada, estão embutidos nos três custos transacionais trazidos por Coase (1973). Todavia, as transações possuem características são determinantes para o aumento ou a redução dos custos de transação envolvidos. Williamson (1985) afirma que essas categorias são baseadas em três dimensões, sendo elas: a frequência transacional, o grau de incerteza subjacente e a especificidade do ativo envolvido.

## 2.2.2 Categorias dos custos de transação

Após abordar os fatores comportamentais que estão por trás do surgimento dos custos de transação, Williamson (1985) explica as características que formam esses custos, e que influenciam positivamente ou negativamente no aumento dos custos de transação.

Nesse cenário, três dimensões estabelecem as características das transações: a frequência transacional, o grau de incerteza subjacente e a especificidade do ativo envolvido.

Entende-se por frequência transacional a frequência com que um agente transaciona com outro agente. Essa frequência gera maior capacidade analítica por parte desses agentes com relação a seus comportamentos, dado o histórico de interações entre ambos. De acordo com Farina *et al.* (1997), a frequência transacional influencia no custo de transação de duas maneiras, sendo elas:

- a) Ao diluir, nas várias transações, os custos de adoção de um mecanismo complexo;
- b) Ao gerar uma reputação entre os agentes envolvidos na transação, dada a repetição das transações.

Em outras palavras, quanto maior a frequência de interação entre os agentes envolvidos, menor o risco percebido e, conseqüentemente, menor o custo envolvido para mitigar as incertezas da transação. Os autores ainda classificam as frequências das transações em diferentes níveis de frequência, e, dentro dessas classificações, a transação efetuada para a entrega de um produto adquirido no e-commerce é tratado como transação *spot*, caracterizada por ser uma transação que não se repete continuamente entre os agentes e, por isso, não se justifica uma estrutura complexa de controle de possíveis atitudes oportunistas. Todavia, apesar de as transações *spot* não justificarem uma estrutura de controle muito rígida e custosa, o fato de as transações não serem tão frequentes, dificulta a geração de reputação entre os envolvidos, aumentando o risco percebido pelos clientes com relação a um possível comportamento oportunista por parte das transportadoras no que diz respeito às janelas de entrega, como comentado na seção 2.2.1.

Por sua vez, o grau de incerteza subjacente está ligado à instabilidade do meio no qual os agentes da transação se encontram. De acordo com Williamson (1985), o problema está nas sociedades conseguirem se adaptar a essas mudanças no tempo e espaço. Nesse sentido, o grau de incerteza subjacente está intimamente ligado à questão da racionalidade limitada, uma vez que, mediante cenário incerto e nebuloso, a percepção do risco envolvido aumenta, e, conseqüentemente, aumentam os custos para o controle desse risco, como mecanismos de governança e contratos que tentem reduzir os riscos *ex post*.

Endossando esse argumento, Farina *et al.* (1997) explicam que muitas vezes esse grau de incerteza deriva da assimetria de informações entre os agentes envolvidos na transação. Há incerteza na formação, elaboração e conteúdo dos contratos e daquilo que foi acordado, e principalmente, há incerteza sobre qual o nível de informação o agente com quem se está transacionando possui. Traçando um paralelo com a questão do LM no contexto do e-commerce, há a complexa situação relacionada com a janela do horário de entrega do bem adquirido, ou a garantia de um produto entregue em boas condições. Essas situações fazem com que o custo da transação possa ser maior ou menor, dependendo do nível de incerteza desses fatores apresentados. Logo, quanto mais informações disponíveis para os agentes envolvidos com relação ao momento de entrega, menor é o custo da transação.

Por fim, a especificidade do ativo está relacionada com o quão específico o ativo é no sentido de não poder ser reaproveitado ou realocado de outra maneira, caso a transação não aconteça mais, sem que esse ativo perca valor ou capacidade, levando assim os agentes a terem menor condições de controle de riscos sobre possíveis encerramentos de contrato. Esses ativos são classificados em seis tipos (WILLIAMSON, 1985):

- a) Especificidade locacional: quando a proximidade dos elos da cadeia produtiva contribui para a diminuição dos custos de transporte e inventário;
- b) Especificidade dos ativos físicos: são instalações necessárias para a produção do ativo específico, gerando assim mais custos;
- c) Especificidade de ativos humanos: recursos humanos necessários para a produção do ativo específico, gerando assim mais custos;
- d) Especificidade de ativos dedicados: recursos de investimentos necessários para que a transação ocorra com um cliente ou fornecedor em particular;
- e) Especificidade da marca: exclusividade que a marca tem no mercado, aumentando a dependência desse agente;
- f) Especificidade temporal: quando o tempo é um fator crítico para a transação.

Nesse sentido, diante de um cenário de risco onde não há certeza de retorno total sobre o investimento em um ativo específico, são inseridos mecanismos de segurança nos contratos, de modo que o ônus de um problema relacionado a especificidade de um ativo não seja empregado à apenas uma das partes.

No contexto do LM, a especificidade do ativo pode ser do tipo temporal. Por exemplo, o cliente pode optar por comprar um produto que precise ser entregue até uma determinada data, como um presente de aniversário ou qualquer outro item que perca sua utilidade após essa data

específica. Nesse sentido, o tempo é um fator crítico para a transação, aumentando o custo da mesma. Levando em consideração os estudos de Chen *et al.* (2018) e Le e Ukkusuri (2019), o prazo de entrega do PUP geralmente é menor que o do HD, fazendo com que o custo da transação devido a especificidade temporal seja menor para ao escolher o PUP.

Portanto, no contexto do LM, a ausência da total compreensão racional das consequências das escolhas dos indivíduos e a existência de um possível comportamento oportunista na transação, dentro das características determinantes dos custos de transação (frequência transacional, grau de incerteza subjacente e especificidade do ativo), fazem com que os indivíduos ponderem, entre PUP e HD, qual a opção que o satisfaz com menor custo de transação possível.

Na próxima seção, é apresentado o modelo de utilidade aleatória e como ela justifica e embasa a escolha do cliente no contexto de entrega de produtos comprados no e-commerce.

### 2.3 MODELO DE UTILIDADE ALEATÓRIA

A PD é um conjunto de diferentes técnicas que tem como objetivo modelar o comportamento dos consumidores (tomadores de decisão) ao se confrontarem com uma situação hipotética de escolha, e aparece como uma opção para estimar a estrutura de preferência de um respondente em uma hipotética situação de tomada de decisão (KROES; SHELDON, 1988). A técnica projeta as preferências do consumidor por meio da observação do indivíduo e suas preferências, partindo do pressuposto que, no papel de tomador de decisão, este indivíduo faz suas escolhas após analisar as características e atributos das opções disponíveis, verificando e elegendo as que prefere, visando maximizar sua utilidade (VON NEUMANN; MORGENSTERN, 1944; KROES; SHELDON, 1988; HENSHER *et al.*, 2005).

Dado este cenário, a tomada de decisão é realizada a partir da análise e percepção do indivíduo com relação a utilidade total de uma escolha. Kroes e Sheldon (1988) explicam que a utilidade total é quantificada por meio de uma função matemática conhecida por Função Utilidade. Essa função, apresentada na equação 1, representa a percepção geral que o indivíduo tem com relação ao conjunto de utilidades de cada opção de escolha, de acordo com as características dos fatores que compõe cada uma dessas opções.

$$U = \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_n x_n \quad (1)$$

Onde:

U = Utilidade total;

$\alpha_1$  até  $\alpha_n$  = Peso da utilidade dos fatores 1 até n;

$x_1$  até  $x_n$  = Fatores de 1 até n.

Os autores explicam que os fatores são características que compõe a formação de percepção do indivíduo sobre determinada escolha e que podem influenciar a percepção de utilidade desse indivíduo. Por exemplo, a utilidade total na tomada de decisão quando se quer escolher entre comprar uma camisa na loja A ou na loja B pode estar ligado a fatores como preço em cada loja, possuir ou não atendimento, formas de pagamento, distância até a loja, entre outros.

Todavia, apesar da utilidade representar a percepção que o indivíduo usa para selecionar a opção que vai maximizar a recompensa social obtida por este indivíduo em um processo decisório, nem sempre o indivíduo seleciona a opção que mais lhe agrada e isso se dá por uma espécie de componente aleatória na Função Utilidade. Por exemplo, utilizando a mesma situação de compra de camisa na loja A ou B, o observador pode questionar a preferência do indivíduo por meio do preço, distância e forma de pagamento. Entretanto, como o observador não inseriu nesse contexto todos os outros fatores que, por algum motivo, possam influenciar na tomada de decisão (por ex.: dia do mês, proximidade do cliente com os donos da loja, reputação da loja entre outros), a escolha absorve um certo nível de aleatoriedade, e essa inconsistência nos padrões de comportamento é captada pelo que se conhece por Parcela Aleatória.

Manski (1977) comenta que os psicólogos, ao perceberem essa inconsistência nos padrões de comportamento dos indivíduos ao realizarem uma tomada de decisão, passaram a desenvolver modelos aleatórios dessa utilidade para tentar captar esse fenômeno. Ou seja, passaram a tratar a Utilidade como uma variável aleatória, uma vez que, ao olhar do observador, dada a diferença de características dos indivíduos e das alternativas postas e, dada a ausência de total informação por parte dos indivíduos, e do pesquisador, com relação aos atributos e características de cada opção posta em um cenário de escolha, a utilidade percebida por cada um é diferente e, por isso, gera comportamentos de escolhas diferentes buscando sempre a alternativa que lhe ofereça a mais alta utilidade. A equação que representa esse modelo então é a Equação 2:

$$U_{i_n} = \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_n x_n + \varepsilon_i \quad (2)$$

Onde:

$U_{i_n}$  = Utilidade aleatória da alternativa “i” para o indivíduo “n”;

$\alpha_1$  até  $\alpha_n$  = Peso da utilidade dos fatores 1 até n;

$x_1$  até  $x_n$  = Fatores de 1 até n.

$\varepsilon_i$  = Parcela Aleatória da alternativa “i”

Apresentado o Modelo de Utilidade Aleatória, suas bases teóricas e como elas explicam a predição do comportamento do consumidor na tomada de decisão, o próximo capítulo destrincha o método utilizado para a realização dessa pesquisa.

### 3. PROCEDIMENTO METODOLÓGICO

Este capítulo tem como objetivo apresentar o método utilizado na pesquisa para desenvolver a exploração necessária e responder ao problema de pesquisa. Para atingir esse objetivo, optou-se por utilizar o método de PD. A opção por essa técnica se deu pelo fato dela trabalhar a preferência das pessoas mediante situações hipotéticas de escolha (KROES; SHELDON, 1988), sendo assim ideal para o cenário desse estudo onde o PUP ainda não é uma realidade existente.

A seguir, é apresentado uma breve conceituação da técnica de PD, bem como o desenho do método que servirá de base para a obtenção dos dados primários a serem utilizados na pesquisa.

#### 3.2 PREFERÊNCIA DECLARADA

O termo PD é um termo designado por Kroes e Sheldon (1988) para a família de técnicas que buscam estimar a utilidade auferida por um indivíduo às opções envolvidas em um cenário real ou hipotético de escolhas, por meio da declaração das suas preferências.

A DCE, uma das técnicas de PD, é definida por Louviere *et al.* (2010) como sendo uma técnica quantitativa que utiliza como pressuposto teórico o Modelo de Utilidade Aleatória, visa elicitare a preferência dos indivíduos sobre alternativas de bens ou serviços em determinados cenários hipotéticos. Ou seja, após observar e internalizar quais vantagens pode obter com cada produto ou serviço, o consumidor mensura a utilidade de cada um deles, por meio de *trade-offs* entre os benefícios e as desvantagens de cada opção. Ao final, acredita-se que este consumidor optará pela opção que ofertar uma melhor relação vantagem *versus* desvantagem, levando em consideração algumas restrições como, por exemplo, o orçamento que dispõe (LOUVIERE *et al.*, 2000).

A importância da técnica de PD na realização de estudos está na possibilidade de observar os custos e benefícios que são esperados pelos consumidores ao se depararem com a possibilidade de escolher um novo produto ou alguma mudança num processo ou serviço. Essa possibilidade é viável por meio do levantamento de dados sobre o que os consumidores fariam em situações hipotéticas desse tipo e quais seriam suas escolhas nesses cenários. Todavia, uma crítica que se faz a esse modelo é que os dados obtidos não refletem com confiança total a realidade, uma vez que são dados obtidos por meio de uma situação, como dito, hipotética. Entretanto, um

contraponto é o de que as técnicas de PD possibilitam analisar situações hipotéticas, uma vez que permite observar cenários ainda não existentes (LOUVIERE *et al.*, 2000).

Modelos DCE podem utilizar dados secundários ou primários, obtidos por meio de *survey*, seja ela presencialmente, por telefone ou internet. De acordo com Louviere *et al.* (2000), existem quatro tipos de respostas que podem ser exigidas dos respondentes nessa *survey*: escolha discreta, *rankeamento*, *rating* e aceitação de uma alternativa. A escolha discreta é uma forma de resposta onde o respondente escolhe apenas uma das opções a ele apresentadas. Todavia, o ponto negativo dessa forma de resposta é que ela não fornece informação sobre as alternativas não escolhidas, ou seja, faz-se necessário mais respostas por respondente e, conseqüentemente, maior número de respondentes para se conseguir obter a variação nos atributos da amostra. O *rankeamento* é uma forma de resposta onde o respondente deverá elencar, em ordem de preferência, as alternativas a ele apresentadas. O ponto negativo dessa forma é que, além de muitas vezes o respondente encarar o *rankeamento* das alternativas como algo complexo, não se tem informação sobre o grau de importância que o consumidor atribui à preferência de uma alternativa sobre a outra. O *rating* por sua vez é uma forma de resposta onde o respondente deve dar uma nota, dentro de uma escala definida pelo pesquisador, para avaliar cada uma das alternativas. Essa forma de resposta é, de acordo com os autores, a forma mais confiável de se compreender as preferências dos respondentes por cada uma das alternativas. Todavia, existem críticas a esse formato, uma vez que não é possível garantir que as preferências possam ser transformadas em escalas numéricas, criando assim certo nível de inconfiabilidade (ORTÚZAR; WILLUMSEN, 2001). Por fim, a aceitação de uma alternativa é uma forma de resposta onde o respondente classifica como aceitável ou não-aceitável as alternativas, por meio da análise dos atributos e dos níveis dos atributos. Todavia, como essa forma não oferece uma informação mais acurada com relação ao nível de aceitação ou de não-aceitação, o formato requer um tamanho de amostra maior. É possível observar um exemplo de como se daria cada um dos quatro tipos de respostas em uma *survey* na Tabela 1.

Tabela 1 - Formas de resposta de uma survey de PD

Alternativas	Escolha discreta	Rankeamento	Rating (de 1 a 5)	Aceitação
Bicicleta		4	2	Não
Ônibus		3	2	Não
Carro	X	1	5	Sim
Transporte Escolar		2	4	Sim

Fonte: adaptado de Louviere *et al.* (2000)

Partindo do pressuposto que a questão de pesquisa desta pesquisa envolve apenas duas opções, o HD e o PUP, a forma de resposta escolhida para obter as respostas dos respondentes foi a “Escolha discreta” que permite ao respondente optar por uma opção em detrimento a outra.

No caso do “Rating”, essa opção não foi escolhida pois, de acordo com os argumentos de Hensher et al. (2005), esse modelo de resposta fornece para o pesquisador a preferência do respondente e não necessariamente sua escolha. Além disso, a escala que se utiliza é arbitrariada pelo pesquisador e isso pode causar viés nas informações obtidas. Por fim, os autores argumentam que não é possível determinar como cada respondente mensura o peso de uma nota e de outra. Por exemplo, não é possível afirmar que um respondente que der uma nota 4, em uma escala de 10 pontos, prefere mais uma opção do que outro respondente que der a nota 3, na mesma escala.

A opção de resposta do tipo “Aceitação” foi descartada porque as informações obtidas a partir desse modo acabam por demonstrar, de maneira individual, se o respondente utilizaria ou não determinada opção. No caso dessa pesquisa, o objetivo é o de confrontar, em um cenário experimental, qual das duas opções o respondente escolheria, no intuito de observar quais fatores afetam mais essa escolha. Uma resposta do tipo “Aceitação” mostra, na realidade, se a opção é ou não interessante para o respondente que acaba demonstrando esse interesse de forma independente por cada opção a ele fornecida.

A seguir, são apresentados o modo de fazer e as preocupações que se deve ter ao criar o desenho de uma PD, bem como o desenho de DCE utilizado nesta pesquisa.

### 3.3 DESENHO EXPERIMENTAL

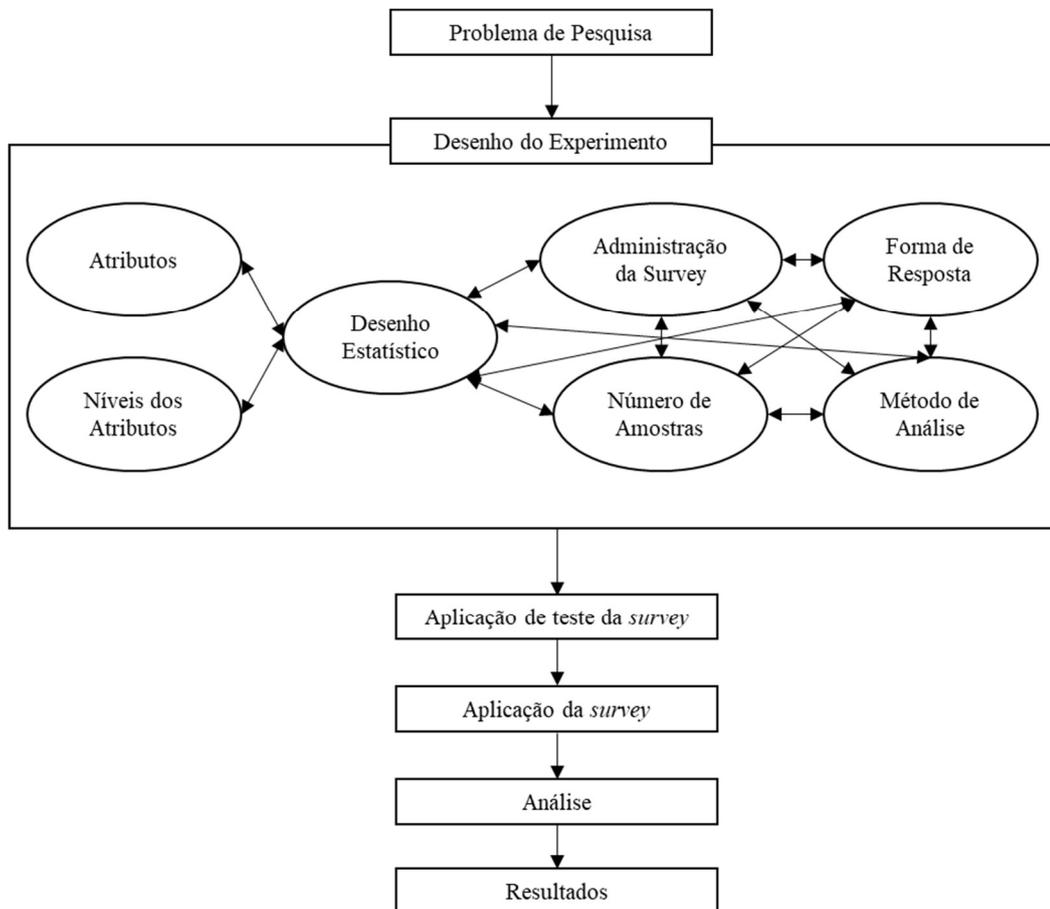
As técnicas de PD precisam seguir um procedimento que inclui o desenvolvimento de um desenho experimental onde são abordados alguns temas de preocupação, tais como a forma de resposta, o método de análise, número de amostras, as alternativas, seus atributos e os níveis dos atributos e, por fim, a administração da aplicação da *survey* (SANKO, 2001). O procedimento das técnicas de PD é observado na Figura 4, com o acréscimo de aplicação de teste da *survey*.

