

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

JOÃO PEDRO ARAUJO DOMINGUES

**PREVISÃO DE VENDAS NA PRESENÇA DE UM ELEVADO
NÚMERO DE VARIÁVEIS:
UM ESTUDO DE CASO DE ITENS INTRA E INTER-
CATEGORIA**

**VITÓRIA
2018**

JOÃO PEDRO ARAUJO DOMINGUES

**PREVISÃO DE VENDAS NA PRESENÇA DE UM ELEVADO
NÚMERO DE VARIÁVEIS:
UM ESTUDO DE CASO DE ITENS INTRA E INTER-
CATEGORIA**

Dissertação apresentada ao Programa de Pós-graduação em Administração do Centro de Ciências Jurídicas e Econômicas da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do título de Mestre em Administração, na linha de pesquisa Estratégia, Inovação e Desempenho organizacional.

Prof. Dr. Hélio Zanquetto Filho

VITÓRIA
2018

Dados Internacionais de Catalogação-na-publicação (CIP)
(Biblioteca Central da Universidade Federal do Espírito Santo, ES, Brasil)

Domingues, João Pedro Araujo, 1990-
D671p Previsão de vendas na presença de um elevado número de
variáveis : um estudo de caso de itens intra e inter-categoria /
João Pedro Araujo Domingues. – 2018.
82 f. : il.

Orientador: Hélio Zanquetto Filho.
Dissertação (Mestrado em Administração) – Universidade
Federal do Espírito Santo, Centro de Ciências Jurídicas e
Econômicas.

1. Previsão de vendas. 2. Comércio varejista. 3. Análise de
regressão. 4. Controle de estoque. I. Zanquetto Filho, Hélio. II.
Universidade Federal do Espírito Santo. Centro de Ciências
Jurídicas e Econômicas. III. Título.

CDU: 65

Elaborado por Perla Rodrigues Lôbo – CRB-6 ES-527/O

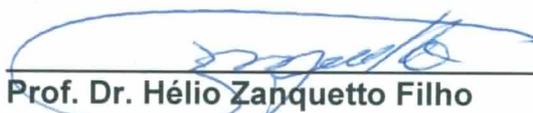
JOÃO PEDRO ARAUJO DOMINGUES

**PREVISÃO DE VENDAS NA PRESENÇA DE UM ELEVADO NÚMERO
DE VARIÁVEIS: UM ESTUDO DE CASO DE ITENS INTRA E INTER-
CATEGORIA**

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

Aprovada em 4 de maio de 2018.

COMISSÃO EXAMINADORA



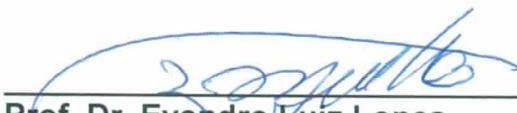
Prof. Dr. Hélio Zanquetto Filho
Universidade Federal do Espírito Santo
Orientador



Prof. Dr. Marcelo Moll Brandão
Universidade Federal do Espírito Santo



Prof. Dr. Alexandre Loureiros Rodrigues
Universidade Federal do Espírito Santo



Prof. Dr. Evandro Luiz Lopes
Universidade Nove de Julho – membro
remoto

DEDICATÓRIA

Dedico este trabalho aos meus pais, Ana e José, pelo incentivo e apoio incondicional para sua realização.

AGRADECIMENTOS

À minha família, pelo grande apoio que me deram em todas as fases da minha vida, que se revelaram cruciais, principalmente nesta etapa, para a elaboração desta dissertação. Agradeço-lhes o constante incentivo e a habitual motivação com que pude contar ao longo deste trabalho.

Ao meu orientador, Professor Doutor Hélio Zanquetto Filho, por acreditar na minha capacidade, bem como pelo apoio, pela motivação, pelas críticas construtivas e pela paciência que teve comigo nos momentos difíceis de ansiedade. Agradeço-lhe pela sua compreensão, ensinamentos, conselhos e total disponibilidade que revelou em todas as fases do trabalho e, acima de tudo, pela confiança que depositou em mim.

Aos professores do PPGADM (Programa de Pós-Graduação em Administração), pelos ricos ensinamentos relacionados à disciplina, como também conselhos relacionados à vida pessoal e profissional. A todos os funcionários do PPGADM (Programa de Pós-Graduação em Administração), por demonstrarem atenção e paciência na resolução de problemas ao me auxiliarem.

A todos os meus amigos e colegas, por estarem perto, sempre me apoiando e motivando.

A Empresa estudada, por tornar possível a realização da minha pesquisa, principalmente, fornecendo todos os dados necessários.

A FAPES (Fundação de Amparo à Pesquisa do Espírito Santo) pela concessão da bolsa durante todo o período de realização deste mestrado.

*“Não adianta olhar para o céu
com muita fé e pouca luta.”*

Gabriel o Pensador

RESUMO

A presente pesquisa teve como objetivo verificar se séries históricas referentes a Intra e Inter-categorias são capazes de melhorar o modelo de previsão de vendas de curto prazo para o varejo. Trata-se de um estudo de caso utilizando regressões múltiplas e o método de seleção de variáveis LASSO (*Least Absolute Shrinkage And Selection Operator*). Os objetivos específicos consistiram em: (1) confirmar empiricamente a existência de itens complementares e substitutos em Intra e Inter-categorias; (2) propor um modelo de previsão que leve em consideração séries de Intra e Inter-categorias; (3) comparar os resultados encontrados entre o modelo de previsão com somente uma série histórica e o modelo proposto com Intra e Inter-categorias; (4) identificar se existe diferença entre os resultados do modelo com séries Intra e Inter-categorias. Assim, os principais resultados identificados revelam que foi comprovada a existência de itens complementares e substitutos em Intra e Inter-categorias no nível de gramatura. Além disso, os resultados demonstraram maior prevalência de itens complementares, o que representa em média 88,8% das interações, sendo os demais 11,2% substitutos; os resultados apontam que 83,8% da melhoria dos resultados do RMSE são provenientes das séries Intra-categoria, o que representa a maioria expressiva da contribuição. Dentro deste percentual, a redução média do RMSE foi de 56,30%. Entretanto, o estudo destaca que séries Inter-categorias também são capazes de contribuir com 16,2% para melhorar a acurácia, demonstrando assim uma redução do erro e comprovando a existência de interação entre séries ao longo das categorias. Por fim, conclui-se que a utilização de séries pertencentes apenas a Intra-categoria para compor o modelo de previsão consegue melhorar a acurácia na maioria dos casos, e que a redução alcançada atinge resultados satisfatórios.

Palavras Chave: Previsão de vendas; Varejo supermercadista; Regressão múltipla, Business analytics.

ABSTRACT

In the present work the main objective was verify if time series of Intra and Inter-category were able to improve sales forecasting model in short term for retail. This is a study case that used multiple regression and the methodological selection of variables Lasso (*Least Absolute Shrinkage And Selection Operator*). The specific objectives were: (1) Confirm empirically the existence of complementarity and substitutability in Intra and Inter-category; (2) Propound a sales forecasting model that uses time series from both Intra and Inter-category; (3) Compare the results between the model which uses only one time series and the model which uses Intra and Inter-category; (4) Identify if there are any difference between the results using Intra or Inter-category. Therefore, the major results have shown the presence of complementarity and substitutability among Intra and Inter-category at weight level. Furthermore, the outcomes have shown the more prevalence of complementarity, with 88.8% of the interactions, and the rest 11.2% were substitutability. Other results displayed that 83.8% of the improvements of RMSE comes from time series from Intra-category, which represents the prime majority. Inside this percentage, the decrease average of RMSE was 56.60%. Meanwhile, the research highlighted that Inter-category was capable to improve 16.2% of accuracy, showing the reduction of the error and proving the series interaction cross-category. Finally, the study concluded that the usage of time series from Intra-category can improve the accuracy in the majority of the cases, and the reduction reached satisfactory outcomes.

Keywords: Sales forecasting; Grocery; Retail; Multiple Regression; Business Analytics.

LISTA DE FIGURAS

Figura 1 - Segmentação dos fatores que influenciam na demanda.	22
Figura 2 – Relação Intra e Inter-categoria.	25
Figura 3 - Modelo proposto de previsão.	33
Figura 4 – Interação entre variáveis para composição do modelo de regressão.	34
Figura 5 - Separação dos dados nas fases de “treino” e “estimação”.	35
Figura 6 – Classificação do banco de dados por nível.	40
Figura 7 – Diagrama de relacionamento entre séries complementares e substitutas.	49
Figura 8 – Matriz de correlação das séries históricas da categoria Frios e Laticínios.	50

LISTA DE GRÁFICOS

Gráfico 1 – Demonstração de demanda sazonal e tendência ao longo do tempo	32
Gráfico 2 – Separação de dados para treino, estimação e mensuração da acurácia da previsão.....	42
Gráfico 3 - Classificação das categorias pelo método de Pareto.	45
Gráfico 4 – Demonstração do ajuste do modelo de regressão Lasso aos dados reais.	52

LISTA DE QUADRO

Quadro 1 - Técnicas de previsão	26
Quadro 2 – Categorias inseridas nos respectivos cálculos.	44
Quadro 3 – Legenda de códigos.	48
Quadro 4 – Cenários das categorias inseridas nos respectivos cálculos.	53
Quadro 5 – Cenários para cálculos Intra e Inter-categorias.	57

LISTA DE TABELAS

Tabela 1 - Quantidades de Séries por nível.	46
Tabela 2 – Matriz de complementares e substitutos.	48
Tabela 3 – Relação complementar e substituta Intra e Inter-categorias.	51
Tabela 4 - Comparação entre resultados para previsão Holt-Winters e LASSO.	55
Tabela 5 - Comparação entre resultados da Intra-categorias.	56
Tabela 6 – Contribuição da Intra e Inter-categorias para redução do RMSE.	57

LISTA DE SIGLAS

IGBE – Instituto Brasileiro de Geografia e Estatística

LASSO - *Least Absolute Shrinkage And Selection Operator*

PIB - Produto Interno Bruto

RMSE – *root mean square error*

SBVC - Sociedade Brasileira de Varejo e Consumo

SKU – *Stock Keep Unit*

SUMÁRIO

1	INTRODUÇÃO	12
1.1	OBJETIVOS	14
1.1.1	Objetivo Geral	14
1.1.2	Objetivos Específicos	15
1.2	JUSTIFICATIVA	15
1.3	ORGANIZAÇÃO DA DISSERTAÇÃO	18
2	REVISÃO BIBLIOGRÁFICA	20
2.1	PREVISÃO DE VENDAS	20
2.2	INTERAÇÃO ENTRE PRODUTOS	21
2.3	MODELOS QUANTITATIVOS DE PREVISÃO	25
2.3.1	Modelo de previsão com elevado número de variáveis	28
2.4	MODELO UTILIZADO NA PESQUISA	31
2.4.1	Avaliação da previsão	34
3	PROCEDIMENTO METODOLÓGICO	37
3.1	ABORDAGEM METODOLÓGICA	37
3.2	COLETA E ORGANIZAÇÃO DE DADOS	38
3.3	PROCEDIMENTOS DE ANÁLISE DOS DADOS	41
4	APRESENTAÇÃO E ANÁLISE DE DADOS	45
4.1	ANÁLISE PRELIMINAR DOS DADOS	45
4.2	IDENTIFICAÇÃO DE COMPLEMENTARES E SUBSTITUTOS	46
4.3	PROPOSTA DO MODELO DE PREVISÃO	52
5	CONSIDERAÇÕES FINAIS	59
5.1	CONTRIBUIÇÕES DA PESQUISA	60
5.2	LIMITAÇÕES E SUGESTÕES DE FUTURAS PESQUISAS	62
	REFERÊNCIAS	64
	APÊNDICE A	71

1 INTRODUÇÃO

Com o aumento do volume diário de dados gerados em praticamente todos os setores da economia, organizações são forçadas, cada vez mais, a melhorar a análise de seus dados como forma de vantagem competitiva. A extração de informações para embasar a tomada de decisão, a partir de um vasto banco de dados, vem crescendo ao longo dos anos, e se tornando cada vez mais popular (CHEN; CHIANG; STOREY, 2012; COVINGTON, 2016; JAMES et al., 2013; PROVOST; FAWCETT, 2013; SCHLÄFKE; SILVI; MÖLLER, 2013). Com isso, destaca-se a importância da análise dos dados das organizações, para formulações de modelos de previsão que auxiliem na tomada de decisão (SCHLÄFKE; SILVI; MÖLLER, 2013).