A seguir, são apresentados cada etapa do desenho experimental, bem como sua estrutura e elaboração.

### 3.2.1 Problema de pesquisa

A primeira etapa consiste na definição do Problema de Pesquisa. Essa etapa tem como objetivo deixar explícito qual o problema de pesquisa que a técnica de PD irá apoiar. No caso dessa pesquisa, a questão de abordada é a seguinte: Quais são os fatores significantes (atributos e seus níveis), a sensibilidade do consumidor a alterações internas dos atributos e a ordem de influência que estes fatores exercem, na preferência do consumidor pelo PUP em detrimento ao HD, nas entregas de compras online?

Figura 4 – Procedimentos de Preferência Declarada



Fonte: adaptado de Sanko (2001)

Os autores Hensher *et al.* (2005) explicam que, nesse momento, é necessário definir quais serão as alternativas envolvidas no processo de escolha. Para essa seleção, os autores sugerem fazer uma lista com todas as alternativas passíveis de escolhas pelos consumidores no contexto da pesquisa. No caso dessa pesquisa, as opções envolvidas já são todas as opções disponíveis, ou seja, HD e PUP, sendo essas duas, portanto, as duas alternativas que serão apresentadas aos respondentes em todas as situações de escolha.

A seguir, é apresentado os atributos dessas alternativas, como sugere a Figura 4.

### 3.2.2 Atributos das alternativas

Uma vez determinadas as alternativas que serão apresentadas aos respondentes, é o momento de determinar os atributos dessas alternativas que comporão os cenários. De acordo com Louviere *et al.* (2000), a seleção desses atributos deve ser feita de modo a identificar aqueles atributos que influenciam de fato a preferência do respondente com relação às alternativas apresentadas. Alguns cuidados devem ser tomados nessa etapa, como por exemplo a clareza com que os atributos são descritos aos respondentes, de modo a não causar dúvida ou dubiedade sobre a interpretação do atributo – como por exemplo a confusão cognitiva que pode ser feita entre as definições de preço e valor, por exemplo, caso um dos dois fossem atributo – além disso, é importante estar atento às possíveis correlações cognitivas inter-atributo (HENSHER *et al.*, 2005) – como por exemplo em uma possível relação feita pelo respondente entre qualidade do serviço e preço cobrado.

É importante destacar também que existem inúmeros atributos que podem influenciar a tomada de decisão de um consumidor ao optar por HD ou pelo PUP. Todavia, Louviere *et al.* (2000) afirmam que é preciso parcimônia na quantidade de atributos inseridos em um estudo de PD, uma vez que o excesso de atributos apresentados ao respondente pode gerar cansaço do mesmo e, conseqüentemente, obter respostas que não representem fidedignamente a realidade.

Portanto, na intenção de evitar os riscos citados de se utilizar uma quantidade excessiva de atributos no estudo, foram levantados os principais pontos utilizados em alguns trabalhos, citados como os principais motivos de insatisfação dos clientes do varejo online com relação ao LM ou principais atributos que são valorizados no processo de delivery. Tais estudos são apresentados no Quadro 1.

Quadro 1 – Quadro sinótico de atributos e estudos anteriores

<b>Autores</b>	<b>Ano da Publicação</b>	<b>Atributos utilizados</b>	<b>Observação</b>
Nguyen <i>et al.</i>	2019	Janela do tempo de entrega, prazo de entrega e taxas de frete	Autores descobriram que as taxas de frete era a variável de maior influência.
Yuen <i>et al.</i>	2019	Distância do PUP, flexibilidade do horário de entrega, preço do frete e experiências anteriores	O estudo concluiu que todas as variáveis agiram como mediadoras entre o valor percebido e a intenção de utilizar o PUP.

<b>Autores</b>	<b>Ano da Publicação</b>	<b>Atributos utilizados</b>	<b>Observação</b>
Nguyen <i>et al.</i>	2018	Velocidade na entrega	Os autores trazem a velocidade na entrega de um produto como ponto essencial na percepção de qualidade do serviço por parte do consumidor.
Silva	2018	Preço do frete, prazo de entrega, janela de espera e acessibilidade aos PUP	O autor observou que o atributo Janela de Entrega foi estatisticamente não significante.
Chen; Wei	2018	Conveniência do local, pressão do tempo de entrega, interação com humanos, grau de inovatividade, grau de otimismo, intenção de uso	Os autores observaram que, se o PUP está localizado em local conveniente ao consumidor, isso afeta positivamente a intenção de uso. A pressão do tempo não afetou a intenção de uso. Os fatores de inovatividade e otimismo estão fortemente ligados à intenção de uso.
Eurobarometer	2013	Preço, velocidade de entrega e distância percorrida para receber a encomenda	O estudo mostrou que essas variáveis representavam os motivos de maior insatisfação por parte dos consumidores.
Feng; Huang	2005	Preço do frete, qualidade do serviço, ação de marketing	Os autores observaram que a lealdade aumenta quando as ações de marketing existem e o preço do frete e a qualidade do serviço agem como variáveis de controle no modelo.

Fonte: elaborado pelo autor.

Optou-se nessa pesquisa por utilizar os atributos de preço do frete, distância para o PUP e prazo de entrega, pois foram atributos citados em praticamente todos os estudos como significantes e importantes no que diz respeito à influência sobre o comportamento de escolha do consumidor de e-commerce que vai escolher o método de entrega do produto.

A definição conceitual de cada um desses atributos escolhidos é apresentada no Quadro 2.

Quadro 2 – Atributos utilizados no instrumento de Preferência Declarada

<b>Atributo</b>	<b>Definição</b>	<b>Autores que utilizaram as variáveis</b>
Frete	Valor pago pelo consumidor pelo serviço de frete, seja ele para entrega da encomenda em casa, no caso de optar pelo HD; ou para entrega da encomenda em um <i>pick-up point</i> , no caso de optar pelo PUP.	Nguyen <i>et al.</i> , 2019; Yuen <i>et al.</i> , 2019; Silva, 2018; Eurobarometer, 2013; Feng; Huang, 2005

Atributo	Definição	Autores que utilizaram as variáveis
Prazo	Tempo informado pelo varejista para a realização da entrega da encomenda, seja na casa do consumidor, em caso de optar pelo HD; ou para entrega da encomenda em um <i>pick-up point</i> , no caso de optar pelo PUP.	Nguyen <i>et al.</i> , 2019; Nguyen <i>et al.</i> , 2018; Silva, 2018; Chen; Wei, 2018; Eurobarometer, 2013
Localização	Distância que o consumidor precisará percorrer fora da sua rotina diária (casa, ida para o trabalho, ida para a escolha e etc.) para receber a encomenda.	Yuen <i>et al.</i> , 2019; Silva, 2018; Chen; Wei, 2018; Eurobarometer, 2013

Fonte: elaborado pelo autor.

Definidos os atributos, o próximo passo é a definição de níveis para esses atributos.

### 3.2.3 Níveis dos atributos

A quantidade de níveis de atributos envolvidos no desenho afeta a habilidade do pesquisador na distinção de possíveis relacionamentos não lineares entre o valor do atributo e a utilidade derivada desse atributo. Isto é, se o pesquisador monta um desenho onde um ou mais dos atributos possui apenas dois níveis – por exemplo, atributo “Temperatura” e níveis do atributo como “Quente” ou “Frio” – conclui-se que a relação entre esses níveis de atributos e a utilidade derivada deles é linear. Por outro lado, quando a quantidade de níveis de atributos aumenta para três, quatro ou mais variáveis, é possível observar relações mais complexas entre esses atributos e suas utilidades. Todavia, é preciso ser parcimonioso, uma vez que quanto mais atributos envolvidos no experimento, maior o tamanho da amostra necessária para se obter resultados estatisticamente confiáveis (HENSHER *et al.*, 2005).

Para a definição dos níveis de atributos do experimento, Hensher *et al.* (2005) explicam que não há uma forma definida na literatura para realizar essa definição. Os autores comentam que essa definição pode ser feita com base em dados de preferência revelada ou, também, por meio de entrevistas com grupos focais.

Todavia, dado a impossibilidade de se obter dados de preferência revelada no caso dessa pesquisa, e também a impossibilidade de se montar grupos focais presenciais, dado o ambiente de pandemia de covid-19 na época da pesquisa, optou-se por gerar os níveis de atributos com base numa discussão entre o orientador dessa pesquisa, o co-orientador e o pesquisador mestrando.

Nessa discussão, foram levantadas possibilidades de percentual de desconto no valor do frete e redução do prazo de entrega para os cenários nos quais o entrevistado optasse pelo PUP. Nesse sentido, entendeu-se que 15% e 30% de desconto seriam possibilidades viáveis e realistas para descontos em entregas via PUP. No caso do prazo de entrega, entendeu-se que 2 ou 4 dias a menos seriam possibilidades viáveis e realistas para entregas via PUP, de modo a captar a sensibilidade do respondente com relação à variação desses atributos.

No caso do atributo Localização, uma vez que o PUP deve, por conceito, estar próximo aos locais de maior concentração de consumidores de e-commerce, entendeu-se que 2 quilômetros seria uma distância razoável para medir a sensibilidade dos respondentes com relação a esse atributo. Por isso, para efeito de comparação, inseriu-se também a possibilidade de 1 quilômetro de distância. Todos os níveis de atributos montados podem ser observados pelo Quadro 3.

Quadro 3 – Níveis de atributos utilizados

Nível	Frete	Prazo	Localização
-1	PUP e HD com preços iguais	PUP e HD com prazo de entrega iguais	O PUP está localizado entre 1 e 2 km da sua residência
0	PUP 15% mais barato	Previsão de entrega 2 dias mais rápido.	O PUP está localizado em até 1 km da sua residência
1	PUP 30% mais barato	Previsão de entrega 4 dias mais rápido.	O PUP está localizado ao lado da sua residência

Fonte: elaborado pelo autor.

Por fim, nessa pesquisa, optou-se por apresentar os níveis de atributos de forma rotulada, ou seja, apresentando os termos e os valores envolvidos. Hensher *et al.* (2005) explicam que a opção de rotular ou não os níveis de atributos devem partir do pesquisador e de sua intenção com a pesquisa, cuidando apenas para que o respondente não expresse juízo de valor sobre as alternativas apenas por conta dos rótulos que elas recebem. Os níveis são codificados como -1, 0 e 1 para manterem uma estrutura ortogonal, uma vez que os níveis dos atributos são equidistantes, em termos absolutos, entre si.

### 3.2.4 Desenho Estatístico

Definidas as alternativas, os atributos destas e os níveis que serão usados para esses atributos, é o momento de definir o desenho estatístico. Hensher *et al.* (2005) comentam que existem

vários desenhos, mas que os pesquisadores tendem a utilizar o desenho fatorial completo para os trabalhos, uma vez que esse método aborda todas as opções possíveis para os respondentes. Os autores definem esse desenho fatorial completo como um desenho onde todas as possíveis combinações de cenários – obtidas pelas combinações dos níveis de atributos e as alternativas – são apresentadas aos respondentes. Para o cálculo da quantidade de combinações possíveis com atributos com diferentes quantidades de níveis basta seguir a equação 3 (SANKO, 2001):

$$S^{ff} = \prod_{L=1}^L \prod_{n=1}^M L^n \quad (3)$$

Onde:

$S^{ff}$  = Número de combinações possíveis no desenho fatorial completo;

$L$  = Quantidade de níveis do atributo  $n$ ;

$n$  = Quantidade de atributos com  $L$  níveis;

$M$  = Quantidade de atributos envolvidos.

Utilizando a equação 3, é possível calcular a possibilidade de 27 combinações nos cenários dessa pesquisa, pois são 3 atributos de 3 níveis, como mostrado na equação 4:

$$S^{ff} = 3^3 = 27 \quad (4)$$

Hensher *et al.* (2005) explicam que, quando o desenho fatorial completo apresenta uma alta quantidade de combinações, essa alta quantidade pode reduzir a qualidade dos resultados, uma vez que uma grande quantidade de cenários apresentados a cada respondente pode gerar fadiga e, conseqüentemente, comprometer o comprometimento do respondente com a *survey*, afetando assim a qualidade das respostas. Nesse sentido, é importante que, caso não seja viável apresentar todos os cenários possíveis ao respondente por conta deste risco de fadiga e redução do comprometimento do respondente, seja reduzido a quantidade de cenários apresentados a cada respondente.

Alguns desenhos experimentais utilizam apenas uma fração dos desenhos fatoriais completos. Tais desenhos são conhecidos na literatura como desenhos fatoriais fracionados (HENSHER *et al.*, 2005) e tem como intenção representar o modelo fatorial completo com menos cenários, sem perder a capacidade de captar os efeitos não-lineares do modelo.

Dado o fato de que o desenho fatorial completo dessa pesquisa contém 27 cenários e que esse número parece ser suficientemente grande para causar fadiga e reduzir o comprometimento do respondente com a *survey*, optou-se inicialmente por tentar reduzir essa quantidade de cenários de 27 para 9, com um desenho fatorial fracionado elaborado por meio do Método dos Efeitos

Principais (HENSHER *et al.* p. 127, 2005) e auxílio do software IBM SPSS (SPSS INC, 2017). Todavia, após a aplicação da *survey* teste, que é comentada na subseção 3.2.5, ao tentar executar o modelo de regressão logística utilizando o software STATA (STACORP, 2013), o modelo não convergiu, mostrando que os coeficientes não eram confiáveis.

Um dos principais fatores que explicam a não conversão de um modelo é a insuficiência de casos em relação ao número de variáveis independentes incluídas nesse modelo, uma vez que o método de máxima verossimilhança, adotado pelo STATA (STACORP, 2013) para gerar o modelo de regressão logística, procura ajustar o modelo através de interações de repetições. Todavia, mesmo tentando replicar a amostra para aumentar a quantidade de casos e mitigar a questão da insuficiência de dados, o modelo continuou não convergindo e, por isso, foi rejeitado.

A solução encontrada foi, portanto, utilizar o desenho fatorial completo. Todavia, para resolver a questão da fadiga e descomprometimento com o questionário, cada um dos respondentes responderia a apenas um dos cenários. A seguir é comentado a aplicação da *survey* teste e suas implicações para, posteriormente, a aplicação da *survey*.

### 3.2.5 Aplicação da *survey* teste

Definidas as alternativas – PUP ou HD – os atributos envolvidos e seus níveis, além do desenho estatístico que foi utilizado para o instrumento de PD, o próximo passo, de acordo com a Figura 4 é a elaboração de como a *survey* teste será aplicada. Nesse sentido, levando em consideração que no momento do desenvolvimento desta pesquisa o mundo passava por uma crise sanitária devido à pandemia do Coronavírus, abordar pessoas presencialmente não era um ato recomendado pelos órgãos sanitários, dado o risco de contaminação. Por isso, optou-se por realizar uma coleta de dados de maneira virtual e a plataforma escolhida para a coleta das respostas foi a plataforma do Google Forms.

Nesta *survey* teste, foram apresentados aleatoriamente à cada indivíduo 3 cenários. Em cada um deles o indivíduo precisava fazer uma escolha se preferia receber o produto em casa – opção HD – ou ir buscar em um *pick-up point* – opção PUP. A aplicação do teste foi feita via plataforma Google Forms e as respostas tabuladas no software Microsoft Excel.

A intenção da *survey* teste era observar se haveria alguma dificuldade por parte dos respondentes em preencher o questionário. Todavia, não foi relatado por nenhum participante

dificuldade em preencher a *survey* e, por esse motivo, nenhuma alteração foi realizada no método, inicialmente.

### 3.2.6 Aplicação da *survey*

Cada *survey* fora composta por quatro etapas. A primeira etapa tinha a intenção de delimitar se os participantes teriam ou não suas respostas registradas para fins de análise dos dados. Com esse objetivo, duas perguntas eram feitas: uma para saber se o respondente concordava em participar da pesquisa e outra para saber se o respondente já havia comprado alguma vez online. Em ambos os casos, caso a resposta fosse negativa, o respondente era direcionado à página de agradecimento da *survey* e sua participação desconsiderada no banco de dados.

A segunda etapa teve como objetivo designar aleatoriamente um cenário para que o participante respondesse. Todavia, levando em consideração que se tratam de 27 cenários diferentes e que cada respondente responderia a apenas 1 desses cenários, fez-se necessário a criação de uma estratégia que fizesse com que o próprio respondente escolhesse o cenário que responderia, sem saber qual seria nem quais as características desse cenário. Para isso, criou-se uma relação de 27 números pseudoaleatórios, contendo 3 dígitos cada, que eram apresentados em ordem diferente à cada respondente que participava do experimento. Cada um desses 27 números estava atrelado a um cenário, cuja informação não era de conhecimento do respondente.

Com o intuito de equalizar a quantidade de respostas obtidas por cada um dos 27 cenários, optou-se por retirar do questionário os cenários, a medida em que esses fossem atingindo a marca de 15 respostas válidas. Esse método reduziu a quantidade de cenários que alguns respondentes puderem escolher, uma vez que, no momento em que participaram do questionário, alguns cenários já haviam sido retirados das possibilidades de escolha.

Nesse sentido, respondente escolhia um dos cenários disponíveis de maneira pseudoaleatória e, de acordo com o número escolhido, era direcionado para a página do cenário respectivo. A relação entre os 27 cenários e os números aleatórios é exemplificada no Quadro 4.

Na etapa 3, o respondente se deparava com a *survey* de Preferência Declarada. Era apresentado a ele uma situação hipotética em que esse participante realizava uma compra online. Após a leitura da situação hipotética, havia um quadro abaixo do texto, contendo um dos 27 cenários, preenchido com os níveis dos atributos Frete, Localização e Prazo daquele cenário específico. Ao final da leitura, o respondente precisava analisar e realizar uma escolha discreta entre receber o produto em casa – opção que representava o HD – ou ir buscar em um *pick-up point*

– opção que representava o PUP –. Todos os 27 cenários e a situação hipotética da *survey* de PD podem ser encontradas no APÊNDICE A.

Após responder qual era a sua preferência no determinado cenário, o respondente seguia para a etapa 4 da *survey*. Esta etapa tinha como objetivo fazer um levantamento da característica da amostra que participou do experimento. Todas as cinco questões do questionário sociodemográfico podem ser encontradas no APÊNDICE B.

A primeira questão tinha como objetivo levantar a característica da amostra do ponto de vista do comportamento dela com relação à inovação em serviços. Uma vez que o PUP ainda não está tão difundido no Brasil como outras formas de *delivery*, a intenção foi a de caracterizar a amostra respondente do ponto de vista da abertura dos indivíduos para a utilização serviços inovadores.

Quadro 4 – Relação de cenário versus número

Cenário	Número
1	201
2	192
3	983
4	404
5	635
6	226
7	007
8	018
9	809
10	800
11	651
12	502
13	093
14	874
15	195
16	596
17	277
18	058
19	149
20	200
21	901
22	552
23	333
24	484
25	315
26	206

27	907
----	-----

Fonte: elaborado pelo autor.

A segunda questão teve como intenção caracterizar a amostra respondente do ponto de vista de indivíduos que possuem ou não alguma restrição com relação ao HD. Nesse sentido, era apresentado ao respondente uma pergunta que questionava se ele já havia tido alguma experiência negativa com relação à entrega de um produto comprado em e-commerce. Essa pergunta foi inspirada no estudo de Rao *et al.* (2011) que, em seus resultados, mostravam que os indivíduos costumam reduzir a frequência e o tamanho de futuras compras online quando esses experienciavam alguma falha na entrega.

A próxima pergunta tinha como intenção caracterizar a amostra respondente com relação à sua idade. O objetivo da pergunta era caracterizar a amostra com relação à sua capacidade de compra e possível dificuldade com novos serviços ou tecnologias. Nesse sentido, foi definido que essa pergunta possuiria 3 opções: Menos que 21 anos, Entre 21 e 30 anos, Acima de 30 anos. Supõe-se que o indivíduo com menos de 21 anos ainda esteja em seu período de formação profissional e, portanto, com provável pequena renda própria. Entre 21 e 30 anos, o indivíduo já está na fase de amadurecimento profissional e, conseqüentemente, provável média renda. Por fim, acima dos 30 anos o indivíduo já está possivelmente com sua formação profissional consolidada e, conseqüentemente, consolidada renda individual.

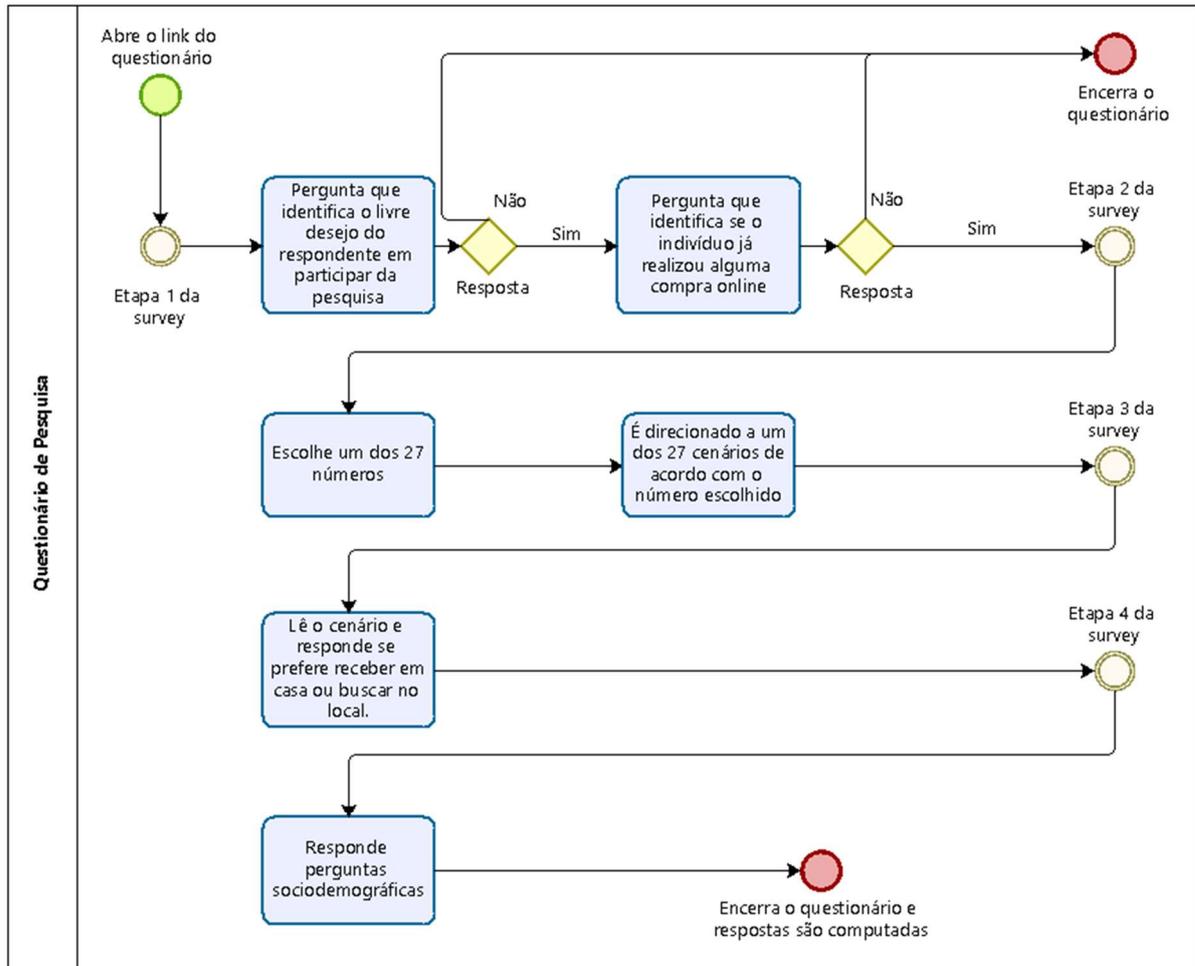
A quarta pergunta sociodemográfica tinha como intenção levantar se a amostra era composta por pessoas que já experimentaram algum tipo de serviço onde você compra o produto pela internet e busca na loja, não necessariamente o PUP. Para isso, questionou-se o respondente se ele já havia comprado alguma vez pela internet e optado por retirar o produto em loja.

Por fim, a última questão do questionário sociodemográfico tinha como intenção caracterizar a amostra com relação a quantidade de habitantes da cidade onde residem. O objetivo dessa pergunta reside na ideia de que as pessoas possuem diferentes percepções sobre distância. Os indivíduos que residem em grandes cidades, com alta taxa populacional por exemplo, tendem a fazer deslocamentos maiores em suas rotinas e, por isso, estariam mais pré-dispostos a percorrer 1 ou 2 quilômetros para buscar um produto, por exemplo. Por outro lado, indivíduos de cidades menores não tem esse mesmo hábito e, por isso, podem entender que andar 1 ou 2 quilômetros para buscar um item comprado é uma ação de muito esforço.

Nesse sentido, utilizando a classificação do Instituto Brasileiro de Geografia e Estatística (IBGE) para cidades pequenas, médias e grandes (SENADO FEDERAL, 2009), foi desenvolvida a quinta pergunta.

Ao final da Etapa 4 da *survey*, o respondente confirmava o envio das suas respostas e elas eram computadas em uma planilha fornecida pela própria plataforma do Google Forms. Na Figura 5, feita pelo software Bizagi, é possível conferir o fluxograma de como se deu o processo onde os respondentes faziam sua participação no questionário.

Figura 5 – Fluxograma de processo da survey



Powered by  
bizagi  
Modeler

Fonte: elaborado pelo autor.

Após a aplicação da *survey*, a próxima subseção trata do modelo de escolha discreta, que foi o modelo escolhido para identificar a preferência declarada dos respondentes, de modo a alimentar uma análise de regressão logística dos dados obtidos.

### 3.3 MODELO DE ESCOLHA DISCRETA E REGRESSÃO LOGÍSTICA

Para analisar quais situações o cliente opta pelo PUP ou pelo HD, esta pesquisa, inspirada nos estudos de Hsiao (2009), Twaddle (2011) e Silva *et al.* (2019), fez uso do modelo de escolha discreta, utilizando a regressão logística para compreender a probabilidade de um respondente escolher o PUP em detrimento ao HD como método de entrega. A seguir, é apresentado a técnica de regressão logística, o porquê de sua escolha como técnica de tratamento dos dados, bem como os pressupostos que essa técnica apresenta, o modo como ela foi utilizada nessa pesquisa e o tamanho da amostra mínima necessária para sua execução.

#### 3.3.1 Modelo de Regressão Logística

A Regressão Logística tem como objetivo gerar o mais otimizado modelo possível que descreva a relação entre uma variável resposta, seja ela dependente ou responsiva, e um conjunto de variáveis independentes, sejam elas preditoras ou exploratórias (HOSMER; LEMESHOW, 2000; HAIR *et al.*, 2009).

A utilização desta técnica justifica-se pelo fato dela possibilitar a utilização de variáveis categóricas como variáveis de saída (HOSMER; LEMESHOW, 2000). Isso significa que, dado o fato dessa pesquisa ter como objetivo estudar a relação dos atributos Preço do Frete, Distância e Prazo de Entrega com a preferência dos consumidores, em relação a dois itens categóricos, PUP ou HD, a Regressão Logística se mostra ajustada a essa meta, e, portanto, ideal para a mesma.

Além dessa necessidade de ter sua saída pautada em uma escolha discreta, ou seja, uma alternativa em detrimento à outra, a regressão logística requer ausência de multicolinearidade entre as variáveis independentes, bem como a inexistência de outliers nos resíduos (HOSMER; LEMESHOW, 2000).

Citado os pressupostos de uma regressão logística, faz-se necessário abordar o fato de que o modelo de regressão logística dessa pesquisa utilizou o modelo fatorial completo do desenho experimental.

Ao utilizar o desenho fatorial completo, alguns dos 27 cenários envolvidos são tão favoráveis a escolha de HD ou de PUP que acabaram por produzir outliers nos resíduos, o que foi identificado pela medida DFBETA ( $> 2/\sqrt{N}$ ) após análise dos dados da *survey* teste. Para verificar as saídas das variáveis DFBETA, veja APÊNDICE C.

Nesse sentido, a solução encontrada foi a de realizar uma regressão logística robusta por clusters, onde cada cenário foi tomado como um cluster composto por respostas de vários respondentes. Essa solução se alinha ao recomendado por Farcomeni e Ventura (2012), para os quais excluir os valores outliers pode ser uma solução simplista e uma regressão robusta acaba por lidar com mais eficiência com esse desvio, sem comprometer a generalização dos resultados. Acrescenta-se que a regressão logística foi processada no software STATA (STATACORP, 2013) (cf. script no APÊNDICE D).

### 3.3.2 Variáveis independentes e coeficientes logísticos

Um detalhe que Hosmer e Lemeshow (2000) explicam e que se aplica a essa pesquisa é dada a utilização de variáveis independentes nominais, é necessário transformar essas variáveis em variáveis *dummies*, uma vez que não é possível mensurar de maneira contínua essas variáveis nominais. O processo de transformação consiste em criar outras variáveis que assumam valores de 0 ou 1 e que passem a representar as variáveis dependentes. Nesse sentido, os autores comentam que a quantidade de variáveis necessárias para representar uma determinada variável independente é sempre uma unidade a menos do que a quantidade de categorias da determinada variável. Por exemplo, no caso dessa pesquisa, a variável PREÇO possui 3 categorias, as quais aqui são chamadas de níveis de atributos. Portanto, para representar a variável PREÇO, são necessárias duas variáveis *dummies*. As variáveis *dummy* criadas para essa pesquisa estão dispostas nos quadros 5, 6 e 7.

Quadro 5 – Variável dummy Preço

PREÇO	X1	X2
PUP e HD com preços iguais	0	0
PUP 15% mais barato	0	1
PUP 30% mais barato	1	0

Fonte: elaborado pelo autor.

Quadro 6 – Variável dummy Distância

DISTÂNCIA	X3	X4
O PUP está localizado entre 1 e 2 km da sua residência.	0	0
O PUP está localizado em até 1 km da sua residência.	0	1
O PUP está localizado ao lado da sua residência.	1	0

Fonte: elaborado pelo autor.

Quadro 7 – Variável dummy Prazo

PRAZO	X5	X6
PUP e HD com prazo de entrega iguais	0	0
Previsão de entrega 2 dias mais rápido.	0	1
Previsão de entrega 4 dias mais rápido.	1	0

Fonte: elaborado pelo autor.

Uma vez exposto a construção das variáveis *dummies*, parte-se para a estimação dos coeficientes logísticos desconhecidos ( $\beta_0, \beta_1, \beta_2, \dots, \beta_n$ ). Para essa tarefa, Hosmer e Lemeshow (2000) explicam que a regressão logística utiliza o método da máxima verossimilhança, que consiste na busca por valores para os coeficientes desconhecidos ( $\beta_0, \beta_1, \beta_2, \dots, \beta_n$ ) que maximizem a função de verossimilhança. O beta que maximiza a função de verossimilhança é aquele que iguale a zero a derivada da função e, portanto, é o beta determinado como coeficiente logístico na regressão.

Por fim, explicada a maneira como se estimou os coeficientes logísticos, é preciso entender como se dá a interpretação desses coeficientes de regressão, e esse tópico é abordado na próxima seção.

### 3.3.3 Interpretação dos coeficientes de regressão

Uma vez desenvolvido o modelo de regressão logística e seus coeficientes, é necessário interpretar esses coeficientes. No caso da regressão logística, a alteração de uma unidade da variável independente  $X_i$  vai alterar em  $\beta_i$  unidades a variável *logit* do modelo de regressão. Ou seja, cada incremento do valor de  $X_i$  vai incrementar  $\beta_i$  unidades no logaritmo da razão de desigualdade. Sendo a interpretação de uma unidade a mais de logaritmo uma tarefa difícil, o que se faz é um aplicar o antilogaritmo na função *Logit*, como expresso na equação 5 para o caso dessa pesquisa.

$$\text{Razão das desigualdades} = e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6} \quad (5)$$

No caso do modelo exponenciado, como exposto na equação 5, cada incremento de unidades nas variáveis  $X_i$  incrementam  $\beta_i$  unidades na razão de desigualdades, ficando assim uma maneira mais simples de observar a influência das variáveis independentes na razão de probabilidades. Levando em consideração que essa pesquisa precisou utilizar variáveis

independentes *dummies* no modelo de regressão logística, a interpretação desses coeficientes se dará de uma maneira um pouco diferente. Sendo o atributo original Preço expresso em termos de  $X_1$  e  $X_2$ , o atributo original Distância expresso por  $X_3$  e  $X_4$  e o atributo original Prazo expresso por  $X_5$  e  $X_6$ , como mostrado nos quadros 5, 6 e 7, a variação de 0 para 1 em cada uma dessas variáveis expressa a influência do grupo ao qual elas pertencem na razão de desigualdade. Por exemplo, quando a variável  $X_1$  passa de 0 para 1, temos uma influência de  $\beta_1$  unidades na razão de desigualdades da probabilidade de o consumidor escolher o PUP quando o preço do frete deixa de ser igual entre o PUP e o HD e passa a ser ter um desconto de 30% no valor caso o consumidor opte pelo PUP.

Na próxima seção é abordado o tamanho da amostra necessária para se executar a análise por meio da regressão logística.

### 3.3.4 Tamanho da amostra

O tamanho da amostra necessária para estudos de regressão logística foi determinado utilizando o estudo de Bujang *et al.* (2018) onde, segundo os autores, é possível determinar o tamanho da amostra por meio da equação 6:

$$n = 100 + 50i \quad (6)$$

Onde  $n$  representa o tamanho da amostra e  $i$  o número de variáveis independentes do modelo final.

Dado o fato de o modelo de regressão dessa pesquisa utilizar 6 variáveis *dummies*, logo, o tamanho da amostra mínima é de 400 respondentes.

Todavia, levando em consideração que o experimento contém 27 cenários, cada cenário forma um cluster que deve ter o mesmo número de observações. Nesse sentido, tendo como restrição que a amostra mínima definida por Bujang *et al.* (2018) na equação 6 foi de 400 respondentes, foi necessário acrescentar mais 5 respondentes para que todos os 27 cenários possuíssem a mesma quantidade de observações.

#### 4. ANÁLISE DOS RESULTADOS

Neste capítulo, é apresentado a análise dos resultados que foram obtidos a partir das saídas da técnica de regressão logística. Sua organização se dá inicialmente com um resumo sobre a aplicação da *survey* de PD, seguido de uma análise de caracterização da amostra participante. Na sequência, foi realizada uma análise descritiva da *survey* e, por fim, uma análise da regressão logística. As discussões desse capítulo são feitas a luz apenas da técnica utilizada, enquanto a discussão teórica dos resultados obtidos, a luz da TCT, é feita no capítulo 5.

A *survey* de PD foi aplicada entre os meses de Abril e Maio de 2021. Ao todo, participaram 422 pessoas respondendo, cada uma delas uma única vez. Como explicado no fluxograma apresentado na seção 3.2.6 pela Figura 5 foram excluídos da amostra 17 participantes que responderam “Não” para ao menos uma das perguntas eliminatórias. Portanto, foram levados em consideração 405 respondentes e a análise de suas resposta se deu em duas etapas. Todos os 405 respondentes completaram todas as etapas do questionário e, portanto, tiveram suas respostas computadas.

Em seguida, foi feita uma categorização da amostra com base nas respostas obtidas na quarta parte do questionário. Na sequência, é realizada a análise por regressão logística dos dados obtidos na segunda etapa do questionário, bem como a avaliação do modelo, dos coeficientes logísticos e das variáveis logísticas. Para tal análise, foi utilizado o software STATA (STATACORP, 2013) e o script utilizado pode ser conferido no APÊNDICE D.

A seguir é apresentado a etapa da análise de caracterização da amostra.