Na área de estudos organizacionais, a gestão da demanda é considerada uma das atividades mais relevantes (AYE et al., 2015; COSTANTINO et al., 2015; SUSTROVA, 2016), podendo influenciar a tomada de decisão estratégica em diferentes áreas incluindo *marketing*, aquisições, operações de produção e logística (BALLOU, 2008). A gestão da demanda busca a adequação da operação da organização às necessidades do mercado, a partir de previsões e políticas de gestão de estoque (ACHABAL et al., 2000; ARUNRAJ; AHRENS, 2015; MELO; ALCÂNTARA, 2011). Com a demanda conhecida, é possível tomar decisões mais precisas para melhor atendê-la, auxiliar no posicionamento de mercado e nas interações da cadeia de suprimentos (BALLOU, 2008; SEREL, 2009). Com isso, percebe-se a importância que a gestão da demanda exerce nas organizações, e que no presente trabalho atua como pano de fundo.

Considerando-se a grande diversidade de produtos comercializados em redes varejistas, notam-se certas relações entre produtos, de forma direta ou indireta, possibilitando que determinados produtos possam complementar ou substituir uns aos outros. A relação complementar é caracterizada por produtos que são vendidos em conjunto – quando determinado produto aumenta suas vendas, seu complementar também aumenta. Oposta a essa relação está a substituição – quando a venda de um produto aumenta, o outro tende a diminuir, mesmo que em proporções diferentes. Tais relações devem ser observadas dentro do mesmo

período de tempo, sejam elas diárias ou semanais (MA; FILDES; HUANG, 2016; SHOCKER; BAYUS; KIM, 2004).

Compreender tal relação entre substitutos e complementares é crucial para o gerenciamento do negócio, haja vista o impacto que determinado produto exerce sobre os demais (BANDYOPADHYAY, 2009; GREWAL; ROGGEVEEN; NORDFALT, 2017; MA; FILDES; HUANG, 2016; ZHAO; ZHANG, 2017). Com isso, varejistas são capazes de analisar interações entre as distribuições de vendas dos produtos ao longo do tempo, possibilitando a análise de como eles estão se relacionando, para tomadas de decisões estratégicas e previsões mais assertivas (BANDYOPADHYAY, 2009; BRADLOW et al., 2017; GREWAL; ROGGEVEEN; NORDFALT, 2017; MA; FILDES; HUANG, 2016; PETERS, 2012; RIBEIRO, 2015).

Tais relações de produtos complementares e substitutos podem existir dentro de uma categoria específica (Intra-categoria), a qual é composta somente por produtos similares, como por exemplo, entre leites desnatados e integrais. Além disso, existem relações entre produtos de categorias diferentes (Inter-categorias), como a relação entre macarrão e molho de tomate, os quais estão alocados em diferentes categorias, mas se relacionam (BANDYOPADHYAY, 2009; GELPER; WILMS; CROUX, 2016; MA; FILDES; HUANG, 2016). Vale destacar que, na presente dissertação, “categoria” é entendido como conjunto de itens que possuem semelhanças acerca de sua composição. Deste modo, a interação não ocorre com a mesma intensidade entre produtos dentro e fora das categorias, e compreender tais relações se torna cada vez mais importante (BANDYOPADHYAY, 2009; BRADLOW et al., 2017; GELPER; WILMS; CROUX, 2016; GREWAL; ROGGEVEEN; NORDFALT, 2017; MA; FILDES; HUANG, 2016; RIBEIRO, 2015; SHOCKER; BAYUS; KIM, 2004).

Portanto, a pesquisa almeja verificar a possibilidade de inserção de séries históricas relacionadas a séries de Intra e Inter-categorias em modelos de previsão, como forma de estabelecer melhores resultados de previsão de curto prazo, quando comparados modelos de previsão Holt-Winters. Tal modelo foi utilizado como base de comparação por utilizar somente uma série histórica e não leva em consideração os demais itens relacionados, além de ser vastamente utilizado na literatura (HANSUN, 2017). Para realização do estudo, considera-se que a literatura acerca do

tema tratado dispõe de estudos empíricos que utilizaram a relação Intra e Inter-categorias entre produtos no varejo alimentício, como forma de melhoria da acurácia de modelos de previsão de curto prazo (MA; FILDES; HUANG, 2016; RIBEIRO, 2015).

Por conseguinte, destaca-se a necessidade de identificar, entre os diversos produtos existentes em um supermercado, quais podem influenciar as distribuições de vendas de determinada série focal – assim, a seleção de séries históricas para composição do modelo se faz necessária (HUANG; FILDES; SOOPRAMANIEN, 2014; MA; FILDES; HUANG, 2016; RIBEIRO, 2015). Com o intuito de reconhecer conjuntos de séries que apresentem características ou comportamentos similares em um vasto banco de dados, a seleção de séries históricas os separa em conjuntos homogêneos, para que sejam tratados da mesma maneira, alcançando-se então melhor acurácia na previsão (BOYLAN; SYNTETOS; KARAKOSTAS, 2008; GARCIA; MEDEIROS; VASCONCELOS, 2017; KESTEN C. GREEN, 2012; MA; FILDES; HUANG, 2016; NG, 2007; REIS, 2014; RIBEIRO, 2015). A seleção de séries históricas é indicada quando se deseja trabalhar com um grande número de itens distintos (BOYLAN; SYNTETOS; KARAKOSTAS, 2008; KESTEN C. GREEN, 2012; MA; FILDES; HUANG, 2016; NG, 2007; REIS, 2014), como no caso da presente pesquisa.

Sendo assim, com a finalidade de verificar empiricamente se a inclusão de séries históricas referentes a Intra e Inter-categorias é capaz de melhorar os resultados da previsão de vendas que utilizam somente séries históricas individuais, a pergunta que norteia o trabalho é: *A inclusão de variáveis referentes a Intra e Inter-categorias é capaz de melhorar a acurácia dos modelos de previsão de séries históricas?* Para responder à pergunta, uma rede supermercadista localizada no interior do estado do Espírito Santo foi utilizada para realização do estudo.

1.1 OBJETIVOS

1.1.1 Objetivo Geral

O objetivo geral deste trabalho é verificar se séries históricas referentes a Intra e Inter-categorias são capazes de melhorar o modelo de previsão de vendas de curto prazo para o varejo.

1.1.2 Objetivos Específicos

- a) Confirmar empiricamente a existência de itens complementares e substitutos em Intra e Inter-categorias.
- b) Propor um modelo de previsão que leve em consideração séries de Intra e Inter-categorias.
- c) Comparar os resultados encontrados entre o modelo de previsão com somente uma série histórica e o modelo proposto com Intra e Inter-categorias.
- d) Identificar se existe diferença entre os resultados do modelo com séries Intra e Inter-categorias.

1.2 JUSTIFICATIVA

O tema de previsão de vendas vem sendo amplamente estudado no meio acadêmico e profissional, com o propósito de alcançar melhores resultados no atendimento da demanda. Demonstra-se, assim, a crescente preocupação por parte das organizações no que se refere ao gerenciamento de suas atividades (LO; WANG; LIN, 2008; SCHWARTZ; RIVERA, 2010; TIACCI; SAETTA, 2009; XU et al., 2016; ZOTTERI; KALCHSCHMIDT; CANIATO, 2005).

Neste sentido, do ponto de vista da contribuição conceitual, autores citam a possibilidade de inclusão das variáveis referentes aos produtos complementares e substitutos da Intra e da Inter-categoria, como forma de melhorar a acurácia das previsões (BANDYOPADHYAY, 2009; BRADLOW et al., 2017; GELPER; WILMS; CROUX, 2016; HUANG; FILDES; SOOPRAMANIEN, 2014; MA; FILDES; HUANG, 2016; PETERS, 2012; RIBEIRO, 2015; SONG; CHINTAGUNTA, 2006). Logo, esta

pesquisa não visa apenas realizar previsões de vendas por séries históricas de uma única série, uma vez que já se confirmou sua vasta utilização no campo acadêmico e profissional, nos mais diversos setores (LO; WANG; LIN, 2008; SCHWARTZ; RIVERA, 2010; TIACCI; SAETTA, 2009; XU et al., 2016; ZOTTERI; KALCHSCHMIDT; CANIATO, 2005). Então, espera-se que a pesquisa contribua para a Academia, na tentativa de melhorar os modelos de previsão, e gerar mais dados que possam ser considerados relevantes para a realização de previsões.

Corroborando o exposto, se faz necessário desenvolver um modelo que considere os fatores influenciadores das vendas como forma de melhorar a acurácia da previsão, haja vista que não existe um único modelo que possa ser aplicado para todos os tipos de problemas (ARUNRAJ; AHRENS, 2017). Ademais, observa-se que no varejo supermercadista, diversos produtos comercializados apresentam características, como presença de produtos complementares e substitutos na Intra e Inter-categoria, e também o grande número de unidades para serem gerenciadas. Isso torna a previsão de vendas cada vez mais complexa e desafiadora, fazendo com que o assunto seja de interesse para futuras investigações (ABURTO; WEBER, 2005; ALI et al., 2009; ARUNRAJ; AHRENS, 2015, 2017; GELPER; WILMS; CROUX, 2016; HUANG; FILDES; SOOPRAMANIEN, 2014; MA; FILDES; HUANG, 2016; RIBEIRO, 2015; VAN DONSELAAR et al., 2006).

Neste sentido, a tentativa de realizar previsões no contexto do varejo supermercadista se torna relevante, pela possibilidade de identificar importantes fatores operacionais e gerenciais para o campo teórico e empírico, uma vez constatada a insuficiência de estudos que utilizem itens da Intra e Inter-categorias como forma de melhorar os resultados das previsões (BANDYOPADHYAY, 2009; GELPER; WILMS; CROUX, 2016; MA; FILDES; HUANG, 2016; RIBEIRO, 2015). Assim sendo, a presente pesquisa busca o preenchimento desta lacuna, como oportunidade de contribuição teórico-empírica.