##### 4.1 CARACTERIZAÇÃO DA AMOSTRA

A quarta etapa do questionário fora composta por 5 perguntas que buscaram caracterizar a amostra. A primeira pergunta tinha como intenção observar se a amostra era formada por pessoas com alguma tendência a experimentar coisas novas, uma vez que o PUP é um serviço ainda novo no Brasil. A pergunta foi: “Quando um novo serviço é lançado, você costuma consumir ele:” e as opções foram: “Antes que outras pessoas conhecidas experimentem” ou “Após outras pessoas conhecidas experimentarem”. A Figura 6 mostra o gráfico que caracteriza a amostra nesse sentido.

Figura 6 – Caracterização da amostra inovação



Fonte: elaborado pelo autor.

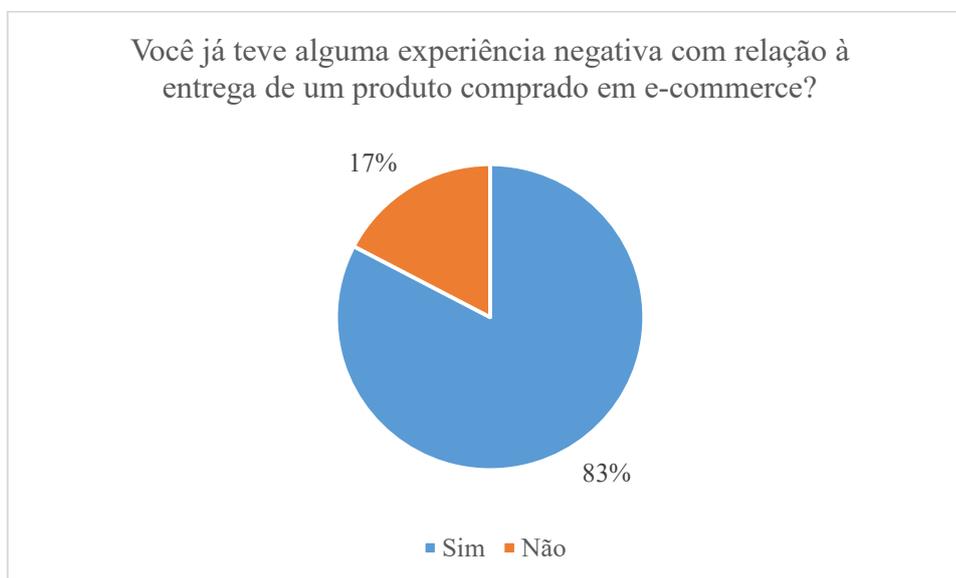
Os resultados observados na Figura 6 com relação a proporção de pessoas que experimentam um novo serviço antes de outras pessoas conhecidas e de pessoas que experimentam um novo serviço após outras pessoas conhecidas experimentarem mostra que a amostra segue uma proporção muito semelhante à aquela definida por Rogers (1995) em seu trabalho que caracterizou a inovatividade dos indivíduos em grupos com frequência de distribuição normal. No trabalho, o autor sugere que cerca de 16% dos indivíduos adotam uma inovação antes de outras pessoas – caracterizado por dois grupos: os Inovativos e os Primeiros a Adotar – e esse número se assemelha aos 23% verificados pela amostra dessa pesquisa. Isso significa que essa é uma amostra que, do ponto de vista da inovatividade, está de acordo com a população no geral.

A próxima pergunta dizia respeito a característica da amostra com relação a experiências com compras em e-commerce. Nesse sentido, a pergunta feita foi: “Você já teve alguma experiência negativa com relação à entrega de um produto comprado em e-commerce? Por exemplo: falha na entrega, atraso na entrega...” e as opções foram “Sim” ou “Não”. A Figura 7 mostra o gráfico que caracteriza a amostra nesse sentido.

Ao compararmos o perfil da amostra, apresentado na Figura 7, é possível fazer um comparativo com a satisfação geral do brasileiro com relação à entrega de um produto comprado no comércio eletrônico. Nesse sentido, analisando os resultados do relatório anual Webshoppers (2018), desenvolvido pela Ebit, e sendo um dos relatórios de maior credibilidade sobre o tema de e-commerce em território brasileiro, é possível observar que a amostra dessa pesquisa apresenta

semelhanças com relação ao percentual de consumidores que já experienciaram algum problema na entrega do produto.

Figura 7 – Caracterização da amostra experiências anteriores



Fonte: elaborado pelo autor.

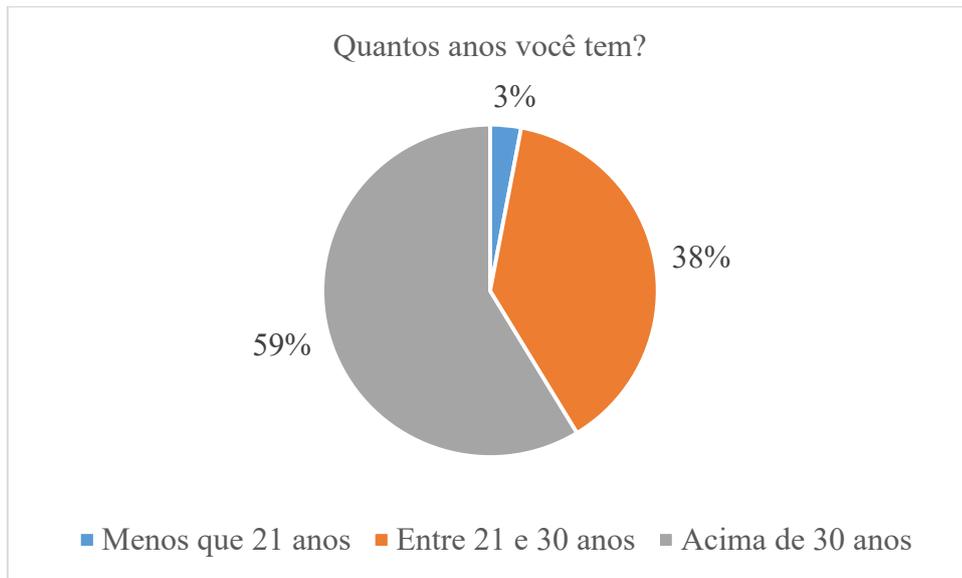
No relatório, por exemplo, o percentual de clientes satisfeitos com a entrega do produto foi de 14%. Isso significa que apenas 14% dos entrevistados não acredita ter tido uma experiência negativa com relação à entrega do produto, enquanto 86% afirmam ter tido uma experiência negativa. Esses números são muito semelhantes ao da amostra dessa pesquisa, o que significa que essa amostra se assemelha a população brasileira no que diz respeito a experiências anteriores com entrega de produtos adquiridos no comércio eletrônico.

Em seguida, caracterizou-se a amostra com base na idade dos respondentes. A categorização e sua frequência da idade dos respondentes podem ser observadas na Figura 8.

É possível fazer um comparativo da amostra dessa pesquisa, com relação à idade dos respondentes, e o relatório apresentado pela Neotrust (2021). Nesse relatório, a empresa apresenta dados com relação à idade dos consumidores de e-commerce do Brasil. Apesar da categorização utilizada pelo relatório ser diferente da utilizada por essa pesquisa, é possível tirar algumas conclusões em prol da caracterização da amostra dessa pesquisa.

Por exemplo, o relatório mostra que 19% dos consumidores tem menos de 25 anos, 33,1% tem entre 26 e 35 anos, 33,9% tem entre 36 e 50 anos e apenas 14% tem acima de 51 anos. Isso significa que a grande maioria dos consumidores está concentrada na faixa etária entre 25 e 50 anos, com um total de 67% dos respondentes. Ao ser comparado com os resultados dessa pesquisa, é possível observar que a maior parte da população está acima dos 30 anos, com 59%.

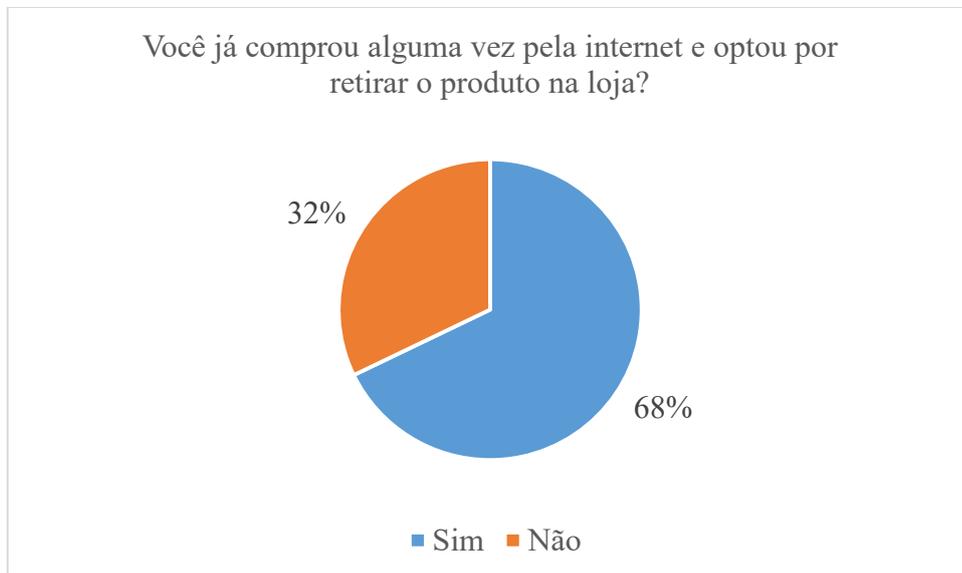
Figura 8 – Caracterização da amostra por idade



Fonte: elaborado pelo autor

Foi também caracterizada a amostra com base na experiência na utilização de serviços do tipo PUP. A caracterização pode ser observada pela Figura 9.

Figura 9 – Caracterização da amostra – utilização do PUP

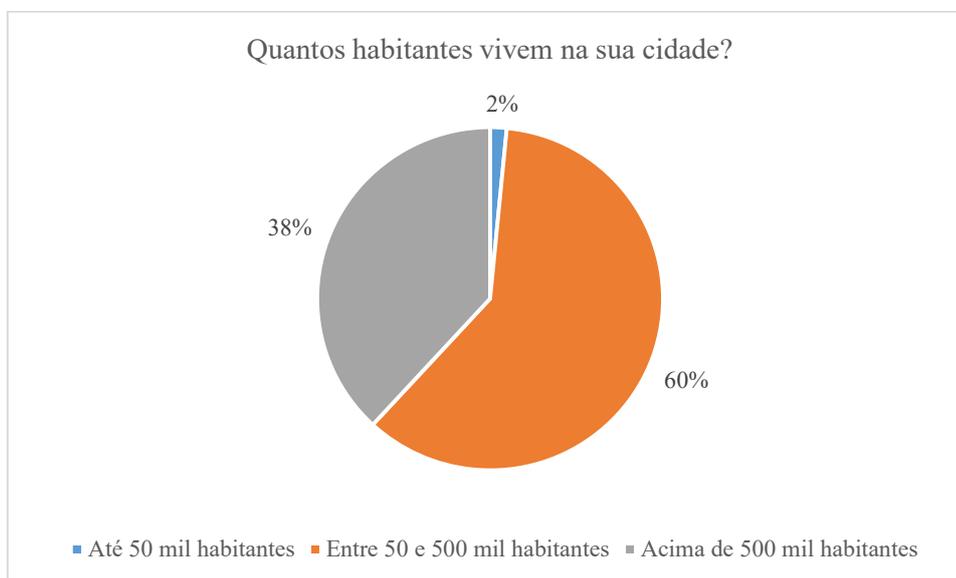


Fonte: elaborado pelo autor.

A partir da categorização observada na Figura 9, é possível fazer um comparativo com o relatório “Connected Shoppers” (SALESFORCE, 2019). O relatório mostra que cerca de 69% dos brasileiros que compraram online já optaram por retirar o produto na loja. Esse dado é muito semelhante aos 68% apresentados pela amostra dessa pesquisa, o que leva a conclusão de que essa amostra pode ser caracterizada muito semelhante à população brasileira, em termos de utilização do serviço de PUP em algum momento.

Por fim, a amostra foi caracterizada de acordo com o tamanho da cidade onde os respondentes moravam. É possível observar essa caracterização na Figura 10.

Figura 10 – Caracterização da amostra – tamanho da cidade



Fonte: elaborado pelo autor.

Ao ser comparado com a característica da população brasileira, a amostra dessa pesquisa encontra-se desbalanceada com relação a quantidade de respondentes que vivem em cidades com até 50 mil habitantes e entre 50 e 500 mil habitantes, a partir de dados do IBGE (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA, 2020). Todavia, essa amostra se mostra mais semelhante a regiões metropolitanas, como a região da Grande Vitória, onde a maioria das cidades inclusas tem entre 50 e 500 mil habitantes – Cariacica, Guarpari, Viana e Vitória –, apenas uma cidade abaixo de 50 mil habitantes – Fundão – e duas cidades com população acima de 500 mil habitantes – Serra e Vila Velha –, de acordo com os mesmos dados do IBGE (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA, 2020).

Dado o fato da dificuldade de execução da *survey* por conta das restrições impostas com relação ao distanciamento social, não foi possível obter uma amostra maior que fosse mais semelhante a todo o território brasileiro.

Caracterizada a amostra de respondentes, a próxima etapa é a análise dos resultados. Na próxima seção, são apresentados os resultados da análise de estatística descritiva, bem como os resultados da regressão logística e a avaliação do modelo logístico.

## 4.2 ANÁLISE DESCRITIVA DA SURVEY

Como explicado na seção de procedimentos metodológicos, exemplificado na Figura 4, o próximo passo é a análise dos resultados. Nessa seção é apresentada a análise estatística descritiva dos resultados das respostas obtidas pela *survey* de preferência declarada.

A primeira análise feita, com auxílio do software STATA (STATACORP, 2013) foi a sumarização da estatística descritiva. Como pode ser observado pela Tabela 2, ao serem apresentados um dos 27 cenários a cada um dos 405 respondentes, a frequência de indivíduos que responderam preferir PUP em detrimento ao HD foi de, aproximadamente, 69,87%. Essa variável de escolha é representada como R1 na Tabela 2.

Tabela 2 - Estatística descritiva

Variável	Nº de Observações	Média	Mínimo	Máximo
X1	405	0,3333333	0	1
X2	405	0,3333333	0	1
X3	405	0,3333333	0	1
X4	405	0,3333333	0	1
X5	405	0,3333333	0	1
X6	405	0,3333333	0	1
R1	405	0,6987654	0	1

Fonte: elaborado pelo autor.

De uma maneira geral, isso significa que, caso seleccionássemos um indivíduo aleatoriamente dentro dessa amostra de 405 respondentes, a chance desse indivíduo ter escolhido o PUP no seu cenário é de 69,87%, como mostra a distribuição de frequência na Tabela 3.

Tabela 3 - Distribuição de frequência

R1	Frequência	Percentual	Percentual acumulado
0	122	30,12	30,12
1	283	69,88	100,00
Total	405	100,00	

Fonte: elaborado pelo autor.

Feita a análise da estatística descritiva, o próximo passo é a realização da análise da *survey* por meio da regressão logística. A próxima seção trata dessa análise, bem como a avaliação do modelo de regressão logística desenvolvido.

### 4.3 ANÁLISE DA REGRESSÃO LOGÍSTICA

As análises de regressão logística realizadas foram feitas com base nas saídas obtidas pelo software STATA (STATA CORP, 2013) e foram baseadas nos pressupostos apresentados por Hosmer e Lemeshow (2000) e Hair *et al.* (2009). Para realizar essa análise, foi desenvolvido um *script*, que pode ser encontrado no APÊNDICE D. Após inserido esse *script* no *software*, geraram-se as saídas à medida em que essas foram demandadas. Todas essas saídas do *script* podem ser encontradas no APÊNDICE E.

A primeira análise feita foi a regressão logística binomial, já de maneira robusta e utilizando os 27 cenários como 27 clusters. Os resultados obtidos por meio do software STATA (STATA CORP, 2013), utilizando o modelo *logit*, são apresentados pela Tabela 4 e mostram os coeficientes beta –  $\beta_1, \beta_2, \beta_3, \dots, \beta_6$  – relacionados a cada uma das variáveis *dummies* –  $X_1, X_2, \dots, X_6$  – bem como o coeficiente  $\beta_0$ , representado pela variável “\_cons” e os clusters representados pela variável CENRIO.

Como pode ser observado pela Tabela 4, o p-valor, representado pelo “P> |z|”, dos coeficientes  $\beta_3$  e  $\beta_4$ , sugere que os mesmos foram estatisticamente não significantes, de acordo com o ponto de corte estabelecido por Hosmer e Lemeshow (2000) de valores acima de 0,05. Isso significa que há evidências de que a alteração do valor de 0 para 1 nas variáveis  $X_3$  ou  $X_4$  não influenciam de maneira estatisticamente significativa a tomada de decisão do consumidor pelo PUP com relação à variável *dummy* de referência.

Uma vez que o p-valor das variáveis  $X_3$  e  $X_4$  apresentaram evidências de que tais variáveis não são estatisticamente significantes, utilizou-se a técnica de *stepwise forward* para verificar a formação de um novo modelo e compará-lo com o modelo completo, visto até então. Essa técnica avalia a “importância” que cada variável tem para o modelo em termos da medida da significância estatística do coeficiente dessa variável, inserindo-as uma a uma no modelo (HOSMER; LEMESHOW, 2000).

Tabela 4 - Saídas da regressão logística

Variável	Coefficientes	P >  z
X1	$\beta_1 = 1,007323$	0,000
X2	$\beta_2 = 0,8723332$	0,000
X3	$\beta_3 = 0,2896366$	0,159
X4	$\beta_4 = 2,44e^{-16}$	1,000
X5	$\beta_5 = 1,520399$	0,000
X6	$\beta_6 = 1,424122$	0,000
Constante	$\beta_0 = -0,7434009$	0,000

Fonte: elaborado pelo autor.

Nesse sentido, com um intervalo de confiança de 95%, as variáveis não estatisticamente significantes no modelo completo foram as variáveis X<sub>3</sub> e X<sub>4</sub>, como de fato a evidência do p-valor sugeria. Os resultados do *stepwise forward* são apresentados na Tabela 5, onde é possível observar como se comportaram os parâmetros e os coeficientes.

Tabela 5 - Regressão logística stepwise forward

Variável	Coefficientes	P >  z	Limite inferior (95% intervalo de confiança)	Limite superior (95% intervalo de confiança)
X5	1,514689	0,000	1,15798	1,871397
X6	1,418698	0,000	0,9644086	1,872986
X1	1,003702	0,000	0,6477679	1,359636
X2	0,8691699	0,000	0,4219327	1,316407
Constante	-0,644669	0,000	-0,9256327	-0,3637054

Fonte: elaborado pelo autor.