Um estudo com a utilização de técnicas de previsão de vendas, para compreensão dos fatores capazes de influenciar os resultados da previsão, em uma organização varejista, pode vir a contribuir no campo empírico sobre o assunto. Pode oferecer, ainda, contribuições práticas, dado que os resultados esperados, a partir do estudo de caso, podem gerar dados que auxiliem na gestão da organização. Oferece,

assim, contribuições teóricas para pesquisas sobre previsão de vendas (GELPER; WILMS; CROUX, 2016; GREWAL; ROGGEVEEN; NORDFALT, 2017; MA; FILDES; HUANG, 2016; RIBEIRO, 2015).

Para comprovação empírica do estudo, foi escolhido o segmento de varejo, pela sua importância econômica no país. Esse segmento também é um dos mais impactados pela atual crise financeira enfrentada no Brasil (SBVC, 2016). Pesquisas recentes da Sociedade Brasileira de Varejo e Consumo (SBVC) destacam o cenário atual do varejo como um dos piores da história, sofrendo forte desaceleração ao longo dos anos, por conta da crise econômica brasileira. O varejo, que emprega cerca de 19 milhões de pessoas, no ano de 2015 viu as vendas recuarem 8,6%; já em 2016, as vendas recuaram 8,7%, e 2017 não apresenta tendência de reação, podendo apresentar números semelhantes, ou até mesmo piores (SBVC, 2017).

Assim sendo, a previsão de vendas é importante para a gestão e tomada de decisões no varejo, uma vez que em momentos de crise é fundamental perceber novas oportunidades, captar novos clientes e preservar os atuais. Uma previsão acurada é fator-chave para a organização, contribuindo para o controle de falta de produtos e manutenção de estoques de segurança (ALI et al., 2009; BABAI et al., 2013; CHOPRA; MEINDL, 2006; JAIPURIA; MAHAPATRA, 2014; MERSEREAU, 2013; NAGASHIMA et al., 2015; RAMOS; SANTOS; REBELO, 2015; SEREL, 2009). Uma previsão precisa impacta positivamente uma organização, assim como uma imprecisa a impacta negativamente. Uma previsão de vendas imprecisa pode acarretar efeitos negativos diretos na rentabilidade, gestão da organização e posicionamento no mercado (RAMOS; SANTOS; REBELO, 2015; SUSTROVA, 2016).

A aplicação de decisões assertivas auxilia a cadeia a se ajustar às mudanças econômicas e do mercado, para então alinhar suas metas internas, de produção, *marketing* e financeiras por toda a cadeia (WHITTEN; GREEN; ZELBST, 2012). A previsão de vendas permite determinar os níveis apropriados dos produtos utilizados pela organização, sejam eles para produção ou comercialização (ACHABAL et al., 2000). Portanto, impede que ocorra falta de produtos (*understocking*) ou excesso de estoque (*overstocking*), ambos acarretando desperdícios de capital. Especialmente no ramo de varejo alimentício, esses tipos de erros operacionais são mais

impactantes, com o *overstocking*, que provoca a disposição inadequada de produtos, redução dos preços finais e desperdício de produtos; já o *understocking* resulta em perda de vendas, diminuição da confiança do consumidor e possível redução do *market share* (ARUNRAJ; AHRENS, 2015).

Desta forma, a partir de mudanças nos ambientes político e econômico, diversas instituições começaram a utilizar métodos de previsão e de gestão de demanda para se destacar competitivamente em um cenário globalizado, cada vez mais dinâmico e exigente (ABURTO; WEBER, 2005; BERNARDI et al., 2010; LO; WANG; LIN, 2008; SIMCHI-LEVI, 2003; WHICKER et al., 2009). Em períodos turbulentos, a eficiência da gestão da demanda influencia, ainda mais, o futuro de uma organização (AYE et al., 2015; VAN ZELST et al., 2009). Em consonância, o aumento da competitividade no mercado interfere diretamente na performance das organizações, forçando a utilizar todas as oportunidades para melhoria do processo (SUSTROVA, 2016).

1.3 ORGANIZAÇÃO DA DISSERTAÇÃO

Esta dissertação foi dividida em cinco capítulos. O primeiro enfoca a contextualização do tema e como os diversos fatores impactam a previsão de vendas, para então apresentar a pergunta que irá nortear a pesquisa. Seguem-se os objetivos específicos e o geral, para demonstrar o que se pretende alcançar com a realização da pesquisa. Logo depois, apresentam-se a justificativa da pesquisa e relevância do tema.

O segundo capítulo visa realizar uma revisão bibliográfica sobre o tema. O assunto-chave da dissertação inicia este capítulo, que é a previsão de vendas. São então abordados os temas, importância na organização, como ela é realizada e as dificuldades para sua implantação. Este capítulo mostra em detalhes a importância do entendimento dos fatores influenciadores da previsão de vendas e seu impacto para a gestão da organização. Logo após, o assunto dos produtos complementares e substitutos é abordado, demonstrando seu impacto na organização. E assim, visa embasar teoricamente a construção de resultados que estejam em consonância com as necessidades da organização estudada e a literatura.

Para que a execução da previsão seja embasada teoricamente, as seções na sequência explicam os modelos de previsões presentes na literatura, como a previsão se comporta em situações nas quais há um número elevado de variáveis, foco da pesquisa; como a previsão é avaliada em comparação com os dados reais, e como o modelo de regressão foi realizado. Elas esclarecem os itens que têm influência nos resultados da previsão e que são parte integrante do problema de pesquisa. Por fim, demonstrou-se o modelo de pesquisa proposto e como as previsões foram avaliadas.

Na sequência, o terceiro capítulo da dissertação explora o procedimento metodológico que irá guiar a pesquisa, descrevendo quais são as ferramentas metodológicas utilizadas no estudo de caso, e como foram realizados os procedimentos de coleta e análise dos dados.

O quarto capítulo tem por objetivo apresentar os dados, as análises do estudo empírico e, na sequência, discutir os resultados encontrados. Tal sequência visa explorar os dados e discuti-los conceitualmente, como forma de contribuição para a Academia, assim como para o meio gerencial. Por fim, o último capítulo são as considerações finais da dissertação e sugestões para pesquisas futuras.

Todos os capítulos giram em torno de um argumento central: se as séries referentes a Intra e Inter-categoria que compõem o modelo de previsão são capazes de impactar na acurácia da previsão de vendas. A dissertação é baseada em diversos livros e estudos acadêmicos, nacionais e internacionais, atuais e clássicos, com alto índice de citações e fator de impacto. A construção da narrativa levou em consideração o que os autores mais influentes da área estão escrevendo sobre o assunto e os temas mais atuais da área. Ela aborda o tema em sua definição técnica, para que o estudo de caso esteja embasado no que a literatura aborda sobre o tema, em consonância com achados empíricos da realidade a qual se pretende estudar.

2 REVISÃO BIBLIOGRÁFICA

Os conceitos sobre previsão de vendas, produtos complementares e substitutos, Intra e Inter-categorias, modelos quantitativos de previsão utilizados na literatura, previsão com elevado número de variáveis, além de formas de avaliação dos erros de previsões, são apresentados neste capítulo, com vistas a uma construção teórica que esteja em consonância com a proposta de pesquisa.

2.1 PREVISÃO DE VENDAS

A previsão de vendas é considerada uma tarefa essencial para a gestão do negócio no ramo do varejo, podendo impactar de forma direta na gestão do estoque e lucratividade da organização. Os resultados das previsões são considerados importantes *inputs* para diversos tipos de tomada de decisões. A realização de previsões é considerada complexa, devido à grande quantidade de fatores capazes de impactar na demanda (ZLIOBAITE; BAKKER; PECHENIZKIY, 2012).

Diversas são as variáveis que podem afetar a demanda de determinado bem ou serviço, podendo ser a distribuição de renda, preço de bens relacionados, fatores como sazonalidade, investimento em *marketing*, rotinas e preferências dos consumidores e meios de pagamento (VASCONCELLOS, 2011).

A realização de previsões é considerada pré-requisito do planejamento, pois, para o planejamento ser realizado, é necessário obter estimativas dos acontecimentos futuros. Desta forma, a previsão impacta na elaboração do planejamento de curto prazo, produção e vendas, que por sua vez têm o poder de influenciar nas decisões da organização (ARNOLD, 1999; BOWERSOX; CLOSS, 2010; SLACK; CHAMBERS; JOHNSTON, 2007). Logo, se as decisões da organização são influenciadas pelas previsões, e as previsões por sua vez são influenciadas pela demanda, a demanda dita o rumo das operações dentro de uma organização.

Dada a importância das previsões, observa-se a necessidade de escolher os métodos com cautela, para que o mais coerente com o contexto seja selecionado

(ALFTAN; KAIPIA; LOIKKANEN, 2015). Os métodos de previsão de vendas quantitativos são considerados mais eficientes, pelo fato de serem procedimentos indispensáveis para que gestores consigam gerenciar suas organizações de forma mais eficiente e moderna. Além disso, as previsões são essenciais na realização do planejamento estratégico das organizações (AYE et al., 2015). Para que seja possível projeção dos cursos de ações que serão tomadas no futuro, baseado em dados coletados no passado (RAMOS; SANTOS; REBELO, 2015).

Uma previsão de vendas assertiva, que consiga captar as vendas futuras de produtos com características próprias, é considerada uma tarefa complexa e se torna cada vez mais essencial para atuação no mercado (ALFTAN; KAIPIA; LOIKKANEN, 2015; ARUNRAJ; AHRENS, 2015; AYE et al., 2015; NAGASHIMA et al., 2015). Em consonância, gestores convivem constantemente com a dificuldade de realizar previsões assertivas por conta de particularidades dos produtos comercializados, como ciclos de vida curtos e demanda incerta. Com estoques apresentando alto número de produtos perecíveis, redes varejistas convivem diariamente com a incerteza na demanda desses produtos. Já os gestores que ignoram tais características particulares dos produtos enfrentam resultados operacionais e financeiros negativos; e ainda, lidam com a insatisfação do cliente (NAGASHIMA et al., 2015).

Para determinadas operações, é possível prever com certa precisão quais serão os recursos necessários para a realização das tarefas; entretanto, em outros negócios, esse tipo de previsão se torna imprevisível, até mesmo em curtos períodos de tempo (SLACK; CHAMBERS; JOHNSTON, 2007). Em momentos em que a acurácia da previsão não pode ser garantida, gestores tendem a aumentar o volume de estoque, para que o risco de falta de abastecimento seja reduzido, aumentando consequentemente os custos operacionais (NAGASHIMA et al., 2015).

2.2 INTERAÇÃO ENTRE PRODUTOS

Pesquisas realizadas apontam que os fatores que podem influenciar as vendas são passíveis de serem divididos a partir dos mais diversos meios, sendo que alguns são

controlados pela organização, e outros, não. Com isso, alguns dos fatores que influenciam as vendas foram identificados, como sazonalidade, preço, venda de itens substitutos e complementares, características do produto, número de visitas dos clientes à loja, eventos e clima (ARUNRAJ; AHRENS, 2015), conforme Figura 1.

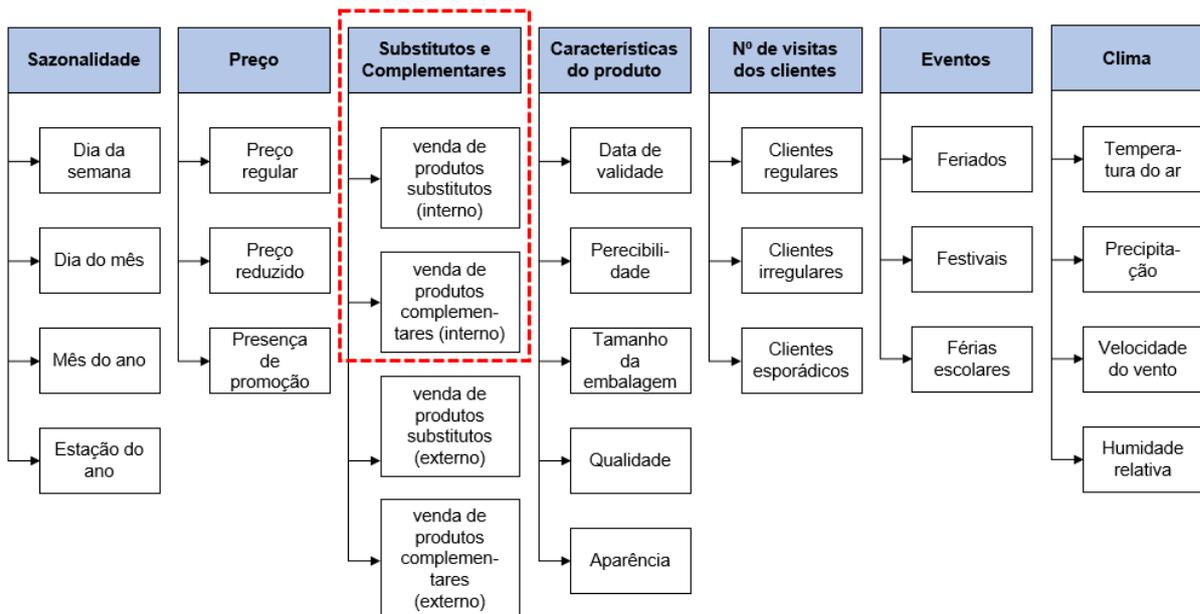


Figura 1 - Segmentação dos fatores que influenciam na demanda.
Fonte: (ARUNRAJ; AHRENS, 2015).

Como a previsão de vendas pode ser influenciada por diversos fatores externos ou internos à organização (ARUNRAJ; AHRENS, 2015; CAMPBELL, 2006; JISANA, 2014; KOTLER; ARMSTRONG, 2000; MA; FILDES; HUANG, 2016; PETER; OLSON, 2009), a presente dissertação foi direcionada para incluir no modelo somente as variáveis referentes às vendas dos produtos complementares e substitutos da organização estudada, conforme destacado na Figura 1. A utilização de todos os fatores inviabiliza a pesquisa, ao passo que a utilização destes dificulta a coleta de dados e execução do modelo, além de exigir mais tempo de pesquisa, o que se caracteriza como uma das limitações da pesquisa.

O processo de comercialização de um produto gera diversos tipos de informações ao varejista, como preço pago, quantidade comprada, *ticket* médio, composição dos produtos comprados, entre outros. A partir da análise desse banco de dados, varejistas conseguem extrair informações de relações entre produtos, possibilitando

analisar como eles estão se relacionando (GREWAL; ROGGEVEEN; NORDFALT, 2017).

Produtores e comerciantes, dos mais diversos segmentos e níveis, estão cientes de que determinados produtos podem complementar ou substituir uns aos outros, compreensão que é de interesse de ambas as partes (BANDYOPADHYAY, 2009). Com o entendimento detalhado da relação entre tais produtos, existe a possibilidade de gerar previsões mais precisas, auxiliando na tomada de decisões gerenciais assertivas e auxílio para decisões estratégicas (BANDYOPADHYAY, 2009; BRADLOW et al., 2017; GELPER; WILMS; CROUX, 2016; GREWAL; ROGGEVEEN; NORDFALT, 2017; MA; FILDES; HUANG, 2016; PETERS, 2012; RIBEIRO, 2015). Entretanto, apesar da existência de estudos relacionados, notam-se poucos trabalhos empíricos acerca do tema, o que gera a necessidade de fazer a literatura evoluir.

Produtos são considerados como complementares quando sua compra é feita em conjunto de outro. Exemplificando, em situações nas quais determinado produto apresenta aumento nas vendas, seu complementar também deve apresentar o mesmo efeito positivo, mesmo que não seja exatamente nas mesmas proporções. O mesmo se aplica aos produtos substitutos, mas de forma inversa. Quando determinado produto vende mais, seu substituto tende a vender menos, apresentando uma relação negativa (MA; FILDES; HUANG, 2016; SHOCKER; BAYUS; KIM, 2004).

Desta forma, emerge a necessidade de classificação dos produtos comercializados em categorias. Nesta pesquisa, entende-se “categoria” como o conjunto de itens que possuem semelhanças em sua composição. A separação das categorias seguiu a que é utilizada pela organização estudada, a qual utiliza a segregação de produtos similares como forma de auxílio para o controle do estoque.

Dito isto, nota-se a possibilidade de haver relações complementares ou substitutivas entre produtos que estão classificados nas mesmas categorias ou categorias distintas. Alguns desses relacionamentos podem ser mais facilmente identificados, como por exemplo, a margarina e a manteiga, que podem ser substitutos ou complementares entre si e estão na mesma categoria (Intra-categoria). Porém, a

identificação dessas relações entre todos os outros produtos comercializados, entre as outras categorias, se mostra uma tarefa árdua, devido à quantidade de produtos distintos. Verifica-se assim uma possibilidade de identificação de efeitos Inter-categoria (BANDYOPADHYAY, 2009; GELPER; WILMS; CROUX, 2016; MA; FILDES; HUANG, 2016; RIBEIRO, 2015).

Além do conceito de interação entre os produtos de forma geral, nota-se que tais relações podem existir entre os produtos de mesma marca, mesmo produto com sabores distintos ou que estão somente presentes na mesma categoria por serem semelhantes, o que é classificado como Intra-categoria. Estudos apontam que esse tipo de relação é mais facilmente encontrado no setor estudado (GELPER; WILMS; CROUX, 2016; MA; FILDES; HUANG, 2016). Mas também existem relacionamentos de um produto de categorias distintas, classificado como Inter-categoria. Esses tipos de relacionamentos são mais fracos em relação à Intra-categoria, e não são observados com grande frequência em pesquisas na área (GELPER; WILMS; CROUX, 2016; MA; FILDES; HUANG, 2016).

A classificação da interação entre produtos em suas categorias foi exemplificada na Figura 2, a qual demonstra a relação entre produtos. No exemplo dado, o produto A está se relacionando com o B; independentemente se for complementar ou substituta, é classificado como Intra-categoria, porque ambos pertencem à mesma categoria. De forma análoga, a relação do produto A e o C é classificada como Inter-categoria, porque ambos pertencem a categorias distintas.

Em vista disso, a presente dissertação se propõe a: (1) identificar a existência de itens complementares e substitutos; não faz parte do escopo, no entanto, explicar o porquê de tais relações; (2) identificar se a interação Intra-categoria é mais relevante para os resultados da previsão do que Inter-categorias. E ainda, (3) identificar se o modelo Intra ou Inter-categoria é capaz de melhorar o modelo de previsão Holt-Winters.

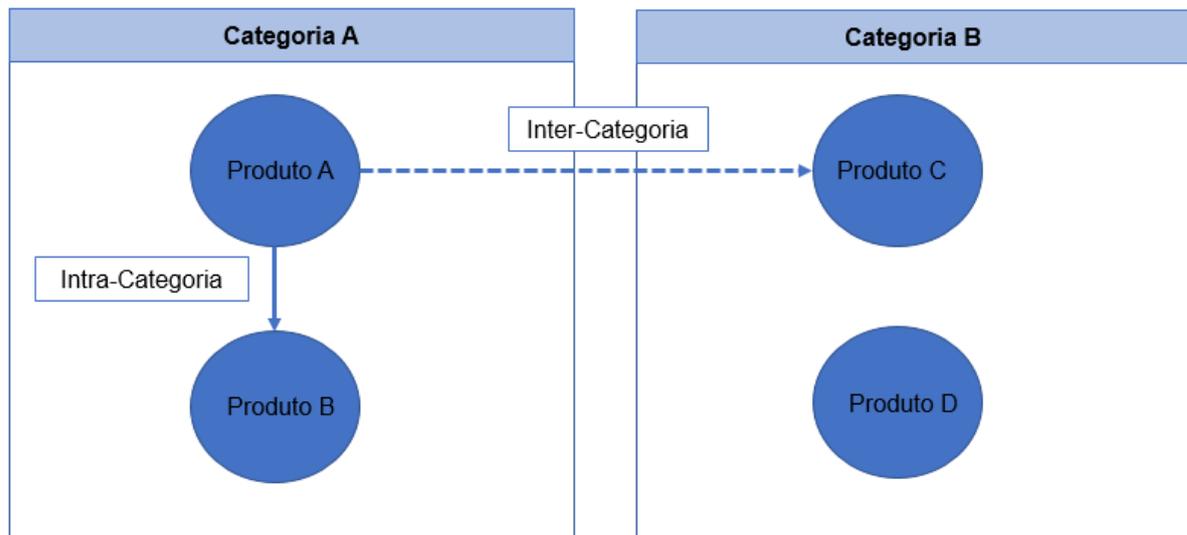


Figura 2 – Relação Intra e Inter-categoria.
Fonte: Próprio autor.

2.3 MODELOS QUANTITATIVOS DE PREVISÃO

Para realizar previsões de vendas, deve-se levar em consideração as três categorias nos quais os métodos são classificados, que variam conforme suas especificações, a saber: métodos qualitativos, de projeção histórica e causais. Dentre os métodos citados, nota-se uma grande quantidade de técnicas de previsão, as quais podem ser utilizadas para os mais variados fins e tipos de dados (BALLOU, 2008).

O Quadro 1 apresenta algumas dessas técnicas, que por sua vez podem ser separadas pelo horizonte de tempo que a previsão é capaz de realizar. O Quadro 1 é organizado de modo decrescente em relação ao horizonte de tempo da previsão. Destacam-se na parte superior as técnicas que executam a previsão de curto prazo até o fim do quadro, que demonstram as técnicas com objetivos de médio-longo prazo (BALLOU, 2008). Além disso, é demonstrado como são classificados os métodos em quantitativos e qualitativos. Sendo que o método quantitativo possui o erro conhecido, já o qualitativo não.

Como cada uma das técnicas demonstradas possui seu horizonte de previsão, nota-se uma diferença entre sua utilização. Por exemplo, as de curto prazo são utilizadas

para tomadas de decisões mais próximas do presente. Já as que executam previsões com horizonte de tempo maior, visando saber o que irá acontecer no futuro em um período de médio-longo prazo, são utilizadas para fins estratégicos e de posicionamento organizacional ao longo dos meses e anos (BALLOU, 2008).