O primeiro item a ser comentado é a função de utilidade, obtida após a execução da regressão logística com *stepwise forward*. Os coeficientes atrelados às variáveis independentes formam a nova função linear de utilidade, expressa pela equação 7.

$$Y = -0,644 + 1,003 \cdot x_1 + 0,869 \cdot x_2 + 1,514 \cdot x_5 + 1,418 \cdot x_6 \quad (7)$$

Como pode ser observado na equação 7, o modelo tem seu intercepto com sinal negativo. Isso significa que, se não houvessem nenhum dos atributos envolvidos no modelo, a intenção dos consumidores seria de optar pelo HD. Isso se mostra coerente com o que foi apresentado até aqui, uma vez que o modelo de HD oferece ao consumidor a comodidade de receber o produto em casa.

Além disso, o modelo se mostra coerente com a proposta de influência das variáveis, uma vez que todos os outros coeficientes apresentaram sinal positivo e isso segue a mesma lógica das variáveis *dummies* apresentadas pelos Quadros 5 e 7. Isso significa que, a cada inserção de uma unidade a mais, de maneira independente, em cada uma das variáveis do modelo, essa inserção influencia positivamente na opção do respondente pelo PUP.

Por exemplo, como mostrado o Quadro 5 de variáveis *dummy*, a alteração de 0 para 1 unidades na variável *dummy*  $X_1$  significa um comparativo entre os cenários em que o preço do frete sai de igual entre HD e PUP para os cenários onde o preço do frete é de 30% menor quando se opta pelo PUP. Teoricamente, pressupõe-se que, ao analisar individualmente essa variável, racionalmente o consumidor tende a optar pelo PUP e a lógica é seguida pela equação 7, onde o coeficiente que acompanha a variável  $X_1$ , o  $\beta_1$ , é um coeficiente positivo, mostrando que essa variável de fato influencia positivamente na escolha pelo PUP em detrimento ao HD.

Encontrado um novo modelo a partir da técnica de *stepwise forward*, Hosmer e Lemeshow (2000) explicam que é necessário realizar o teste de razão de verossimilhança para verificar se a retirada das duas variáveis – nesse caso,  $X_3$  e  $X_4$  – do modelo completo afetaria a qualidade do modelo. Entretanto, o teste de razão de verossimilhança não é possível de ser realizado para regressões robustas por cluster, uma vez que as observações individuais não são mais independentes e a verossimilhança não reflete mais isso. Por isso, optou-se por realizar o Teste de Wald, que também é sugerido por Hosmer e Lemeshow (2000) como uma alternativa.

O Teste de Wald testa se os parâmetros de interesse são simultaneamente iguais a zero. Se eles forem, isso sugere que removê-los do modelo não afeta significativamente a qualidade desse modelo. No caso dessa pesquisa, o interesse está em testar se a qualidade do modelo é afetada ao retirar as variáveis *dummies*  $X_3$  e  $X_4$ .

Ao executar o teste, os resultados mostraram que o qui-quadrado foi de 3,41 para 2 graus de liberdade e o p-valor associado a esse qui-quadrado foi de 0,1818, falhando, portanto, em rejeitar a hipótese nula, o que comprova que a retirada de  $X_3$  e  $X_4$  não afeta a qualidade do modelo. As saídas do Teste de Wald podem ser conferidas no APÊNDICE E.

Como pode ser observado na Tabela 5, a análise do intervalo de confiança não permite afirmar que uma variável seja maior que a outra. Nesse sentido, a solução encontrada foi a Para criar uma classificação de cenários com maior probabilidade de escolha por parte do consumidor, utilizando a equação 8, utilizando dados da Tabela 5 de regressão logística.

$$P(1) = \frac{1}{1 + e^{-[\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_5 X_5 + \beta_6 X_6]}} \quad (8)$$

Onde:

$P(1)$  = Probabilidade do evento 1 ocorrer – no caso, o evento escolha PUP.

Essa equação foi retirada de Hosmer e Lemeshow (2000) e, a partir dela, foi possível calcular a probabilidade do evento 1 ocorrer em todos os vinte e sete cenários. Com esse número em mãos, foi montada a Tabela 6 que mostra a influência da alteração de cada variável *dummy* de 0 para 1, em comparação com o cenário referência de cada atributo, quando todas as outras variáveis *dummy* são iguais a zero.

Tabela 5 - Percentual de influência das variáveis dummy

Cenário	Probabilidade de optar pelo PUP	Diferença percentual para o cenário referência	Diferença percentual interna entre atributos
Cenário referência			
Frete do PUP e HD tem o mesmo valor Prazo de entrega para o PUP e o HD é o mesmo (X1 = 0; X2 = 0; X5 = 0; X6 = 0)	34,42%	0,00%	-
Frete no PUP é 15% mais barato (X1 = 0; X2 = 1; X5 = 0; X6 = 0)	55,59%	61,50%	-
Frete no PUP é 30% mais barato (X1 = 1; X2 = 0; X5 = 0; X6 = 0)	58,88%	71,06%	15,54%
Prazo 2 dias a menos pelo PUP (X1 = 0; X2 = 0; X5 = 0; X6 = 1)	68,44%	98,84%	-
Prazo 4 dias a menos pelo PUP (X1 = 0; X2 = 0; X5 = 1; X6 = 0)	70,47%	104,74%	5,97%

Fonte: elaborado pelo autor.

Como pode ser observado, quando todas as outras variáveis são nulas, a variável *dummy* que mais surtiu efeito positivo na probabilidade de o respondente optar pelo PUP foi a variável  $X_5$ , que corresponde ao prazo de entrega como sendo de 2 dias mais rápido via PUP, o que causou um aumento percentual de 104,74% na probabilidade de escolha pelo PUP, quando comparado ao cenário referência.

Entretanto, é possível observar que, a influência do frete 30% mais barato no PUP é 15,54% maior do que a influência do frete 15% mais barato no PUP na escolha do respondente pelo PUP. Enquanto isso, a influência do prazo de 4 dias a menos pelo PUP é de apenas 5,97% maior do que a influência do prazo com 2 dias a menos pelo PUP. Isso significa que a variação percentual de influência que as variáveis *dummies* relacionadas ao Preço do Frete tem sobre a intenção do respondente em utilizar o PUP foi maior do que o percentual de influência entre as variáveis *dummies* relacionadas ao Prazo de Entrega.

Em regressão logística, a variável dependente segue uma distribuição de Bernoulli e, por isso, é preciso realizar uma transformação de modo a conectar as variáveis independentes à essa distribuição. Essa transformação é denominada por transformação logit e ocorre igualando a função da regressão ao logaritmo natural das *odds ratio*, ou razão de probabilidade. Essas razões de probabilidade são a razão entre a probabilidade de um evento ocorrer pela probabilidade de ele não ocorrer. No intuito de se obter o valor dessas razões de probabilidade, a manipulação matemática realizada é a exponenciação da equação linear, de modo a gerar o antilogaritmo do logaritmo das razões de probabilidade (HOSMER; LEMESHOW, 2000). As saídas geradas pelo software STATA (STATACORP, 2013) podem ser observadas na Tabela 7.

Tabela 6 - Regressão Logística das Razões de Probabilidade

Variável	Razão de Probabilidade	P >  z
X5	4,548005	0,000
X6	4,131735	0,000
X1	2,728364	0,000
X2	2,38493	0,000
Constante	0,5248362	0,000

Fonte: elaborado pelo autor.

Observando as saídas da Tabela 7, a equação 9 é construída e representa a relação de razão de probabilidades com as variáveis independentes.

$$Odds\ ratio = 0,5248 + 2,728 \cdot x_1 + 2,385 \cdot x_2 + 4,548 \cdot x_5 + 4,132 \cdot x_6 \quad (9)$$

A importância de gerar essas razões de probabilidades se dá pelo fato de que ela possibilita conhecer a influência que determinado cenário tem sobre a razão de desigualdades. Peguemos como exemplo a interpretação da variável  $X_1$ . O acréscimo de uma unidade dessa variável, de acordo com o Quadro 5 de variáveis *dummies*, representa o PUP como opção que oferece frete

30% mais barato do que o HD. Se todas as outras variáveis permanecerem constantes e iguais a zero, o que se tem é a equação 10.

$$Odds\ ratio = 0,5248 + 2,728.1 + 2,385.0 + 4,548.0 + 4,132.0$$

$$Odds\ ratio = 0,5248 + 2,728$$

$$Odds\ ratio = 3,2528 \quad (10)$$

Isso significa que, no cenário onde a opção frete é 30% mais barato pelo PUP do que pelo HD, e todas as outras variáveis são nulas, a razão de desigualdade é 3,2528 vezes maior para o consumidor optar pelo PUP em detrimento ao HD.

A fins de comparação do modelo robusto completo com o modelo final, desenvolvido a partir da técnica de *stepwise forward*, verificou-se o erro de predição de ambos os modelos. Com esse intuito, utilizou-se o critério de informação de Akaike (AIC), como sugerido por McElreath (2016, p. 189). O AIC é um estimador de erro de predição e, portanto, da qualidade relativa dos modelos estatísticos para um determinado conjunto de dados. Dado um conjunto de modelos, o AIC determina a qualidade de cada um deles em relação a seus pares escolhendo aquele que minimiza o erro quadrático médio de predição ou estimativa. Os valores de AIC do modelo completo e do modelo final podem ser observados na Tabela 8.

Outro critério de seleção de modelo utilizado foi o critério de informação Bayesiano (BIC), também sugerido por McElreath (2016, p. 192). Ela é uma razão de verossimilhanças médias, utilizadas para comparação entre modelos. Neste caso, o valor BIC encontrado para o modelo completo e para o modelo final podem ser observados na Tabela 8.

Tabela 7 - AIC e BIC modelo final

<b>Modelo</b>	<b>AIC</b>	<b>BIC</b>
Completo	454,4157	482,4429
Final	451,7584	471,7779

Fonte: elaborado pelo autor.

De acordo com McElreath (2016), o critério que se utiliza para analisar a melhoria ou não de um modelo é se o AIC e o BIC foram reduzidos. Ao observar a Tabela 8 que traz os resultados do AIC e do BIC para o modelo completo, vemos que o AIC foi de 454,4157 e o BIC de 482,4429. Para o modelo final, tivemos um AIC de 451,7584, que já mostra ser melhor do que o AIC do modelo completo e um BIC de 471,7779, que também mostra ser um valor melhor

do que o do modelo completo. Isso significa que o modelo final possui maior qualidade e simplicidade.

Para aumentar a robustez do modelo final, realizou-se também o teste de Hosmer-Lemeshow (HOSMER; LEMESHOW, 2000). Esse teste tem como intenção observar o quão ajustado os dados estão ao modelo ao dividir a amostra em grupos de N indivíduos e observar se a predição do modelo está ajustada ao que foi observado. A Tabela 9 mostra os resultados desse teste. De acordo com os autores, a hipótese nula desse teste é a de que o modelo está bem ajustado aos dados observados. Ressalta-se que o teste falhou em rejeitar a hipótese nula, uma vez que seu p-valor foi de 0.866, o que significa que o modelo está bem ajustado aos dados.

Tabela 8 - Teste de Hosmer-Lemeshow

<b>Grupo</b>	<b>Probabilidade</b>	<b>Quantidade de respostas 1 no grupo</b>	<b>Quantidade de respostas 1 esperadas no grupo</b>	<b>Quantidade de respostas 0 no grupo</b>	<b>Quantidade de respostas 0 esperadas no grupo</b>	<b>Total</b>
1	0,3442	14	15,5	31	29,5	45
2	0,5559	25	25	20	20	45
3	0,5888	28	26,5	17	18,5	45
4	0,6844	34	30,8	11	14,2	45
5	0,7047	30	31,7	15	13,3	45
6	0,8380	36	37,7	9	7,3	45
7	0,8506	40	38,3	5	6,7	45
8	0,8554	37	38,5	8	6,5	45
10	0,8669	39	39	6	6	45

Fonte: elaborado pelo autor.

A próxima avaliação do modelo feita foi a análise da matriz de classificação, que pode ser observada na Tabela 10. A primeira análise feita foi a taxa de classificação correta geral do modelo. Ela representa, de acordo com Hosmer e Lemeshow (2000), a taxa de classificação correta que o modelo apresenta. Ou seja, representa a taxa com a qual o modelo classifica corretamente como 1 (opção PUP) ou 0 (opção HD). Como pode ser observado, o modelo apresentou uma taxa de classificação correta geral de 64,44%.

A próxima análise feita foi a de sensibilidade do modelo. De acordo com o Hosmer e Lemeshow (2000), nessa pesquisa a sensibilidade é a taxa que representa a razão entre a quantidade de predições corretas do modelo para indivíduos que escolhem a opção 1 (opção PUP) e a quantidade observada de indivíduos da amostra que escolheram a opção 1 (opção PUP). Nesse caso, o modelo apresentou uma sensibilidade de 64,31%, uma vez que classificou 182 como optantes pelo PUP, mas a amostra possuía 283 optantes do PUP na realidade. Isso significa que o modelo é menos eficiente do que o critério de chance estabelecido, que é de 69,88%.

Uma possibilidade de explicação para esse fato reside no desbalanceamento da amostra, que apresentou uma quantidade de aproximadamente 70% dos entrevistados optantes do PUP e apenas 30% optantes do HD. Sugere-se que em estudos futuros seja utilizado uma amostra melhor balanceada para comparação dos resultados e verificação da melhoria ou não da capacidade preditiva e sensível do modelo.

Por outro lado, uma análise que pode ser feita é que, apesar do modelo ter se mostrado menos eficiente do que o critério de chance estabelecido, ele se mostrou muito eficiente para classificar os indivíduos que optam pelo PUP. A capacidade preditiva nesse caso foi de 80,89%, uma vez que o modelo classificou corretamente 182 de um total de 225 que foram classificados como optantes pelo PUP.

Tabela 9 – Matriz de classificação

Classificado	Observado		Total
	Opção 1 – PUP	Opção 0 – HD	
Opção 1 – PUP	182	43	225
Opção 0 – HD	101	79	180
Total	283	122	405

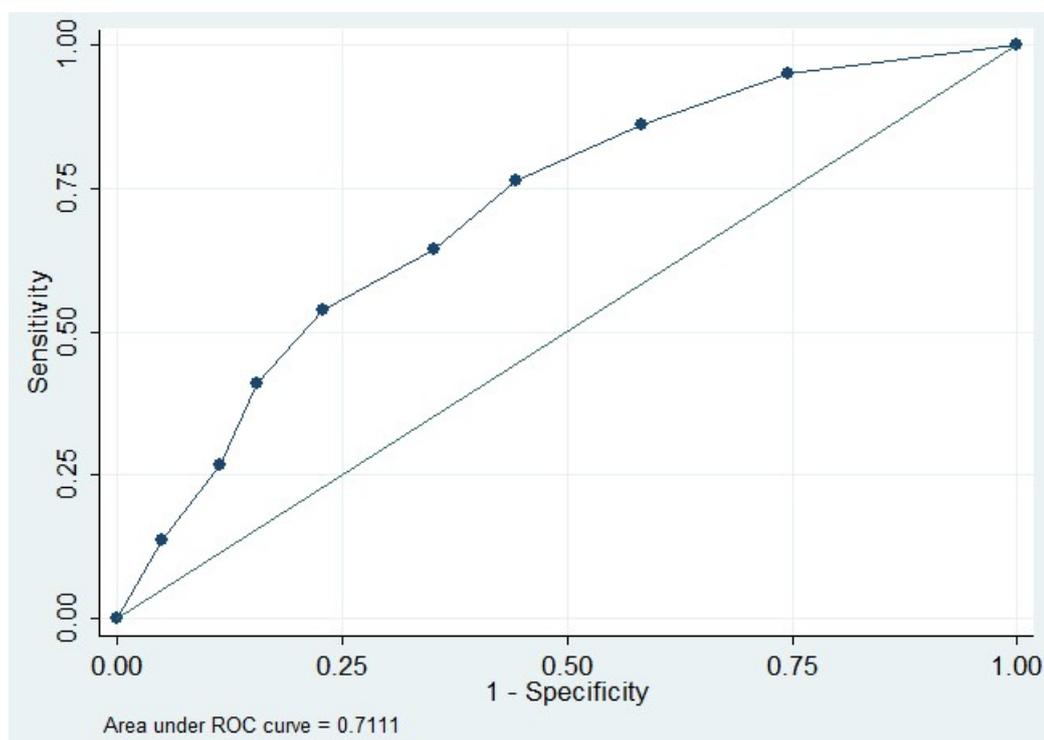
Fonte: elaborado pelo autor.

Uma vez levantadas a matriz de classificação do modelo final, a próxima avaliação feita foi a Curva de Características Operacionais do Receptor, a Curva ROC. De acordo com Hosmer e Lemeshow (2000), a Curva ROC mensura a capacidade preditiva do modelo proposto. Para isso, ela utiliza a relação entre a taxa dos verdadeiros positivos do modelo – valores 1 que o modelo classificou corretamente como 1 – e a taxa dos falsos positivos preditos – valores 1 que o modelo classificou erroneamente como 0 – e gera um gráfico a partir de uma série de

classificações feitas utilizando o modelo proposto. A curva ROC do modelo final é observada na Figura 11.

Como pode ser observado, a área abaixo da curva ROC foi estabelecida, pelo software STATA (STATA CORP, 2013), como sendo de 0,7111. Hosmer e Lemeshow (2000) estabelecem que valores acima de 0,7 são aceitáveis e, por isso, o modelo final pode ser classificado como tendo uma boa capacidade preditiva.

Figura 11 – Curva ROC



Fonte: elaborado pelo autor.

Até aqui, as análises foram focadas na observação de potenciais problemas que o modelo final proposto poderia apresentar. Entretanto, como abordado por Hosmer e Lemeshow (2000), faz-se necessário também investigar e detectar potenciais observações da amostra que tenham impacto significativo no modelo.

Para realizar essa investigação, essa pesquisa utilizou o cálculo do *dfbeta* de Pregibon (HOSMER; LEMESHOW, 2000) pelo software STATA (STATA CORP, 2013). Ao analisar a saída do *dfbeta* de Pregibon, é possível ver que existem *outliers* nas observações, uma vez que o ponto de corte estabelecido é de 1 para amostras pequenas e médias, como proposto por Kutner *et al.* (p. 405, 2005), e existem pontos que atingem o pico máximo de 4,0351. A saída do *dfbeta* de Pregibon pode ser conferida no APÊNDICE C.

A presença de *outliers* reforça a necessidade de ter sido utilizado a técnica de regressão logística robusta por clusters, pois a sua utilização possibilita que seja possível trabalhar a regressão, mesmo com os *outliers* deflagrados pelos valores de *dfbeta*, uma vez que a regressão robusta utiliza os cenários como clusters e reduzem a influência das observações de maneira individualizada no modelo.