Método	Horizonte de Tempo da Previsão	Métodos
Média Móvel	Curto	Quantitativo
Ponderação exponencial	Curto	Quantitativo
Redes Neurais	Curto	Quantitativo
Estimativas da equipe de vendas	Curto-Médio	Qualitativo
Box-Jenkins	Curto-Médio	Quantitativo
Projeção de Tendência	Curto-Médio	Quantitativo
Modelos de Regressão	Curto-Médio	Quantitativo
Decomposição de séries temporais	Curto-Médio	Quantitativo
Delphi	Médio-Longo	Qualitativo
Pesquisa de Mercado	Médio-Longo	Quantitativo/Qualitativo
Painel de consenso	Médio-Longo	Qualitativo
Analogia Histórica	Médio-Longo	Quantitativo
Intenções de Compra	Médio-Longo	Quantitativo

Quadro 1 - Técnicas de previsão
Fonte: Adaptado de (BALLOU, 2008).

Desta forma, o presente trabalho pretende elaborar um modelo de regressão múltipla, o qual visa conciliar elementos dos métodos de séries históricas. O método de regressão múltipla utiliza dados de outras variáveis que são capazes de explicar a variável a qual se pretende prever, a partir de regressões, o que faz dele um bom método de previsão de curto-médio prazo. Já os modelos de previsão de séries temporais partem da proposição de que os mesmos padrões que ocorreram no passado serão repetidos no futuro, com certa proximidade (ARUNRAJ; AHRENS, 2015; BALLOU, 2008). A escolha deste método em detrimento dos demais se dá pelo horizonte de tempo no qual se deseja realizar as previsões, uma vez se constatou que, para horizontes de tempo reduzidos, esse tipo de abordagem

alcança resultados satisfatórios (GARCIA; MEDEIROS; VASCONCELOS, 2017; MA; FILDES; HUANG, 2016; RIBEIRO, 2015)

O método de regressão utiliza a correlação entre uma variável que se pretende prever, com uma ou mais variáveis que estão se relacionando para gerar o resultado da variável prevista, ou dependente. Observa-se que não é necessário encontrar um alto grau de correlação entre as variáveis para a execução da previsão (BOWERSOX; CLOSS, 2010; LANTZ, 2015). Esse método realiza uma correlação entre os fatores influenciadores da demanda da organização – por exemplo, a relação existente entre as vendas de determinados produtos, as relações entre grupos de produtos e como eles interagem entre si. A análise de regressão trabalha a partir da execução de modelos que tenham a capacidade de representar o relacionamento existente entre as variáveis estudadas (WERKEMA; AGUIAR, 1996). Com isso, esse modelo tem o objetivo de relacionar as variáveis e entender seu impacto na variável dependente (LANTZ, 2015).

Nota-se então a necessidade de entender também o comportamento da série como meio de realizar previsões, haja vista que o modelo de regressão não realiza tal tarefa. Os modelos de séries temporais realizam esse tipo de previsões com base no histórico dos dados das atividades realizadas no passado. Uma vez que se supõe que o histórico passado das vendas é um bom indicador para as vendas futuras e que esta venda será similar ao passado (BOWERSOX; CLOSS, 2010; CHOPRA; MEINDL, 2006). Este método é comumente utilizado para reconhecer variações nos fatores sazonais, padrões cíclicos e tendências (BOWERSOX; CLOSS, 2010).

O método de séries históricas possui certas restrições, assim como todos outros (AYE et al., 2015; BALLOU, 2008; CHOI; YU; AU, 2011; FABIANOVA et al., 2016). Nos casos em que os padrões da série se alteram substancialmente, o modelo passa a não ser capaz de acompanhar tais mudanças, sendo necessário lançar mão de outras abordagens em conjunto com a série histórica (BOWERSOX; CLOSS, 2010; MA; FILDES; HUANG, 2016). Logo, observa-se uma possibilidade de melhoria dos modelos que utilizam somente séries históricas, a partir da utilização conjunta de outras abordagens.

As previsões de vendas são fundamentais para o desempenho organizacional. Sua acurácia depende de um uso adequado dos métodos de previsões existentes para cada tipo de dados, a experiência dos executores e qualidade das informações inseridas no banco de dados. Todos esses fatores são relevantes para sua aproximação da realidade, e afetam de forma relevante os resultados das organizações (ALI et al., 2009; CHOI; YU; AU, 2011; CHU; ZHANG, 2003; FABIANOVA et al., 2016; TANAKA, 2010).

Mediante o exposto, para execução do modelo de previsão, é necessário que se lance mão de uma ferramenta capaz de realizar as previsões, levando em consideração demais variáveis que afetam a distribuição da variável dependente, sem sobrecarregar o modelo. Tal ferramenta é denominada de regressão LASSO e será explicada na próxima subseção.

2.3.1 Modelo de previsão com elevado número de variáveis

A análise dos dados de uma organização pode ir de uma simples observação de como foram as vendas de determinados produtos ao longo do ano, até sistemas complexos de relações entre produtos para basear as estratégias de *marketing*, logística e operações no longo prazo. Este tipo de análise de dados pode extrair informações que não são triviais do cotidiano e fornecer novos *insights* para a gestão do negócio (BRADLOW et al., 2017; FINCH, 2017).

O aumento crescente da disponibilidade e facilidade para criação de novos dados intensifica os desafios das análises dessas informações. Quanto maior o número de variáveis no banco de dados, mais aumenta a complexidade para seleção das que são realmente relevantes (BRADLOW et al., 2017; LEDOLTER, 2013; RIBEIRO, 2015). A seleção de dados é recomendada quando se faz necessário avaliar um grande conjunto de itens distintos, tornando possível a identificação de características semelhantes entre eles. É esperado que os grupos de produtos homogêneos respondam da mesma maneira a fatores influenciadores da demanda (KESTEN C. GREEN, 2012; NG, 2007).

Em situações em que a previsão com apenas uma variável não é suficiente, é necessário lançar mão de modelos mais complexos, como o caso da regressão linear múltipla. Esta estende a função da regressão linear simples, porém utiliza mais de uma variável explicativa, conforme demonstrado na equação (1).

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \varepsilon \quad (1)$$

A maioria das situações do cotidiano que necessitam de análise requer uma elaboração de equações com diversas variáveis. Na regressão linear múltipla, utilizam-se duas ou mais variáveis explanatórias, que são capazes de influenciar a variável dependente y (HILL; JUDGE; GRIFFITHS, 2010). Na regressão, a variável y funciona como a variável dependente, a qual se deseja prever. Já as variáveis x são denominadas como variáveis independentes, explicativas, e são utilizadas para prever y . Os parâmetros β_0 e β_k são os interceptos da inclinação da reta. A variável ε é o erro da operação ao se ajustar aos dados exatamente (HILL; JUDGE; GRIFFITHS, 2010; WERKEMA; AGUIAR, 1996). A questão que emerge desta discussão é: quais variáveis são relevantes para serem inseridas no modelo? Para tanto, um método de seleção de variáveis se faz necessário.

Usualmente, a maioria das variáveis coletadas de um banco de dados pode ser irrelevante para previsão, sendo considerada apenas ruído (LEDOLTER, 2013). Desta forma, determinadas variáveis podem ser identificadas como redundantes, enquanto outras não contribuem para a explicação da variável dependente. Assim sendo, apenas algumas podem ser consideradas realmente importantes. A seleção de quais variáveis irão compor o modelo e quais podem ser retiradas é de suma importância para acurácia da previsão (BRADLOW et al., 2017; GARCIA; MEDEIROS; VASCONCELOS, 2017; LEDOLTER, 2013; MA; FILDES; HUANG, 2016; TIBSHIRANI, 1996).

Com a finalidade de reduzir a quantidade de variáveis no modelo, a regressão LASSO (*Least Absolute Shrinkage And Selection Operator*) é introduzida. Esse tipo de regressão executa a tarefa de encolhimento do conjunto de coeficientes. O algoritmo tem o objetivo de encontrar um modelo que seja capaz de estimar a variável dependente a partir de uma pequena parcela das variáveis independentes. Uma das principais características desse tipo de método é que, durante a execução

do algoritmo de encolhimento, quando existem muitas covariáveis, as que são identificadas como irrelevantes para compor o modelo, elas se anulam, sendo excluídas do modelo, logo, restando apenas as variáveis relevantes. Com isso, o resultado é um método automático de seleção de covariáveis, identificando somente as relevantes que serão parte do modelo (PEREIRA, 2017; TIBSHIRANI, 1996).

O estimador LASSO é definido conforme equação (2), sendo que λ é o parâmetro que controla a quantidade de encolhimento e é determinado por técnicas de validação cruzada. β é um vetor $n \times 1$, $Y = (y_1, \dots, y_n)$ é a variável resposta. Já o X é representado por uma matriz $p \times n$ com as variáveis preditoras. (GARCIA; MEDEIROS; VASCONCELOS, 2017; GELPER; WILMS; CROUX, 2016; JAMES et al., 2013; MA; FILDES; HUANG, 2016; TIBSHIRANI, 1996).

$$\hat{\beta}_{lasso} = \arg \min (\|Y - X\beta\|^2 + \lambda\|\beta\|_1) \quad (2)$$

A utilização do LASSO neste tipo de pesquisa possibilita a análise de uma variável focal, baseada em outras variáveis consideradas importantes para sua previsão. Quando o número de variáveis é maior do que o número de observações, considera-se que os dados possuem grandes dimensões. Nesses casos, não é apropriado utilizar modelos de regressão com todas as variáveis da base, para que se evite o *overfitting* dos dados. Desta forma, se faz necessário escolher variáveis a partir da redução (GARCIA; MEDEIROS; VASCONCELOS, 2017; GELPER; WILMS; CROUX, 2016; JAMES et al., 2013; MA; FILDES; HUANG, 2016; TIBSHIRANI, 1996).

A ferramenta apresentada foi escolhida em detrimento das diversas outras disponíveis, pelo fato de ser capaz de identificar apenas as variáveis mais relevantes para composição do modelo de previsão. Dado que o LASSO é capaz de reduzir o número de variáveis para a regressão múltipla, a ferramenta demonstra ser interessante, pois a análise de poucas variáveis é menos complexa e mais assertiva (MA; FILDES; HUANG, 2016).

É interessante realizar previsões a partir da ferramenta apresentada, graças à possibilidade de se trabalhar inicialmente com um elevado número de séries, sem que se saiba *a priori* qual é a relação existente entre elas. Então o algoritmo resulta no modelo composto somente pelas variáveis relevantes – no caso da presente pesquisa, os produtos que se complementam e se substituem entre si da Intra ou

Inter-categoria. Vale ressaltar que esse tipo de cálculo é inviável manualmente, devido à complexidade e ao volume de dados trabalhados.

Considerando que não é o objetivo da presente dissertação discutir matematicamente o algoritmo LASSO, o mesmo foi inteiramente implementado a partir do *software R*, utilizando-se o pacote *HDeconometrics*. Os pacotes contêm procedimentos computacionais eficientes para o ajuste do modelo, executando grande quantidade de cálculos de forma rápida, robusta e automatizada (SIMON et al., 2011).