No intuito de verificar se o modelo final robusto clusterizado atende aos pressupostos básicos da regressão logística, fez-se a análise do erro de especificação. Para isso, foi aplicado o *linktest* no software STATA (STATACORP, 2013). O objetivo do *linktest* é que, caso o modelo esteja devidamente especificado, não haverá nenhum preditor adicional que seja estatisticamente significativo, salvo em situações de acaso. O *linktest* cria duas variáveis novas, a variável “hat” e a variável “hatsq”. A primeira delas representa uma variável de previsão e a segunda seria o quadrado dessa variável de previsão para a reconstrução do modelo. A intenção é que “hat” seja estatisticamente significativo, observando seu p-valor, e “hatsq” não seja. Caso “hatsq” também seja estatisticamente significativo, então *linktest* também é significativo e isso significa que existem variáveis preditoras significantes que não foram adicionadas ao modelo. Na Tabela 11 é possível observar as saídas obtidas pelo STATA (STATACORP, 2013).

Tabela 10 – Saídas do erro de especificação

Variável	Coefficiente	P >  z
Hat	1,172593	0,000
Hatsq	-0,1250523	0,484
Constante	0,0217869	0,894

Fonte: elaborado pelo autor.

Como pode ser observado, a variável “\_hat” foi considerada significativa estatisticamente e a variável “\_hatsq”, em contrapartida, não, utilizando um ponto de corte no p-valor de 0,05. Isso significa que o modelo final é corretamente especificado.

Outra análise feita foi a de validação cruzada, no intuito de validar que os resultados encontrados têm validade externa e interna. Para fazer isso, Hair *et al.* (2009) comentam que a técnica mais utilizada para validação externa é a avaliação da razão de sucessos. Nesse sentido, essa pesquisa utilizou a validação cruzada, por meio do método *jackknife*, replicando com base nos 27 clusters. As saídas obtidas pelo software STATA (STATACORP, 2013) podem ser observadas na Tabela 12.

Tabela 11 – Saídas jackknife

Variável	Coefficiente Jackknife	P >  z
X1	1,003702	0,000
X2	0,8691699	0,004
X5	1,514689	0,000
X6	1,418698	0,000
Constante	-0,644669	0,001

Fonte: elaborado pelo autor.

Como pode ser observado, mesmo após a validação cruzada dos resultados por meio da técnica de *jackknife*, não houve diferenças significativas nos coeficientes das variáveis, nem no p-valor das mesmas. As únicas variações percebidas foram no p-valor da variável *dummy* X2 – passando de 0,000 para 0,004 – e no p-valor da constante – passando de 0,000 para 0,001 –. Isso reforça a ideia de que o modelo final possui validade externa.

Feitas as análises estatísticas com auxílio da regressão logística da amostra de 405 respondentes, partiu-se para a discussão dos resultados. No próximo capítulo é apresentada a discussão, bem como as implicações teóricas e práticas dessa pesquisa para a academia.

## 5. DISCUSSÃO TEÓRICA DOS RESULTADOS E IMPLICAÇÕES

O objetivo dessa pesquisa foi o de verificar dois pontos referentes a preferência do consumidor pelo PUP em detrimento ao HD. O primeiro deles foi o de analisar a sensibilidade do consumidor a alterações internas nos níveis desses atributos. O segundo ponto foi a identificação da ordem de influência que esses fatores exercem na preferência do consumidor pelo PUP em detrimento ao HD.

Para isso, como observado, realizou-se um desenho experimental a partir da técnica de Preferência Declarada utilizando como ferramenta uma *survey* aplicada a uma amostra de 405 respondentes e analisada estatisticamente por meio do método de Regressão Logística.

Uma primeira evidência encontrada por essa pesquisa está relacionada aos três fatores determinantes na preferência do consumidor, citados na literatura: Preço do Frete, Distância para o PUP e Prazo de Entrega. Após a realização da regressão logística robusta por clusters, as saídas dessa regressão mostram que as variáveis *dummy*  $X_3$  e  $X_4$ , que representavam o atributo Localização, eram estatisticamente não significantes. Ou seja, esse achado é uma evidência de que apenas os atributos Preço do Frete e Prazo de Entrega influenciam de maneira significativa a opção do consumidor pelo PUP em detrimento ao HD.

Por fim, ressalta-se que o atributo Localização foi considerado pelo modelo logístico, desenvolvido a partir de dados obtidos utilizando a técnica de PD, como não estatisticamente significativo no processo de escolha do consumidor pelo PUP em detrimento ao HD. Isso significa que o atributo Localização, nessa situação hipotética de preferência declarada do consumidor, é um atributo que não influencia significativamente o consumidor a optar pelo PUP em detrimento ao HD. Nesse sentido, observa-se que não é possível fazer a mesma afirmação em casos onde o consumidor realmente experimentou esses atributos. Portanto, sugere-se para estudos futuros que seja realizado um estudo de preferência revelada com essas mesmas variáveis para dar continuidade a essa discussão.

Para responder o primeiro ponto da questão de pesquisa, foi feita uma análise da Tabela 5, onde é possível encontrar as saídas da regressão logística robusta realizada após a técnica de *stepwise forward*. Ali, é possível verificar, por meio da dimensão dos coeficientes e de suas variações dentro do intervalo de confiança, que as variáveis *dummies* relacionadas ao Prazo de Entrega, variáveis  $X_5$  e  $X_6$ , possuem um peso maior na escolha do consumidor pelo PUP em detrimento ao HD do que as variáveis *dummies* relacionadas ao Preço do Frete,  $X_1$  e  $X_2$ .

Nesse sentido, é possível analisar que o atributo que apresentou maior influência sobre a escolha do consumidor pelo PUP em detrimento ao HD foi o Prazo de Entrega. Isso pode ser explicado pelo fato de ter sido incluso no texto do cenário hipotético do experimento um certo senso de urgência ao expressar que o participante precisaria comprar um presente para algum conhecido importante, gerando assim um aumento do custo da transação, uma vez que a especificidade temporal do ativo (WILLIAMSON, 1985) fez com que os participantes valorizassem escolhas que reduzissem esse risco de atrasar a entrega do presente, gerando assim uma percepção de maior utilidade sobre o PUP quando esse reduzia em 2 ou 4 dias o prazo de entrega.

Partindo da ótica da TCT, a validação do Preço do Frete como atributo estatisticamente significativo no processo de escolha do indivíduo pelo PUP em detrimento ao HD faz todo o sentido, uma vez que valores menores de frete mitigam a percepção de risco por parte do consumidor. Ou seja, ao se deparar com um valor reduzido, o indivíduo tende a equalizar essa escolha, de modo que tanto o comportamento oportunista como a racionalidade limitada, citados por Williamson (1985), sejam mitigados pela vantagem de se ter um valor de frete menor. Entretanto, neste caso, a influência dele foi menor do que a influência do Prazo de Entrega.

Levando em consideração que o cenário apresentado aos respondentes foi de que eles teriam carro e que poderiam utilizar dele para buscar o produto no PUP, supõe-se que esse fato pode ter influenciado a percepção de distância dos participantes e, conseqüentemente, a percepção de utilidade deles com relação às opções onde seria preciso realizar um deslocamento maior ou menor para chegar ao PUP. Sugere-se aqui que estudos futuros busquem analisar a influência de o consumidor ter ou não um carro como variável mediadora na opção pelo PUP em detrimento ao HD, de modo a complementar os achados dessa pesquisa.

Aqui, o que se tem é a possibilidade de que as distâncias escolhidas como níveis de atributo da variável Distância não foram suficientemente distantes a ponto de influenciar a percepção de utilidade do respondente para influenciar na escolha dele entre HD ou PUP. Entretanto, é preciso reforçar que, de acordo com os pressupostos do PUP, o modelo conceitual do PUP pressupõe que ele seja um local físico que fique próximo dos clientes (FRANCKE; VISSER, 2015). Ou seja, o conceito é de que seja um ponto físico que esteja na rota habitual do consumidor, de modo que esse não precise se deslocar grandes distâncias para conseguir buscar o item comprado. Por isso, não faria sentido adotar distâncias maiores do que 1 ou 2 km para abordar o atributo Localização.

O segundo ponto respondido por essa pesquisa está relacionado a evidência de que o consumidor é mais sensível a alterações no preço do frete do que no prazo de entrega. Isso porque, de acordo com a Tabela 6, o aumento percentual da intenção do consumidor em optar pelo PUP, no cenário onde o preço do frete é 30% mais barato no PUP, é de 15,54%, quando comparado à influência do cenário onde o PUP tem preço de frete 15% mais barato. Por outro lado, a influência do cenário onde o prazo de entrega é 4 dias a menos na opção PUP é apenas 5,97% maior do que o cenário onde o prazo de entrega é de 2 dias a menos via PUP.

Na próxima subseção são abordadas as contribuições teóricas e práticas dessa pesquisa para o campo de estudo em questão.

### 5.1 CONTRIBUIÇÕES TEÓRICAS

Os pesquisadores do LM já definiram que as estratégias utilizadas pelos varejistas na operação logística podem definir a satisfação do consumidor pelo processo de compra (ZEITHAML, 2000; FENG; HUANG, 2005; BOYER; HULT, 2005; RAO *et al.*, 2011; VAKULENKO *et al.*, 2019). Além disso, outros autores já definiram o LM como o processo mais custoso e menos eficiente da cadeia logística do e-commerce (GEVAERS *et al.*, 2014; ZHANG *et al.*, 2018; SILVA *et al.*, 2019). Nesse sentido, esse estudo contribui para a literatura de LM na gestão de operações logísticas abordando e explicitando dois atributos como os que tem peso na tomada de decisão do cliente que opta pelo PUP em detrimento ao HD, sendo eles o prazo de entrega e o preço do frete, nessa ordem.

Com relação ao atributo Prazo de Entrega, esse achado vai ao encontro do que pressupunham Nguyen *et al.* (2018), Chen *et al.* (2018) e Le e Ukkusuri (2019), de que o custo da transação ao optar pelo PUP seria menor em situações onde o PUP reduzisse o prazo de entrega, se tornando uma escolha atraente para os consumidores. Por outro lado, a validação do atributo Prazo de Entrega vai em contraponto ao trabalho de Chen e Wei (2018), que concluíram que a pressão do tempo para a entrega não afetou a intenção de utilizar um outro modelo de delivery diferente do tradicional HD.

Quando se observa a literatura, essa pesquisa vai ao encontro a validação de que o prazo de entrega e do preço do frete são atributos significativos na escolha do consumidor. A validação desses atributos reforça outros estudos, como os de Silva (2018), Nguyen *et al.* (2018), Nguyen *et al.* (2019) e Yuen *et al.* (2019).

Além disso, esse estudo gera uma evidência com relação à ordem de influência apresentada pelo modelo logístico desenvolvido. Diferentemente dos estudos de Silva (2018) e Nguyen *et al.* (2019), nessa pesquisa a variável mais influente na tomada de decisão pelo PUP em detrimento ao HD não foi o preço do frete, mas sim o prazo de entrega. Como dito anteriormente, esse fato pode ser explicado pela utilização de um possível senso de urgência causado pela construção do cenário de pesquisa, criando assim uma percepção de utilidade maior por parte do respondente com relação ao prazo de entrega.

Outra evidência abordada por essa pesquisa é a de que o atributo Localização não se mostrou como estatisticamente significativa na escolha do consumidor pelo PUP em detrimento ao HD. Isso vai de encontro ao estudo desenvolvido por Chen e Wei (2018), que afirmaram que a localização do PUP pode influenciar positivamente na intenção de uso do consumidor por esse serviço.

Por fim, destaca-se que essa pesquisa oferece evidências de quais atributos influenciam a tomada de decisão do consumidor pelo PUP em detrimento ao HD após o início da pandemia de covid-19. O estudo vai de encontro a um dos achados de Suguna *et al.* (2021), que abordam o prazo de entrega como um dos atributos importantes para os consumidores no LM após o início da pandemia.

## 5.2 CONTRIBUIÇÕES PRÁTICAS

Como contribuições práticas, partindo do pressuposto do estudo de Rao *et al.* (2011), que abordavam a influência negativa que uma falha ocorrida no LM pode ter na intenção de compra do consumidor, essa pesquisa oferece evidências de quais atributos devem ser observados com maior atenção pelas organizações ao montarem suas estratégias logísticas no LM, de modo a gerar uma maior percepção de valor a luz do consumidor.

Os resultados dessa pesquisa mostram que os atributos Preço do Frete e Prazo de Entrega foram estatisticamente significantes. Nesse sentido, os gestores do setor de logística podem utilizar dessas evidências para montar suas estratégias de *omnichannel*, de modo a focar, por exemplo, em gestão de entregas e frotas para reduzir o prazo de entrega, ou adoção de políticas como Frete Grátis, tida como na literatura como uma das estratégias mais eficazes para reduzir a percepção de risco do consumidor (LEWIS *et al.*, 2006; SHAMPANIER *et al.*, 2007; MA; 2017).

Como discutido pelos resultados obtidos por essa pesquisa, o varejista precisa estar atento ao senso de urgência que o consumidor pode ter quando adquirir um produto. Isso é, o varejista precisa estar atento a não só ofertar um preço de frete competitivo e atrativo, como sugerido majoritariamente por outras literaturas (e.g. SILVA, 2018; NGUYEN *et al.*, 2019), como também abordar o prazo de entrega como elemento essencial para atingir a qualidade de um processo de compra completo.

Por fim, essa pesquisa contribui, de maneira prática, com uma evidência de que a distância necessária a ser percorrida pelo consumidor até o PUP, quando este está em até 2km de distância do mesmo, não influencia significativamente a tomada de decisão do consumidor pelo PUP em detrimento ao HD. Isso significa, se o gestor de logística de um varejista levar em consideração essa evidência, não fará diferença estatisticamente significativa se o PUP estiver ao lado da residência do consumidor, a 1 km de distância ou a 2 km de distância. A percepção de valor, nesse caso, por parte do consumidor, não é significativa a ponto de influenciar a escolha dele pelo PUP em detrimento ao HD.

## 6. CONCLUSÃO

Essa pesquisa buscou identificar, dentre os fatores determinantes na escolha do consumidor entre PUP e HD, apresentados pela literatura, a sensibilidade do consumidor a alterações internas nos níveis desses fatores e qual a ordem de peso de influência que esses fatores exercem na opção do consumidor pelo Pick-up Point em detrimento ao Home Delivery, nas entregas de compras online.

O objetivo da pesquisa, portanto, foi atingido, uma vez os resultados alcançados geraram evidências de que, dentre os fatores determinantes para o consumidor que optam pelo PUP em detrimento ao HD, há uma evidência de que o consumidor é mais sensível a alterações no Preço do Frete do que no Prazo de Entrega. Além disso, outra evidência gerada foi a de que a ordem do peso de influência que esses fatores exercem sobre a preferência do consumidor pelo PUP em detrimento ao HD mostra que o atributo Prazo de Entrega tem maior peso que o atributo Preço do Frete. Por fim, essa pesquisa também aponta uma evidência de que o atributo Distância para o PUP não é um fator determinante para a preferência do consumidor, apesar de ser apresentado na literatura do tema de LM como tal.

É preciso levar em consideração o fato de que um PUP não deve, por conceito, ficar a uma distância muito grande do consumidor. Por isso, supõe-se que as distâncias utilizadas como níveis de atributo nessa pesquisa não foram relevantes o suficiente, a ponto de o modelo estatístico utilizado captar sua influência na escolha do consumidor pelo PUP em detrimento ao HD. A consequência disso foi que o atributo Distância não se mostrou estatisticamente significativo no modelo dessa pesquisa. Nesse sentido, sugere-se que estudos futuros não utilizem a distância a ser percorrida pelo consumidor até o PUP como um atributo do modelo, mas sim o tempo de deslocamento até o PUP, uma vez que o tempo para realizar um deslocamento de 1 quilômetro, por exemplo, pode ser de 5 minutos ou de 30 minutos, a depender da localização do PUP e do tráfego de veículos no local, o que se supõe que influenciaria na tomada de decisão do consumidor.

Outro tópico a ser abordado é o de que pesquisa foi realizada com a participação de 405 respondentes. Essa quantidade atende aos requisitos de amostra mínima para uma regressão logística, como sugerido pela literatura. Entretanto, dado a necessidade de se obter uma amostra equilibrada, a estratégia utilizada de retirar do questionário os cenários, a medida em que eles atingiam a quantidade de 15 respostas, fez com que, a partir de um certo momento, a escolha dos cenários, que idealmente deveria ser aleatória, passasse a ser pseudoaleatória, uma vez que a quantidade de cenários disponíveis para a escolha do respondente estaria reduzida. Nesse

caso, sugere-se que estudos futuros utilizem amostras maiores, de modo que não seja mais necessário a retirada de cenários para equilibrar as amostras. Além disso, a utilização de amostras maiores pode também suportar testes para grupos de confirmação.

Por fim, entende-se que a importância dessa pesquisa está em gerar evidências de quais fatores são significantes para o consumidor que optaria pelo PUP em detrimento ao HD, além de observar seu comportamento de sensibilidade com relação a alterações internas nos níveis desses atributos. Entretanto, sugere-se que estudos futuros realizem a mesma análise, utilizando dados de preferência revelada, de modo a observar se os resultados da preferência declarada abordada nessa pesquisa se confirmaram em situações reais.

## REFERÊNCIAS

- AGATZ, N. *et al.* Time slot management in attended home delivery. **Transportation Science**, v. 45, p. 435-449, 2011.
- AL NAWAYSEH, M. *et al.* An Adaptive Decision Support System for last mile Logistics in E-Commerce: A Study on Online Grocery Shopping. **International Journal of Decision Support System Technology**, v. 5, n. 1, p. 40-65, 2013.
- Associação Brasileira de Comércio Eletrônico (ABCOMM). **Crescimento do e-commerce no Brasil**. 2019. Disponível em: <<https://abcomm.org/noticias/crescimento-do-e-commerce-no-brasil/>>. Acesso em: 27 out 2019.
- BERGMANN, F. M.; WAGNER, S. M.; WINKENBACH, M. Integrating first-mile pickup and last-mile delivery on shared vehicle routes for efficient urban e-commerce distribution. **Transportation Research Part B: Methodological**, v. 131, p. 26-62, 2021.
- BOYER, K. K.; HULT, G. T. M. Extending the supply chain: Integrating operations and marketing in the online grocery industry. **Journal of Operations Management**, v. 23, p. 642-661, 2005.
- BOYER, K. K.; PRUD'HOMME, A. M. The last mile challenge: evaluating the effects of customer density and delivery window patterns. **Journal of Business Logistics**, v. 30, n. 1, 2009.
- BUJANG, M. A. *et al.* Sample Size Guidelines for Logistic Regression from Observational Studies with Large Population: Emphasis on the Accuracy Between Statistics and Parameters Based on Real Life Clinical Data. **Malays J Med Sci**, v. 25, n. 4, p. 122-130, 2018.
- CAMPBELL, A. M.; SAVELSBERGH, M. W. Decision support for consumer direct grocery initiatives. **Transportation Science**, v. 39, p. 313-327, 2005.
- CASPERSEN, E.; NAVRUD, S.; BENGTTSSON, J. Act locally? Are female online shoppers willing to pay to reduce the carbon footprint of last mile deliveries? **International Journal of Sustainable Transportation**, p. 1-15, 2021.
- CHEN, Z.; DUBINSKY, A. A Conceptual Model of Perceived Customer Value in E-Commerce: A Preliminary Investigation. **Psychology and Marketing**, v. 20, n. 4, p. 323-347, 2003.
- CHEN, Y.; WEI, J. Consumer's intention to use self-service parcel delivery service in online retailing. **Internet Research**, v. 28, n. 2, p. 500-519, 2018.
- CHEN, C. *et al.* The adoption of self-driving delivery robots in last mile logistics. **Transportation Research Part E: Logistics and Transportation Review**, v. 146, 2021.
- COASE, R. H. The Nature of the Firm. **Economica**, v. 4, n. 16, p. 386-405, 1937.
- COMI, A.; SAVCHENKO, L. Last-mile delivering: Analysis of environment-friendly transport. **Sustainable Cities and Society**, v. 74, 2021.
- DEUSTCH, Y.; GOLANY, B. A parcel locker network as a solution to the logistics last mile problem. **International Journal of Production Research**, v. 56, n. 1-2, p. 251-261, 2018.
- EDWARDS, J. *et al.* Carbon dioxide benefits of using collection-delivery points for failed home deliveries in the United Kingdom. **Transportation Research Record Journal of the Transportation Research Board**, 2191, p. 136-143, 2010.