2.4 MODELO UTILIZADO NA PESQUISA

A utilização de série histórica é um fator relevante em estudos na área em que se insere a presente pesquisa, por ordenar os dados cronologicamente à medida que eles são gerados, possibilitando uma análise dos eventos reais ocorridos (HAMILTON, 1994). A série histórica pode ser entendida como: “dados que são distribuídos de acordo com o tempo em que se produziram, permanecendo fixos os locais e categorias” (MARCONI; LAKATOS, 2007b, p. 196).

Séries históricas são descritas por processos estocásticos, que podem ser descritos como “processos controlados por leis probabilísticas” (MORETTIN; TOLOI, 2006, p. 19). A série histórica pode ser compreendida como um conjunto de dados que estão ordenados cronologicamente, conforme são produzidos. Ela pode ser descrita como $\{Z_t, t = 1, 2, \dots, n\}$, tal que n se associa a um período de tempo, assim como $n+1$ é seu período precedente e assim por diante (MORETTIN; TOLOI, 2006).

A sazonalidade e a tendência são características, muitas vezes presentes em séries históricas, que estão demonstradas no Gráfico 1. A demanda sazonal é caracterizada pela formação de padrões que se repetem ao longo do tempo. Os dados não necessitam ser necessariamente os mesmos para que uma sazonalidade seja caracterizada, apenas que sejam distribuídos aproximadamente de um valor em comum. Já a tendência é caracterizada por um fator contínuo que interfere nos dados da demanda. No caso apresentado no Gráfico 1, existe um fator que força os dados a aumentarem ao longo do tempo, caracterizando assim, uma tendência

crescente; da mesma forma, ela poderia ser decrescente, ou até mesmo não existir (ARNOLD, 1999).

Na presente pesquisa a tendência e a sazonalidade foram tratadas inteiramente de forma automática pela modelagem computacional a partir do *software R*. Não se fazendo necessário qualquer tratamento especial das séries separadamente.

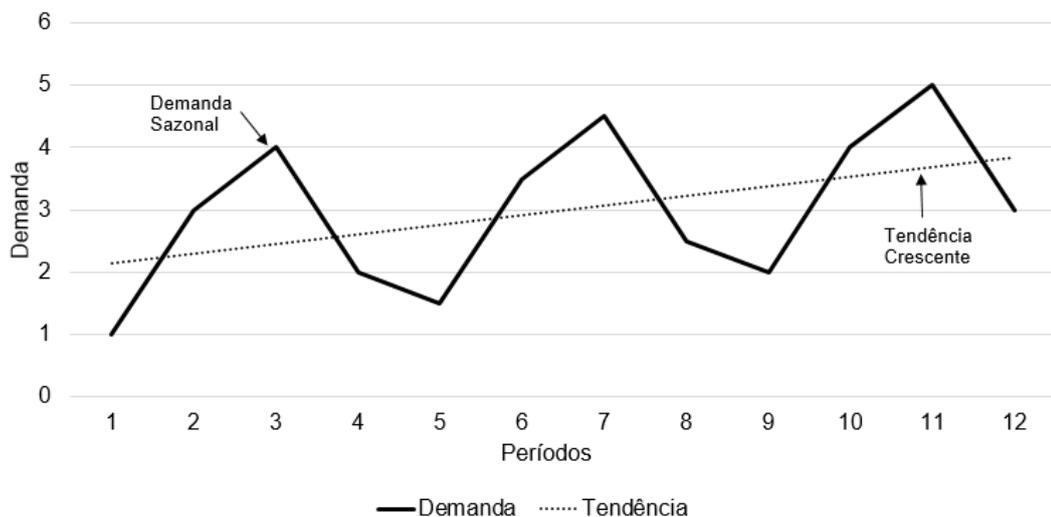


Gráfico 1 – Demonstração de demanda sazonal e tendência ao longo do tempo
Fonte: (ARNOLD, 1999).

De forma geral, um modelo é uma forma simplificada de representação da realidade. Ele é simplificado, baseado em certas premissas sobre o que é e o que não é importante para determinado propósito específico (PROVOST; FAWCETT, 2013). Para o cálculo, o modelo foi composto da série histórica focal, mais as séries históricas identificadas como complementares e/ou substitutas da intra e inter-categorias, a partir do algoritmo Lasso. O modelo leva em consideração fatores capazes de influenciar no resultado da previsão, além de sua própria distribuição de dados no passado. Um único modelo utilizado separadamente, de série histórica ou o de regressão, não é capaz de realizar ambas tarefas (ARUNRAJ; AHRENS, 2015; BALLOU, 2008; MA; FILDES; HUANG, 2016), segundo a Figura 3.

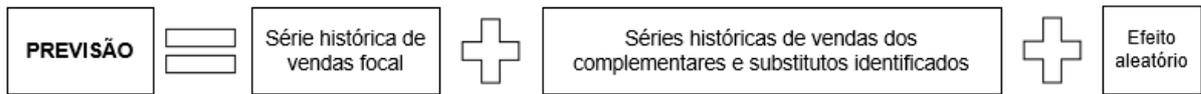


Figura 3 - Modelo proposto de previsão.
Fonte: Próprio autor.

O modelo resultará em uma equação de reta, que é composta pela variável dependente y , representada pela série focal a ser prevista. As variáveis independentes x são as demais séries que compõem a base de dados, identificadas pelo algoritmo. Então, para que o modelo não fique sobrecarregado de variáveis que não são relevantes para a previsão de y , a regressão LASSO foi utilizada para selecionar as variáveis que mais a influenciam. Sendo assim, o modelo é composto apenas por poucas variáveis que influenciam o resultado da previsão, evitando uma sobrecarga de variáveis

A Figura 4 representa o resultado hipotético do modelo proposto, meramente ilustrativo, o qual pode ser identificado no modo como as séries interagem entre si. O mesmo ocorre com suas respectivas influências, sejam elas positivas ou negativas sobre os demais, representadas na cor azul e vermelha, para positivas e negativas, respectivamente.

Exemplificando, o modelo de previsão da série B será composto apenas pelas séries de “C” e “G”, e todas as demais serão descartadas pela regressão LASSO. Essa seleção reduz o tamanho do modelo, facilitando a previsão em situações em que há um número muito elevado de variáveis para serem estudadas (BANDYOPADHYAY, 2009; HUANG; FILDES; SOOPRAMANIEN, 2014; MA; FILDES; HUANG, 2016; RIBEIRO, 2015; SONG; CHINTAGUNTA, 2006; TIBSHIRANI, 1996). Desta forma, o modelo de previsão de “B”, no caso do exemplo dado, será formado pela série histórica de “B”, como variável dependente y , além das séries históricas de “C” e “G”, como variáveis independentes x . Caso o algoritmo não fosse utilizado, o modelo seria composto por todas as séries da base de dados, o que o inviabilizaria.

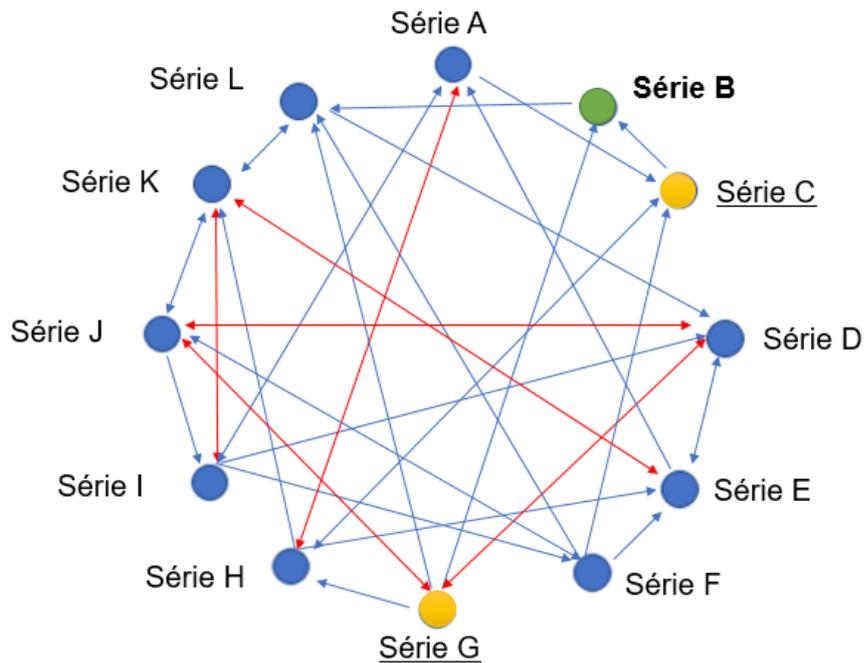


Figura 4 – Interação entre variáveis para composição do modelo de regressão.
 Fonte: Adaptado pelo autor a partir de (MA; FILDES; HUANG, 2016).

Observa-se um avanço na utilização de modelos de séries históricas, pela tentativa de utilizar dados históricos em conjunto com modelos de regressão, que são representados pelos itens complementares e substitutos da Intra e Inter-categorias. Fica então evidente a necessidade de escolher um modelo de regressão que consiga relacionar duas ou mais variáveis para previsão da variável dependente, e que também seja capaz de utilizar as características existentes nas séries históricas. Por esses motivos, a pesquisa pretende realizar tal modelo, de modo a utilizar o melhor dos dois métodos, uma vez que um único modelo separadamente não consegue realizar ambas as tarefas (ARUNRAJ; AHRENS, 2015, 2017; BALLOU, 2008; FABIANOVA et al., 2016).

2.4.1 Avaliação da previsão

Com a finalidade de comparar uma série histórica representada pelo conjunto de observações reais a um modelo de previsão da própria série, surge a questão de se medir o quanto este modelo de previsão é adequado para prever a série original. Para tal, estimativas de erros são utilizadas (PEREIRA, 2017). Existem diversos

meios de estimação da acurácia de uma previsão, e a resposta sobre qual utilizar depende do que está sendo previsto e do tipo de dados que está sendo utilizado (JAMES et al., 2013).

Em geral, para realizar previsões, os dados são separados em duas partes, como forma de mensurar o quanto o modelo consegue se ajustar aos dados reais. A parte inicial da separação, denominada “dados de treino”, é utilizada para treinar, ou “ensinar”, o modelo a estimar os dados futuros – nesta fase é executado o ajuste do modelo.

A segunda parte, denominada “dados de estimação”, é utilizada para mensurar a qualidade da previsão. Esta divisão é demonstrada na Figura 5. A qualidade do ajuste do modelo na fase de treino não é relevante para a análise da acurácia da previsão, e o que realmente importa é como a previsão se realiza na parte de estimação. Nesta parte é que a acurácia da previsão foi mensurada. (JAMES et al., 2013).

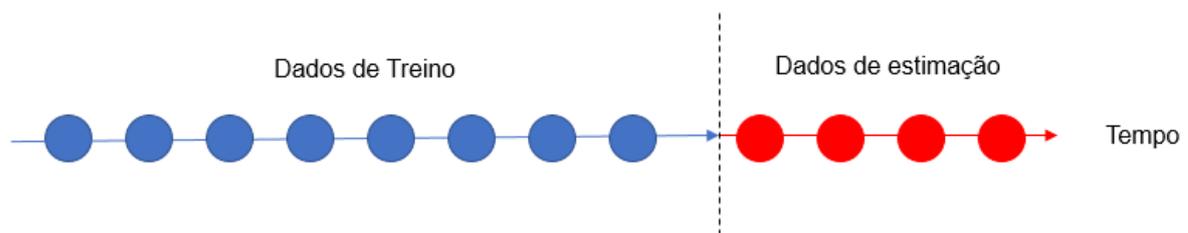


Figura 5 - Separação dos dados nas fases de “treino” e “estimação”.
Fonte: (JAMES et al., 2013).

Desta forma, destaca-se a necessidade de medir a acurácia da previsão em relação aos dados reais. Para realização de tal mensuração, pode ser utilizado medidas na mesma escala dos dados, sendo os mais comuns o MAE (*Mean absolute error*) e o RMSE (*Root Mean Squared Error*). Além disso, medidas de acurácia em percentual também podem ser utilizadas, a mais utilizada é MAPE (*Mean absolute percentage error*). Este tipo de mensuração possui a vantagem de ser independente de escala, facilitando a comparação entre previsões com escalas distintas, porém possui a desvantagem de ser indefinidas se houver observações iguais a zero (HYNDMAN; KOEHLER, 2005), como é o caso dos dados da presente pesquisa.

Ademais, outras medidas de erro são encontradas na literatura, como o *Root Mean Square Error Of Approximation* (RMSEA). O RMSEA mede o ajuste do modelo de equações estruturais, com um índice de correção parcimoniosa que varia de zero até um, penalizando a quantidade de parâmetros estimados (HOOPER; COUGHLAN; MULLEN, 2008). Tal medida não foi utilizada na presente dissertação, optou-se por mensurar o erro de previsão na mesma escala das vendas reais, utilizando-se o RMSE.

Com o intuito de calcular a acurácia da previsão, mensurando o quão próximo uma resposta de previsão está de sua observação real, um dos meios mais comuns, nos casos de regressões, é a utilização do RMSE, dado pela equação (2). É $\hat{f}(x_i)$, a previsão para i observações. Nos casos em que a previsão se aproximar dos valores reais, o valor do RMSE será pequeno e ela será considerada uma “boa” previsão. A mesma analogia é verdadeira para o caso contrário, quando os valores são distantes (JAMES et al., 2013).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{f}(x_i))^2} \quad (2)$$

O resultado do RMSE é apresentado na mesma escala dos dados, sendo popularmente utilizado por sua relevância teórica em estatística (HYNDMAN; KOEHLER, 2005). O RMSE foi o principal valor para se concluir se a inclusão de variáveis referentes a Intra e Inter-categoria foram capazes de melhorar a acurácia da previsão.

3 PROCEDIMENTO METODOLOGICO

O capítulo que se segue procurou expor as abordagens de pesquisa utilizadas e as justificativas da escolha do estudo do caso realizado. Além disso, foi explicado como cada uma das fases da pesquisa foi classificada e realizada, juntamente com a coleta e o método de análise dos dados obtidos.

3.1 ABORDAGEM METODOLÓGICA

A estratégia de pesquisa aqui adotada foi o estudo de caso, que é entendido como sendo uma maneira comum de fazer pesquisa prática e que é bastante utilizado na área de Ciências Sociais e Administração. Sua aplicação é propícia para investigar os fenômenos em profundidade, considerando o contexto em que estão inseridos (YIN, 2010).

No campo científico, o termo “estudo de caso” é utilizado nas mais diversas áreas do conhecimento, o que gera diversas visões sobre o estudo de caso e pouca precisão na utilização do termo (GODOY, 2006). O estudo de caso tem o objetivo de analisar com profundidade determinada unidade social (LAVILLE; DIONNE; SIMAN, 1999; LEITE, 2008), possibilitando também a análise do contexto específico daquele caso e as variáveis que incidem sobre o tema estudado no momento da pesquisa e que podem produzir alterações substanciais no processo (YIN, 2010).

Nesse sentido, nem todo estudo de uma única unidade pode ser considerado estudo de caso, pois além da delimitação do caso, é necessária a aplicação de uma técnica de coleta de dados correspondente ao processo de pesquisa que dê conta daquela situação determinada. Em se tratando de uma realidade dinâmica, uma simples variável pode alterar toda a configuração do problema, assim o estudo de caso pode representar um poderoso instrumento de pesquisa (YIN, 2010).

O estudo do caso da rede supermercadista se enquadra nesse delineamento, por objetivar um estudo em profundidade das vendas da organização, foco da pesquisa, tornando possível propor um modelo de previsão de vendas. Assim sendo, o estudo de caso é conceituado como: “uma investigação que permitirá inicialmente fornecer

explicações no que tange diretamente ao caso considerado e elementos que lhe marcam o contexto” (LAVILLE; DIONNE; SIMAN, 1999, p. 155).

O estudo de caso então é definido como uma investigação empírica que tem por objetivo investigar dado fenômeno dentro de seu contexto, especialmente quando não estão claramente definidos os limites entre o fenômeno e o contexto (YIN, 2010). A principal característica do método estudado é o aprofundamento no contexto da uma rede de varejo supermercadista, que leva em consideração a compreensão do assunto investigado na sua totalidade. A diferença primária entre o estudo de caso e outras alternativas de pesquisa é o foco de atenção do pesquisador, que busca a compreensão de um caso particular em sua complexidade (STAKE, 2011).

Na presente dissertação, verifica-se que a quantidade de informações relevantes sobre o assunto é reduzida, fazendo com que o estudo de caso seja considerado uma parte importante do avanço científico (YIN, 2009). Deste modo, consideram-se a identificação de uma lacuna sobre o assunto estudado e a existência de poucos estudos que auxiliem uma visão geral sobre o tópico (GELPER; WILMS; CROUX, 2016; GREWAL; ROGGEVEEN; NORDFALT, 2017; MA; FILDES; HUANG, 2016; RIBEIRO, 2015). Então, a pesquisa iniciou-se com poucos indícios de quais variáveis poderiam compor o modelo de previsão, considerando-se que a existência de poucos estudos sobre o tema não permite identificá-las *a priori*. Com isso, testou-se se a inclusão de séries históricas referentes a Intra e Inter-categorias pode ser aplicada ao contexto e se pode impactar na acurácia da previsão, assim como a existência de relações complementares e substitutas.

3.2 COLETA E ORGANIZAÇÃO DE DADOS

Organizações produzem uma grande quantidade de dados diariamente, que são gerados para atender determinadas necessidades de tomada de decisão e auxílio ao planejamento estratégico. Esse tipo de informação não é primariamente produzido para servir como fonte de dados de pesquisa, e sim para atender aos interesses das organizações. Porém, são de grande importância para a pesquisa em administração, sendo classificados como “secundários”, por se tratar de uma

construção que não foi realizada pelo pesquisador, e sim por terceiros, em uma prática do cotidiano que resultou na construção de um banco de dados (CATELA; BÊRNI; FERNANDEZ, 2012).

O banco de dados desta pesquisa foi concedido pela organização analisada, e foi construído a partir do seu próprio sistema de vendas. Os funcionários da organização que possuem seus postos de trabalho no *check-out* computam cada venda no sistema, o que permite construir um banco de dados com informações precisas e reais sobre toda comercialização da loja. O banco de dados é organizado a partir da distribuição de vendas diárias de todos os *Stock Keep Units* (SKUs) que foram comercializados, os quais estão dispostos nas colunas, e as linhas são os respectivos dias em que as vendas ocorreram, correspondendo à quantidade comercializada. Desta forma, o pesquisador teve acesso às quantidades de vendas diárias dos SKUs comercializados, juntamente com suas respectivas nomenclaturas específicas.

Ademais, o banco de dados é composto por quatro níveis de separação entre os SKUs comercializados. A maior separação entre eles é o nível de SKU; logo acima, a gramatura representa a quantidade da embalagem comercializada em gramas, quilos, litros, ml e demais unidades. Os níveis de subgrupo e categoria, acima, são meios de separar os itens quanto a suas semelhanças, sendo que a categoria é o primeiro nível de separação entre eles, conforme a Figura 6. Tal separação de itens já foi fornecida *a priori* pela organização.

Com isso, os dados foram codificados pelo pesquisador, a fim de facilitar sua manipulação e análise. O primeiro dígito do código representa a qual categoria o produto pertence; o segundo representa a qual subgrupo ele pertence, e assim por diante, conforme a Figura 6.

Vale ressaltar que o nível da categoria representa a soma de todos os SKUs pertencentes a ela, no nível de subgrupo da mesma forma, até que se chegue ao último nível, com a separação total entre SKUs. Note que um SKU é a série histórica de um único produto específico; já o nível da gramatura representa a soma de todos os SKUs classificados na mesma categoria e subgrupo, porém separados apenas pelas quantidades de suas embalagens. Este nível de separação agrega produtos

semelhantes desconsiderando sua marca ou sabor. Logo a pesquisa seguiu com todos os cálculos posteriores utilizando somente o nível da gramatura, destacado em azul, segundo a Figura 6.

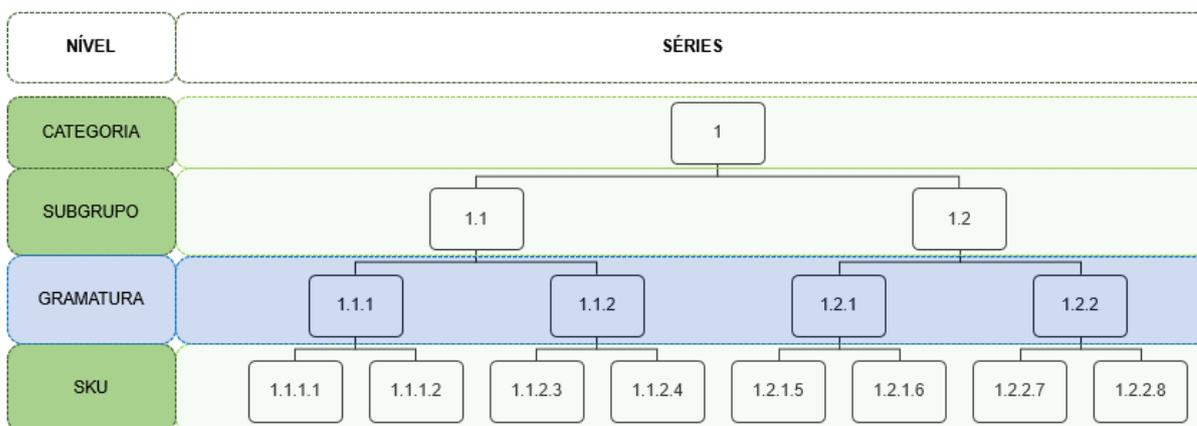


Figura 6 – Classificação do banco de dados por nível.
Fonte: Próprio autor.