- EMARKETER. **E-commerce share of total global retail sales from 2015 to 2021**. 2018. Disponível em: <<https://www.statista.com/statistics/534123/e-commerce-share-of-retail-sales-worldwide/>>. Acesso em: 27 out 2019.
- EMARKETER. **Global Ecommerce 2019: Ecommerce continues Strong gains amid global economic uncertainty**. 2019. Disponível em: <<https://www.emarketer.com/content/global-ecommerce-2019>>. Acesso em: 05 jun 2020.
- European Commission. **Special Eurobarometer 398**. 2013. Disponível em <[https://ec.europa.eu/commfrontoffice/publicopinion/archives/ebs/ebs\\_398\\_en.pdf](https://ec.europa.eu/commfrontoffice/publicopinion/archives/ebs/ebs_398_en.pdf)>. Acesso em: 31 out 2019.
- FAI, K. *et al.* An investigation of customers' intention to use self-collection services for last-mile delivery. **Transport Policy**, v. 66, p. 1-8, 2018.
- FARCOMENI, A.; VENTURA, L. An overview of robust methods in medical research. **Statistical methods in medical research**, v. 21, n. 2, p. 111-133, 2012.
- FARINA, E. M. M. Q. *et al.* **Competitividade: mercado, estado e organizações**. São Paulo: Singular, 1997, 286 p.
- FENG, C. M.; HUANG, Y. K. The choice behavior analysis on the pick-up point for the e-commerce retailing delivery. **Journal of the Eastern Asia Society for Transportation Studies**, v. 6, p. 2778-2793, 2005.
- FRANCKE, J.; VISSER, J. Internet shopping and its impacts on mobility. In: **25th World Road Congress (PIARC)**, p. 1-14, Seoul, Coréia do Sul, 2015.
- GEVAERS, R. *et al.* Characteristics Of Innovations In last mile Logistics - Using Best Practices, Case Studies And Making The Link With Green And Sustainable Logistics. **Association for European Transport and contributors**, Department of Transport and Regional Economics, University of Antwerp, Bélgica, 2009.
- GEVAERS, R. *et al.* Characteristics and typology of lastmile logistics from an innovation perspective in an urban context. In: MACHARIS, C.; MELO, S. **City distribution and Urban freight transport: Multiple perspectives**, Cheltenham, UK: Edward Elgar, 2011, p. 56-71.
- GEVAERS, R. *et al.* Cost Modelling and Simulation of Last-mile Characteristics in an Innovative B2C Supply Chain Environment with Implications on Urban Areas and Cities. **Procedia - Social and Behavioral Sciences**, v. 125, p. 398-411, 2014.
- GREEN, P. E.; SRINIVASAN, V. Conjoint Analysis in Consumer Research: Issues and Outlook. **Journal of Consumer Research**, v. 5, p. 103-123, 1978.
- HAGEN, T.; SCHEEL-KOPEINING, S. Would customers be willing to use an alternative (chargeable) delivery concept for the last mile? **Research in Transportation Business & Management**, v. 39, 2021.
- HAIR, J. F. *et al.* **Análise Multivariada de Dados**. Porto Alegre: Bookman, 2009.
- HENSHER, D. A. *et al.* **Applied Choice Analysis - A Primer**. New York: Cambridge University Press, 2005.
- HOSMER; D. W.; LEMESHOW, S. **Applied Logistic Regression**. 2ª ed, New York: Wiley-Interscience, 2000.

- HSIAO, M. H. Shopping mode choice: Physical store shopping versus e-shopping. **Transportation Research Part E: Logistics and Transportation Review**, v. 45, n. 1, p. 86-95., 2009.
- IBM Corp. Released 2017. IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp.
- IGNAT, B.; CHANKOV, S. Do e-commerce customers change their preferred last-mile delivery based on its sustainability impact? **The International Journal of Logistics Management**, v. 31, n. 3, p. 521-548, 2020.
- International Post Corporation. **State of e-commerce: global outlook 2016-21**. 2017. Disponível em: <<https://www.ipc.be/services/markets-and-regulations/e-commerce-market-insights/e-commerce-articles/global-ecommerce-figures-2017>>. Acesso em: 04 Nov 2019.
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Estimativas da população residente nos municípios brasileiros com data de referência em 1º de julho de 2020**. 2020. Disponível em: <[https://ftp.ibge.gov.br/Estimativas\\_de\\_Populacao/Estimativas\\_2020/POP2020\\_20210331.xls](https://ftp.ibge.gov.br/Estimativas_de_Populacao/Estimativas_2020/POP2020_20210331.xls)>. Acesso em: 19 mai 2021.
- JIANG, L. *et al.* A variable neighborhood search for the last-mile delivery problem during major infectious disease outbreak. **Optimization Letters**, 2021.
- KAPSER, S.; ABDELRAHMAN, M.; BERNECKER, T. Autonomous delivery vehicles to fight the spread of Covid-19 – How do men and women differ in their acceptance? **Transportation Research Part A: Policy and Practice**, v. 148, issue C, p. 183-198, 2021.
- KROES, E.; SHELDON, R. J. Stated preference methods. **Journal of Transport Economics and Policy**, v. 22, n. 1, p. 11-25, 1988.
- KUTNER, M. H. *et al.* **Applied linear statistical models**. 5ed. New York: McGraw-Hill/Irwin, 2005.
- LE, T.; UKKUSURI, S. Influencing Factors That Determine the Usage of the Crowd-Shipping Services. **Journal of the Transportation Research Board**, v. 2673, n. 7, 2019.
- LEE, H. L.; WHANG, S. Winning the last mile of e-commerce. **MIT Sloan Manag. Rev.**, v. 42, n. 4, 2001.
- LEMARDELÈ, C. *et al.* Potentialities of drones and ground autonomous delivery devices for last-mile logistics. **Transportation Research Part E: Logistics and Transportation Review**, v. 149, 2021.
- LEWIS, M. *et al.* An Empirical Study of the Impact of Nonlinear Shipping and Handling Fees on Purchase Incidence and Expenditure Decisions. **Marketing Science**, v. 25, n. 1, p. 51-64, 2006.
- LIU, C. *et al.* Assessing the impacts of collection-delivery points to individual's activity-travel patterns: A greener last mile alternative? **Transportation Research Part E Logistics and Transportation Review**, v. 121, p. 84-99, 2017.
- LOUVIERE, J. *et al.* **Stated Choice Methods: Analysis and applications**. Cambridge: Cambridge University Press, 399p., 2000.
- LOUVIERE, J. *et al.* Discrete Choice Experiments are not Conjoint Analysis. **Journal of Choice Modelling**, v. 3, n. 3, p. 57-72, 2010.

- MA, S. Fast or free shipping options in online and Omni-channel retail? The mediating role of uncertainty on satisfaction and purchase intentions. **The International Journal of Logistics Management**, v. 28, n. 4, p. 1099-1122, 2017.
- MAES, P. *et al.* Agents that buy and sell: Transforming Commerce as we Know It. **Communications of the ACM**, v. 42, n. 3, p. 81-91, 1999.
- MANSKI, C. F. The structure of Random Utility Models. **Theory and Decision**, v. 8, p. 229-254, 1977.
- McELREATH, RICHARD. **Statistical Rethinking: A Bayesian Course with Examples in R and Stan**. CRC Press, p. 189, 2016.
- McKINNON, A.; TALLAM, D. Unattended delivery to the home: an assessment of the security implications. **International Journal of Retail and Distribution Management**, v. 31, n. 1, p. 30-41, 2003.
- Ministério da Ciência, Tecnologia, Inovações e Comunicações (MCTIC). **Estratégia Brasileira para Transformação Digital: E-Digital**. Brasília, 2018. Disponível em: <<http://www.mctic.gov.br/mctic/export/sites/institucional/estrategiadigital.pdf>>. Acesso em: 04 nov 2019.
- MORGANTI, E. *et al.* The Impact of E-commerce on Final Deliveries: Alternative Parcel Delivery Services in France and Germany. **Transportation Research Procedia**, v. 4, p. 178-190, 2014a.
- MORGANTI, E. *et al.* Final deliveries for online shopping: The deployment of pickup point networks in urban and suburban áreas. **Research in Transportation Business & Management**, v. 11, p. 23-31, 2014b.
- MUÑOZ-VILLAMIZAR, A. *et al.* Non-Collaborative versus Collaborative Last-Mile Delivery in Urban Systems with Stochastic Demands. **In: 7th Industrial Product-Service System Conference**, Universidad de La Sabana, Colômbia, 2015.
- NEJADIRANI, F. *et al.* Developing Countries and Electronic Commerce the Case of SMEs. **World Applied Sciences Journal**, v. 15, n. 5, p. 756-764, 2011.
- NEOTRUST. **Novo normal: quais segmentos foram mais impactados durante o último ano?** 7ª ed. 2021. Disponível em: <<https://www.neotrust.com.br/>>. Acesso em: 26 out 2021.
- NGUYEN, D. H. *et al.* Consumer Behaviour and Order Fulfillment in Online Retailing: A Systematic Review. **International Journal of Management Reviews**, v. 20, n. 2, p. 255-276, 2018.
- NGUYEN, D. H. *et al.* What Is the Right Delivery Option for You? Consumer Preferences for Delivery Attributes in Online Retailing. **Journal of Business Logistics**, v. 40, n. 4, p. 299-321, 2019.
- OLIVEIRA, L. K. *et al.* Analysis of the potential demand of automated delivery stations for e-commerce deliveries in Belo Horizonte, Brazil. **Research in Transportation Economics**, v. 65, p. 34-43, 2017.
- ORTÚZAR, J. D.; WILLUMSEN, L. G. **Modelling Transports**. New York: John Willey & Sons, 3ª ed., 2001.
- PATELLA, S. M. *et al.* The Adoption of Green Vehicles in last mile Logistics: A Systematic Review. **Sustainability**, v. 13, n. 1, 2021.

PATCHOU, M.; SLIWA, B.; WIETFELD, C. Flying Robots for Safe and Efficient Parcel Delivery within the COVID-19 Pandemic. In: **15th Annual IEEE International Systems Conference**, 2021.

PERBOLLI, G. *et al.* A new model for Last-Mile Delivery and Satellite Depots management: The impact of the on-demand economy. **Transportation Research Part E: Logistics and Transportation Review**, v. 145, 2021.

Pesquisa Nacional por Amostras de Domicílios Contínua (PNAD). **PNAD Contínua TIC 2017: Internet chega a três em cada quatro domicílios do país**. 2017. Disponível em: <<https://agenciadenoticias.ibge.gov.br/agencia-sala-de-imprensa/2013-agencia-de-noticias/releases/23445-pnad-continua-tic-2017-internet-chega-a-tres-em-cada-quatro-domicilios-do-pais>>. Acesso em: 31 out 2019.

RAI, H. R.; VERLINDE, S.; MACHARIS, C. Who is interested in a crowdsourced last mile? A segmentation of attitudinal profiles. **Travel Behaviour and Society**, v. 22, p. 22-31, 2021.

RAO, S. *et al.* Failure to deliver? Linking online order fulfillment glitches with future purchase behavior. **Journal of Operations Management**, v. 29, p. 691-703, 2011.

RAO, S. *et al.* The role of physical distribution services as determinants of product returns in Internet retailing. **Journal of Operations Management**, v. 32, p. 295-312, 2014.

ROGERS, E. M. **Difussion of Innovations**. New York: Free Press, 3ª ed., 452 p., 1995.

SALESFORCE. **Relatório “Connected Shoppers”**. 3ª ed. 2019. Disponível em: <[https://www.salesforce.com/content/dam/web/pt\\_br/www/documents/e-books/commerce-cloud/Salesforce%20Connected%20Shoppers%20Report%202019.pdf](https://www.salesforce.com/content/dam/web/pt_br/www/documents/e-books/commerce-cloud/Salesforce%20Connected%20Shoppers%20Report%202019.pdf)>. Acesso em: 26 out 2021.

SANKO, N. **Guideliness for Stated Preference Experiment Design**. Dissertação (Mestrado em Administração de Negócios) – School of International Management Ecole Nationale des Ponts et Chaussées, 2001.

SAVRUL, M. *et al.* The Potential of E-commerce for SMEs in a Globalizing Business Environment. **Procedia - Social and Behavioral Sciences**, v. 150, p. 35-45, 2014.

SENADO FEDERAL. **Criados critérios de classificação do espaço urbano e rural**. Disponível em: <https://www12.senado.leg.br/noticias/materias/2009/10/06/criados-criterios-de-classificacao-do-espaco-urbano-e-rural>. Acesso em: 10 mai 2021.

SETTEY, T. *et al.* The Growth of E-Commerce Due to COVID-19 and the Need for Urban Logistics Centers Using Electric Vehicles: Bratislava Case Study. **Sustainability**, v. 13, n. 10, 2021.

SHAMPANIER, K. *et al.* Zero as special price: The true value of free products. **Marketing Science**, v. 26, n. 6, p. 742-757, 2007.

SILVA, J. da. **Avaliação da viabilidade dos Pick-up Points sob o enfoque da logística urbana**. Dissertação (Mestrado em Geotecnia e Transportes – Escolha de Engenharia da Universidade Federal de Minas Gerais, Belo Horizonte, 2018.

SILVA, J. da *et al.* Demand analysis for pick-up sites as an alternative solution for home delivery in the Brazilian context. **Transportation Research Procedia**, v. 39, p. 462-470, 2019.

SIMON, H. A. A behavioral model of rational choice. **The Quarterly Journal of Economics**, v. 69, p. 99-118, 1955.

- SIRAGUSA, C. *et al.* Electric vehicles performing last-mile delivery in B2C e-commerce: An economic and environmental assessment. **International Journal of Sustainable Transportation**, 2020.
- SRIVATSA, S. S.; MARATHE, R. R. Moving towards “mobile warehouse”: Last-mile logistics during COVID-19 and beyond. **Transportation Research Interdisciplinary Perspectives**, v. 10, 2021.
- STACORP. Stata Statistical Software: Release 13. College Station, TX: StataCorp LP, 2013.
- STATISTA. **Retail e-commerce sales worldwide from 2014 to 2023 (in billion U.S. dollars)**. Disponível em: < <https://www.statista.com/statistics/379046/worldwide-retail-e-commerce-sales/>>. Acesso em: 25 nov 2019.
- SUGUNA, M. *et al.* A study on the influential factors of the last mile delivery projects during Covid-19 era. **Operations Management Research**, 2021.
- TA, H. *et al.* Designing crowdsourced delivery systems: The effect of driver disclosure and ethnic similarity. **Journal of Operations Management**, v. 60, p. 19-33, 2018.
- TWADDLE, H. A. **Stated preference survey design and pre-test for valuing influencing factors for bicycle use**. Dissertação (Mestrado em Sistemas de Transporte) – Institute of Transportation, Technische Universität München, 2011.
- TSAI, Y. T.; TIWASING, P. Customers’ intention to adopt smart lockers in last-mile delivery service: A multi-theory perspective. **Journal of Retailing and Consumer Services**, v. 61, n. 2, 2021.
- VAKULENKO, Y. *et al.* What's in the parcel locker? Exploring customer value in e-commerce last mile delivery. **Journal of Business Research**, v. 88, p. 421-427, 2018.
- VAKULENKO, Y. *et al.* Service innovation in e-commerce last mile delivery: Mapping the e-customer journey. **Journal of Business Research**, v. 101, p. 461-468, 2019.
- VAN DUIN, J.; *et al.* Improving Home Delivery Efficiency by Using Principles of Address Intelligence for B2C Deliveries. **Transportation Research Procedia**, v. 12, p. 14-25, 2016.
- VILLA, R.; MONZÓN, A. A Metro-Based System as Sustainable Alternative for Urban Logistics in the Era of E-Commerce. **Sustainability**, v. 13, n. 8, 2021.
- VON NEUMANN, J.; MORGENSTERN, O. **Theory of Games and Economic Behavior**. Princeton: Princeton University Press, 1944.
- VYT, D. *et al.* Grocery pickup creation of value: Customers’ benefits vs. spatial dimension. **Journal of Retailing and Consumer Services**, v. 39, p. 145-153, 2017.
- WEBSHOPPERS. **Webshoppers report**. 37<sup>a</sup> ed., 2018. Disponível em: <<https://company.ebit.com.br/webshoppers>>. Acesso em: 26 out 2021.
- WILLIAMSON, O. E. **The Economic Institutions of Capitalism**. New York: Free Press, 468p., 1985.
- XU, J.; HONG, L. Impact Factors of Choosing Willingness for Picking up Service. **Journal of Applied Sciences, Engineering and Technology**, v. 6, n. 14, p. 2509-2513, 2013.
- YUEN, K. *et al.* The determinants of customers’ intention to use smart lockers for last-mile deliveries. **Journal of Retailing and Consumer Services**, v. 49, p. 316-326, 2019.

ZEITHAML, V. A. Service quality, profitability, and the economic worth of customers: what we know and what we need to learn. **Journal of the Academy of Marketing Sciences**, v. 28, p. 67-85, 2000.

ZHANG, L. *et al.* Simulation-based Assessment of Cargo Bicycle and Pick-up Point in Urban Parcel Delivery. **Procedia Computer Science**, v. 130, p. 18-25, 2018.

ZHOU, L. *et al.* Location-Routing Problem with Simultaneous Home Delivery and Customer's Pickup for City Distribution of Online Shopping Purchases. **Sustainability**, v. 8, n. 8, 2016.

## APÊNDICE A – SITUAÇÃO HIPOTÉTICA E CENÁRIOS DA SURVEY DE PD

### Situação

Imagine que está se aproximando o aniversário de uma pessoa muito querida. Você precisa comprar um presente e, por questão de praticidade, você opta por comprá-lo online. Escolhido o produto, chegou a hora de escolher o método de entrega. O aniversário está próximo e você sabe que seria uma situação desagradável se o presente não chegasse a tempo. As opções que o site te oferece são: Entregar na sua casa ou Buscar em um parceiro da loja (farmácia, padaria, universidades...) a sua escolha. Você tem carro, o produto é facilmente carregável e você mora em prédio com porteiro para receber o produto em horário comercial.

Com base nessa situação hipotética, analise as características oferecidas abaixo pela loja:

### Cenário 1

#### Cenário

- Frete: o valor do frete para receber em casa ou ir buscar no estabelecimento parceiro é o mesmo;
- Localização: o estabelecimento parceiro se encontra entre 1 e 2 km de distância da sua casa;
- Prazo: o prazo de entrega para receber em casa ou ir buscar no estabelecimento parceiro é o mesmo.

### Cenário 2

#### Cenário

- Frete: o valor do frete para ir buscar no estabelecimento parceiro é 15% menor do que o de receber em casa
- Localização: o estabelecimento parceiro se encontra entre 1 e 2 km de distância da sua casa;
- Prazo: o prazo de entrega para receber em casa ou ir buscar no estabelecimento parceiro é o mesmo.

### Cenário 3

#### Cenário

- Frete: o valor do frete para ir buscar no estabelecimento parceiro é 30% menor do que o de receber em casa
- Localização: o estabelecimento parceiro se encontra entre 1 e 2 km de distância da sua casa;
- Prazo: o prazo de entrega para receber em casa ou ir buscar no estabelecimento parceiro é o mesmo.

### Cenário 4

#### Cenário

- Frete: o valor do frete para receber em casa ou ir buscar no estabelecimento parceiro é o mesmo;
- Localização: o estabelecimento parceiro se encontra em até 1km da sua residência;
- Prazo: o prazo de entrega para receber em casa ou ir buscar no estabelecimento parceiro é o mesmo.

### Cenário 5

#### Cenário

- Frete: o valor do frete para ir buscar no estabelecimento parceiro é 15% menor do que o de receber em casa;
- Localização: o estabelecimento parceiro se encontra em até 1km da sua residência;
- Prazo: o prazo de entrega para receber em casa ou ir buscar no estabelecimento parceiro é o mesmo.

### Cenário 6

**Cenário**

- Frete: o valor do frete para ir buscar no estabelecimento parceiro é 30% menor do que o de receber em casa;
- Localização: o estabelecimento parceiro se encontra em até 1km da sua residência;
- Prazo: o prazo de entrega para receber em casa ou ir buscar no estabelecimento parceiro é o mesmo.

**Cenário 7****Cenário**

- Frete: o valor do frete para receber em casa ou ir buscar no estabelecimento parceiro é o mesmo;
- Localização: o estabelecimento parceiro se encontra ao lado da sua residência;
- Prazo: o prazo de entrega para receber em casa ou ir buscar no estabelecimento parceiro é o mesmo.

**Cenário 8****Cenário**

- Frete: o valor do frete para ir buscar no estabelecimento parceiro é 15% menor do que o de receber em casa;
- Localização: o estabelecimento parceiro se encontra ao lado da sua residência;
- Prazo: o prazo de entrega para receber em casa ou ir buscar no estabelecimento parceiro é o mesmo.

**Cenário 9****Cenário**

- Frete: o valor do frete para ir buscar no estabelecimento parceiro é 30% menor do que o de receber em casa;
- Localização: o estabelecimento parceiro se encontra ao lado da sua residência;
- Prazo: o prazo de entrega para receber em casa ou ir buscar no estabelecimento parceiro é o mesmo.

**Cenário 10**

:::

**Cenário**

- Frete: o valor do frete para receber em casa ou ir buscar no estabelecimento parceiro é o mesmo;
- Localização: o estabelecimento parceiro se encontra entre 1 e 2 km de distância da sua casa;
- Prazo: o prazo de entrega para buscar no estabelecimento parceiro é de 2 dias mais rápido do que receber em casa.

**Cenário 11****Cenário**

- Frete: o valor do frete para ir buscar no estabelecimento parceiro é 15% menor do que o de receber em casa;
- Localização: o estabelecimento parceiro se encontra entre 1 e 2 km de distância da sua casa;
- Prazo: o prazo de entrega para buscar no estabelecimento parceiro é de 2 dias mais rápido do que receber em casa.

**Cenário 12****Cenário**

- Frete: o valor do frete para ir buscar no estabelecimento parceiro é 30% menor do que o de receber em casa;
- Localização: o estabelecimento parceiro se encontra entre 1 e 2 km de distância da sua casa;
- Prazo: o prazo de entrega para buscar no estabelecimento parceiro é de 2 dias mais rápido do que receber em casa.

**Cenário 13**

**Cenário**

- Frete: o valor do frete para receber em casa ou ir buscar no estabelecimento parceiro é o mesmo;
- Localização: o estabelecimento parceiro se encontra até 1km de distância da sua residência;
- Prazo: o prazo de entrega para buscar no estabelecimento parceiro é de 2 dias mais rápido do que receber em casa.

**Cenário 14**

...

**Cenário**

- Frete: o valor do frete para ir buscar no estabelecimento parceiro é 15% menor do que o de receber em casa;
- Localização: o estabelecimento parceiro se encontra até 1km de distância da sua residência;
- Prazo: o prazo de entrega para buscar no estabelecimento parceiro é de 2 dias mais rápido do que receber em casa.

**Cenário 15****Cenário**

- Frete: o valor do frete para ir buscar no estabelecimento parceiro é 30% menor do que o de receber em casa;
- Localização: o estabelecimento parceiro se encontra até 1km de distância da sua residência;
- Prazo: o prazo de entrega para buscar no estabelecimento parceiro é de 2 dias mais rápido do que receber em casa.

**Cenário 16****Cenário**

- Frete: o valor do frete para receber em casa ou ir buscar no estabelecimento parceiro é o mesmo;
- Localização: o estabelecimento parceiro se encontra ao lado da sua residência;
- Prazo: o prazo de entrega para buscar no estabelecimento parceiro é de 2 dias mais rápido do que receber em casa.

**Cenário 17****Cenário**

- Frete: o valor do frete para ir buscar no estabelecimento parceiro é 15% menor do que o de receber em casa;
- Localização: o estabelecimento parceiro se encontra ao lado da sua residência;
- Prazo: o prazo de entrega para buscar no estabelecimento parceiro é de 2 dias mais rápido do que receber em casa.

**Cenário 18****Cenário**

- Frete: o valor do frete para ir buscar no estabelecimento parceiro é 30% menor do que o de receber em casa;
- Localização: o estabelecimento parceiro se encontra ao lado da sua residência;
- Prazo: o prazo de entrega para buscar no estabelecimento parceiro é de 2 dias mais rápido do que receber em casa.

**Cenário 19****Cenário**

- Frete: o valor do frete para receber em casa ou ir buscar no estabelecimento parceiro é o mesmo;
- Localização: o estabelecimento parceiro se encontra entre 1 e 2 km de distância da sua casa;
- Prazo: o prazo de entrega para buscar no estabelecimento parceiro é de 4 dias mais rápido do que receber em casa.

### Cenário 20

#### Cenário

- Frete: o valor do frete para ir buscar no estabelecimento parceiro é 15% menor do que o de receber em casa;
- Localização: o estabelecimento parceiro se encontra entre 1 e 2 km de distância da sua casa;
- Prazo: o prazo de entrega para buscar no estabelecimento parceiro é de 4 dias mais rápido do que receber em casa.

### Cenário 21

#### Cenário

- Frete: o valor do frete para ir buscar no estabelecimento parceiro é 30% menor do que o de receber em casa;
- Localização: o estabelecimento parceiro se encontra entre 1 e 2 km de distância da sua casa;
- Prazo: o prazo de entrega para buscar no estabelecimento parceiro é de 4 dias mais rápido do que receber em casa.

### Cenário 22

#### Cenário

- Frete: o valor do frete para receber em casa ou ir buscar no estabelecimento parceiro é o mesmo;
- Localização: o estabelecimento parceiro se encontra até 1km de distância da sua residência;
- Prazo: o prazo de entrega para buscar no estabelecimento parceiro é de 4 dias mais rápido do que receber em casa.

### Cenário 23

#### Cenário

- Frete: o valor do frete para ir buscar no estabelecimento parceiro é 15% menor do que o de receber em casa;
- Localização: o estabelecimento parceiro se encontra até 1km de distância da sua residência;
- Prazo: o prazo de entrega para buscar no estabelecimento parceiro é de 4 dias mais rápido do que receber em casa.

### Cenário 24

...

#### Cenário

- Frete: o valor do frete para ir buscar no estabelecimento parceiro é 30% menor do que o de receber em casa;
- Localização: o estabelecimento parceiro se encontra até 1km de distância da sua residência;
- Prazo: o prazo de entrega para buscar no estabelecimento parceiro é de 4 dias mais rápido do que receber em casa.

### Cenário 25

#### Cenário

- Frete: o valor do frete para receber em casa ou ir buscar no estabelecimento parceiro é o mesmo;
- Localização: o estabelecimento parceiro se encontra ao lado da sua residência;
- Prazo: o prazo de entrega para buscar no estabelecimento parceiro é de 4 dias mais rápido do que receber em casa.

### Cenário 26

**Cenário**

- Frete: o valor do frete para ir buscar no estabelecimento parceiro é 15% menor do que o de receber em casa;
- Localização: o estabelecimento parceiro se encontra ao lado da sua residência;
- Prazo: o prazo de entrega para buscar no estabelecimento parceiro é de 4 dias mais rápido do que receber em casa.

**Cenário 27****Cenário**

- Frete: o valor do frete para ir buscar no estabelecimento parceiro é 30% menor do que o de receber em casa;
- Localização: o estabelecimento parceiro se encontra ao lado da sua residência;
- Prazo: o prazo de entrega para buscar no estabelecimento parceiro é de 4 dias mais rápido do que receber em casa.

## APÊNDICE B – QUESTÕES SOCIODEMOGRÁFICAS

Seção 30 de 30

### Questionário demográfico



Por favor, agora responda rapidamente a essas questões:

Quando um novo serviço é lançado, você costuma consumir ele: \*

- Antes que outras pessoas conhecidas experimentem
- Após outras pessoas conhecidas experimentarem

Você já teve alguma experiência negativa com relação à entrega de um produto comprado em e-commerce? Por exemplo: falha na entrega, atraso na entrega... \*

- Sim
- Não

Quantos anos você tem? \*

- Menos que 21 anos
- Entre 21 e 30 anos
- Acima de 30 anos

Você já comprou alguma vez pela internet e optou por retirar o produto na loja? \*

Sim

Não

Quantos habitantes vivem na sua cidade?

Menos de 50 mil habitantes

Entre 50 e 500 mil habitantes

Acima de 500 mil habitantes

**APÊNDICE C – SAÍDAS EM DFBETA**

```
. sum dbeta
```

Variable	Obs	Mean	Std. Dev.	Min	Max
dbeta	243	9.107459	9.89945	.0002015	28.51396

```
.  
end of do-file
```

## APÊNDICE D – SCRIPT REGRESSÃO LOGÍSTICA

\*Estatística descritiva

summarize X1 X2 X3 X4 X5 X6 R1

\*Distribuição de frequência VD

tab R1

\*Regressão logística binomial

logit R1 X1 X2 X3 X4 X5 X6, vce(cluster CENRIO), level(95)

\*logit R1 X1 X2 X3 X4 X5 X6, level(95) noconstant

\*Teste de Wald

test X3 X4

\*AIC e BIC (modelo completo)

estat ic

\*Salvar estimativas dos parâmetros do modelo completo

lrtest, saving(0)

\*Regressão logística stepwise forward

stepwise, pr(0.10) pe(0.05) forward: logit R1 X1 X2 X3 X4 X5 X6, vce(cluster CENRIO)

\*AIC e BIC (modelo final)

estat ic

\*Odds ratio

logit, or

\*Teste de Hosmer-Lemeshow

estat gof, group(10) table

\*Sensibilidade do modelo

estat class, cutoff(0.6988)

\*Curva ROC

lroc

\*DFBETA

predict dbeta1, dbeta

scatter dbeta1 ID, mlab(ID)

sum dbeta1

\*Linktest (erro de especificação)

linktest

\*Reestimativa de Jackknife

jackknife \_b, level(95) : logit R1 X1 X2 X5 X6, level(95)



```

. *Regressao logistica stepwise forward
. stepwise, pr(0.10) pe(0.05) forward: logit R1 X1 X2 X3 X4 X5 X6, vce(cluster CENRIO)
      begin with empty model
p = 0.0081 < 0.0500 adding X5
p = 0.0000 < 0.0500 adding X6
p = 0.0031 < 0.0500 adding X1
p = 0.0001 < 0.0500 adding X2

Logistic regression
Log pseudolikelihood = -220.87921
Number of obs = 405
Wald chi2(4) = 124.07
Prob > chi2 = 0.0000
Pseudo R2 = 0.1087

      (Std. Err. adjusted for 27 clusters in CENRIO)

```

R1	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
X5	1.514689	.1819975	8.32	0.000	1.15798	1.871397
X6	1.418698	.2317843	6.12	0.000	.9644086	1.872986
X1	1.003702	.1816025	5.53	0.000	.6477679	1.359636
X2	.8691699	.2281864	3.81	0.000	.4219327	1.316407
_cons	-.644669	.1433515	-4.50	0.000	-.9256327	-.3637054

Teste de Wald:

```

. *teste de wald
. test X3 X4

( 1) [R1]X3 = 0
( 2) [R1]X4 = 0

      chi2( 2) = 3.41
      Prob > chi2 = 0.1818

```

Regressão Logística Odds Ratio:

```

. *Odds ratio
. logit, or

Logistic regression               Number of obs   =       405
                                Wald chi2(4)     =       124.07
                                Prob > chi2       =       0.0000
Log pseudolikelihood = -220.87921 Pseudo R2       =       0.1087

                                (Std. Err. adjusted for 27 clusters in CENRIO)

```

R1	Odds Ratio	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
X5	4.548005	.8277256	8.32	0.000	3.183497	6.497369
X6	4.131735	.9576715	6.12	0.000	2.623236	6.507702
X1	2.728364	.4954776	5.53	0.000	1.91127	3.894777
X2	2.38493	.5442087	3.81	0.000	1.524906	3.729996
_cons	.5248362	.075236	-4.50	0.000	.3962806	.695096

AIC e BIC modelo final:

```

Akaike's information criterion and Bayesian information criterion

```

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	405	-247.8222	-220.8792	5	451.7584	471.7779

Note: N=Obs used in calculating BIC; see [R] BIC note

Teste de Hosmer-Lemeshow:

```

Logistic model for R1, goodness-of-fit test

(Table collapsed on quantiles of estimated probabilities)
(There are only 9 distinct quantiles because of ties)

```

Group	Prob	Obs_1	Exp_1	Obs_0	Exp_0	Total
1	0.3442	14	15.5	31	29.5	45
2	0.5559	25	25.0	20	20.0	45
3	0.5888	28	26.5	17	18.5	45
4	0.6844	34	30.8	11	14.2	45
5	0.7047	30	31.7	15	13.3	45
6	0.8380	36	37.7	9	7.3	45
7	0.8506	40	38.3	5	6.7	45
8	0.8554	37	38.5	8	6.5	45
10	0.8669	39	39.0	6	6.0	45

```

number of observations =       405
number of groups      =         9
Hosmer-Lemeshow chi2(7) =       3.19
Prob > chi2          =       0.8666

```

Análise de sensibilidade:

Logistic model for R1

Classified	True		Total
	D	~D	
+	182	43	225
-	101	79	180
Total	283	122	405

Classified + if predicted Pr(D) >= .6988  
True D defined as R1 != 0

Sensitivity	Pr( +   D)	64.31%
Specificity	Pr( -   ~D)	64.75%
Positive predictive value	Pr( D   +)	80.89%
Negative predictive value	Pr( ~D   -)	43.89%
False + rate for true ~D	Pr( +   ~D)	35.25%
False - rate for true D	Pr( -   D)	35.69%
False + rate for classified +	Pr( ~D   +)	19.11%
False - rate for classified -	Pr( D   -)	56.11%
Correctly classified		64.44%

Saídas do dfbeta de Pregibon:

```
. sum dbetal
```

Variable	Obs	Mean	Std. Dev.	Min	Max
dbetal	405	.5652928	1.233797	5.18e-06	4.0351

Saídas erro de especificação:

```
. *Linktest (erro de especificacao)
. linktest
```

Iteration 0: log likelihood = -247.82222  
Iteration 1: log likelihood = -221.16692  
Iteration 2: log likelihood = -220.63361  
Iteration 3: log likelihood = -220.63238  
Iteration 4: log likelihood = -220.63238

Logistic regression	Number of obs	=	405
	LR chi2(2)	=	54.38
	Prob > chi2	=	0.0000
Log likelihood = -220.63238	Pseudo R2	=	0.1097

R1	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
_hat	1.172593	.2880888	4.07	0.000	.6079496 1.737237
_hatsq	-.1250523	.178577	-0.70	0.484	-.4750567 .2249521
_cons	.0217869	.1642065	0.13	0.894	-.3000519 .3436258

Saídas jackknife:

```

. jackknife _b, level(95) : logit R1 X1 X2 X5 X6, vce(cluster CENRIO)
(running logit on estimation sample)

Jackknife replications (27)
-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|-----|-----|-----|-----|-----|-----|-----|-----|-----|
.....

Logistic regression                               Number of obs   =       405
                                                Replications    =        27
                                                F(   4,   26)   =       22.79
                                                Prob > F        =       0.0000
Log likelihood = -220.87921                       Pseudo R2      =       0.1087

                                                (Replications based on 27 clusters in CENRIO)

```

R1	Coef.	Jackknife Std. Err.	t	P> t	[95% Conf. Interval]	
X1	1.003702	.2156616	4.65	0.000	.5604034	1.447001
X2	.8691699	.2734232	3.18	0.004	.3071406	1.431199
X5	1.514689	.2143942	7.06	0.000	1.073995	1.955382
X6	1.418698	.276455	5.13	0.000	.8504362	1.986959
_cons	-.644669	.1722655	-3.74	0.001	-.9987658	-.2905723