Optou-se por utilizar o nível de gramatura em detrimento do SKU individual, pelo fato de existirem muitos itens distintos, e a captação de itens que complementam ou substituem uns aos outros neste nível se torna uma tarefa mais complexa, pela quantidade de variáveis envolvidas. Além do mais, quando produtos específicos são agregados formando o nível de gramatura, possíveis ruídos de produtos específicos são reduzidos, auxiliando na acurácia do resultado final. A utilização de dados desagregados no nível de SKU pode gerar mais ruídos e padrões anormais, dificultando a identificação dos complementares e substitutos na Intra e Inter-categoria, além de dificultar as previsões (RIBEIRO, 2015). Vale ressaltar que, para a série ser incluída no modelo, ela não pode ter mais de 20% de dados faltantes, conforme recomendam alguns autores (MA; FILDES; HUANG, 2016).

Por fim, para a obtenção do banco de dados, todos os contatos realizados com a organização se deram por meio eletrônico. Além disso, com a finalidade de preservar a organização estudada, a presente dissertação tomou o cuidado ético de manter seu nome sob sigilo, assim como ressalta a atenção à privacidade das informações concedidas.

3.3 PROCEDIMENTOS DE ANÁLISE DOS DADOS

A análise dos dados foi dividida em três etapas; na primeira, foram realizados os testes preliminares. Na segunda, os dados foram inseridos no algoritmo para cálculo dos coeficientes do modelo de regressão e identificação dos complementares e substitutos. Por fim, na terceira e última etapa, o modelo de previsão foi proposto.

Na primeira etapa, a análise preliminar dos dados foi realizada para identificar quais categorias poderiam ser mais relevantes quanto ao seu volume de venda, haja vista que a base de dados original é consideravelmente grande para qualquer cálculo manual. Então optou-se por selecionar as categorias com maior impacto no volume de vendas, como forma de estudar produtos de maior relevância para a organização. Para tal, os dados foram analisados utilizando-se o gráfico de Pareto, o qual dispõe a informação com objetivo da priorização, além de permitir a concentração dos esforços nas categorias onde os maiores volumes de vendas são encontrados (WERKEMA, 2006).

Com o resultado do Pareto, quatro categorias foram identificadas como mais impactantes quanto ao volume de unidades vendidas. Dentre elas, foi escolhida a categoria que possuía menos séries para ilustrar os resultados, facilitando a demonstração dos resultados detalhados. Entretanto, os mesmos cálculos foram realizados nas outras três categorias como forma de comparação entre resultados e demonstrados de forma resumida, para que, assim, os achados fossem confirmados a partir de um número considerável de dados.

Logo em seguida, realizou-se a análise das séries a partir de previsões Holt-Winters, com o objetivo de mensurar a qualidade das previsões de modelos que utilizem somente uma série histórica, para então concluir se é possível melhorar a acurácia ou não. Com isso, destaca-se a possibilidade de melhoria; seguiu-se então com a tentativa de inserir os dados referentes a Intra e Inter-categoria no modelo.

A segunda parte da análise de dados se concentrou em realizar os cálculos dos coeficientes de regressão. Para o cálculo, o algoritmo necessita como *input* a variável dependente que foi a série focal, e as demais séries assumem o conjunto de

variáveis independentes, e assim a seleção dos coeficientes da regressão pelo LASSO foi realizada para cada série.

Exemplificando, o procedimento se inicia selecionando a primeira série da matriz como x_1 , a qual se assume como dependente, e as demais colunas, de x_2 até x_n assumem como o conjunto de variáveis independentes. Este processo foi repetido até que todas as séries até x_n fossem tratadas individualmente como x_1 (focal), e calculadas com as demais.

Então, com cada série tratada individualmente como dependente (focal), os resultados são os coeficientes de regressão da série focal cruzadas com as demais séries do conjunto. A leitura deste resultado permite identificar os valores positivos, negativos e zeros, o que deve ser interpretado como as séries se relacionam entre si, possibilitando a identificação de complementares e substitutos.

A terceira e última etapa da análise de dados foi a proposta do modelo de previsão para cada série a partir dos resultados do algoritmo LASSO da etapa anterior. Para tal, o mesmo procedimento de separação dos dados foi realizado para todas as séries trabalhadas. Os dados originais das séries foram divididos inicialmente em dois grupos, o grupo 1 representa os dados utilizados para os cálculos, e por sua vez foram subdivididos em dados para treino e dados para estimação, conforme recomendações de (JAMES et al., 2013). O grupo 2 representa os dados reais da série, os quais foram utilizados para comparar dados reais com os dados previstos e assim, conseguir mensurar a acurácia do modelo.

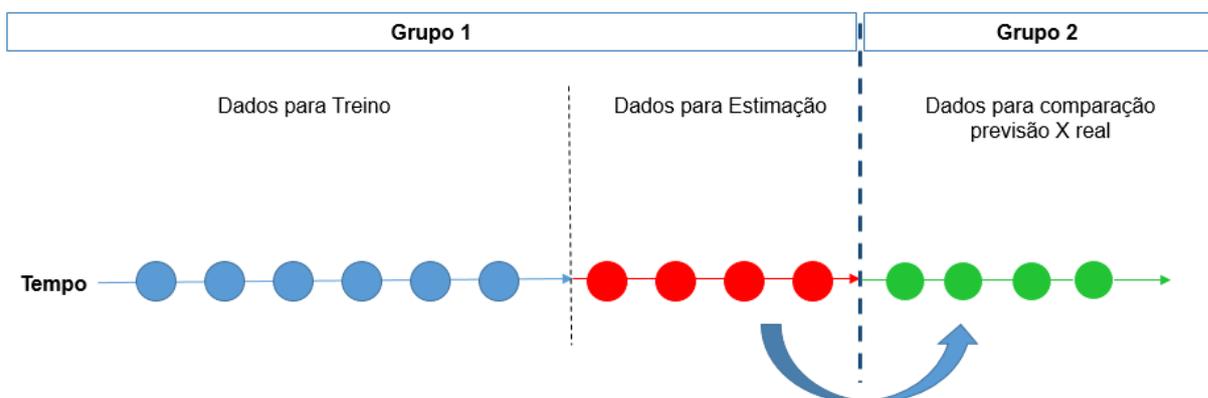


Gráfico 2 – Separação de dados para treino, estimação e mensuração da acurácia da previsão
Fonte: Próprio autor.

Operacionalmente, os dados de treino foram compostos da matriz de séries dependentes, chamados de *x.in* e da série independente que recebeu o nome de *y.in*, estas séries foram utilizadas para realização do cálculo dos coeficientes da regressão LASSO, por meio da função *ic.glmnet*. A mesma lógica foi utilizada nos dados de estimação, onde *x.out* foi a matriz de séries dependentes e *y.out* a série independente. Desta forma, para realização da previsão, utilizou-se a função *predict* com os coeficientes gerados na fase de treino e a matriz de séries dependentes da fase de estimação.

Então foi possível obter a previsão da série focal para ser comparada aos dados reais separados anteriormente, denominados grupo 2, sendo que estes dados não fizeram parte dos cálculos anteriores, serviram somente para fins de mensuração da acurácia da previsão. O algoritmo detalhado pode ser encontrado no Apêndice B – Sumário executivo. Vale ressaltar que toda lógica de cálculo seguiu as recomendações de (GARCIA; MEDEIROS; VASCONCELOS, 2017). Além disso, o teste acf para análise de autocorrelação nas séries foi realizado.

Buscou-se mensurar a acurácia da Intra e Inter-categorias quando comparadas ao Holt-Winters. Para tal realizaram-se seis cenários, sendo o primeiro o Holt-Winters, que serviu de base de comparação com as demais. O segundo foi a Intra-categoria, a qual utiliza somente séries da categoria focal para as previsões. Os outros quatro cenários dizem respeito à relação Inter-categoria, na qual foi testada a categoria focal em conjunto com outras categorias individualmente e, por fim, todas juntas.

Tais cenários Inter-categorias foram realizados porque não há dados que comprovem qual categoria pode auxiliar no resultado da previsão; então decidiu-se testar todos. Desta forma, a execução da análise da relação Intra e Inter-categorias está descrita no Quadro 2, que apresenta, em cada coluna, quais dados foram utilizados para cada cenário, respectivamente.

Todas as previsões foram realizadas para um dia à frente, então foram comparadas com previsões no mesmo horizonte de tempo pelo método de Holt-Winters, como forma de demonstrar se houve redução do erro de previsão, quando comparadas a

um método de previsão difundido na literatura e no meio gerencial. Todas as mensurações de acurácia foram calculadas pela fórmula do RMSE.¹

Como forma de organizar o texto e facilitar o entendimento do leitor, somente os resultados para a categoria com o menor número de séries foi demonstrado detalhadamente. Entretanto, como forma de complementar os resultados, os mesmos cálculos foram realizados para as outras categorias.

	INTRA-CATEGORIA	INTER-CATEGORIAS			
	Categoria Focal	Focal + 2ª Categoria	Focal + 3ª Categoria	Focal + 4ª Categoria	Focal + Todas
	Categoria Focal	Categoria Focal	Categoria Focal	Categoria Focal	Categoria Focal
CATEGORIAS UTILIZADAS NOS CÁLCULOS	-	2ª Categoria	3ª Categoria	4ª Categoria	2ª Categoria
	-	-	-	-	3ª Categoria
	-	-	-	-	4ª Categoria
	-	-	-	-	Categoria

Quadro 2 – Categorias inseridas nos respectivos cálculos.
Fonte: Próprio autor.

Do ponto de vista da execução prática do modelo, e para que as análises fossem confiáveis, o tratamento dos dados foi realizado por meio de *software* estatístico apropriado, o *R*. Este sistema é uma ferramenta muito utilizada para realização de previsões estatísticas e possui vasta utilização ao redor do mundo. Uma de suas características é possuir grande qualidade e abrangência, juntamente com sua gratuidade. Além disso, é notável sua vasta utilização para cálculos estatísticos (R-PROJECT, 2016; REISEN; SILVA, 2011). Por fim, vale ressaltar que todo procedimento para realização dos cálculos estão descritos no sumário executivo no Apêndice B.

¹ Fórmula do RMSE foi descrita no item 2.4.1 - Avaliação da previsão.

4 APRESENTAÇÃO E ANÁLISE DE DADOS

O objetivo deste capítulo foi apresentar o percurso da análise dos dados e demonstrar os resultados encontrados.

4.1 ANÁLISE PRELIMINAR DOS DADOS

A parte inicial da análise se concentrou em identificar as categorias mais relevantes quanto ao volume de unidades vendidas, pelo método de Pareto. Optou-se por utilizar o volume de unidades vendidas, haja vista que, como dito anteriormente, a gestão de estoque atua como pano de fundo da presente pesquisa. Logo, optou-se por utilizar o volume de vendas, por este representar os itens mais movimentados no estoque e, conseqüentemente, os que exigem mais atenção da organização.

Observa-se no Gráfico 3 que as quatro maiores categorias quanto ao volume de vendas são (1) Merceria seca salgada, (2) Merceria líquida, (3) Frios e Laticínios e (4) Merceria Seca e Doce, que são responsáveis por 65,6% do volume de vendas da loja estudada. Sendo assim, tais categorias são consideradas mais relevantes para análise, pelo fato de seu volume de venda ser mais expressivo quando comparado às demais.

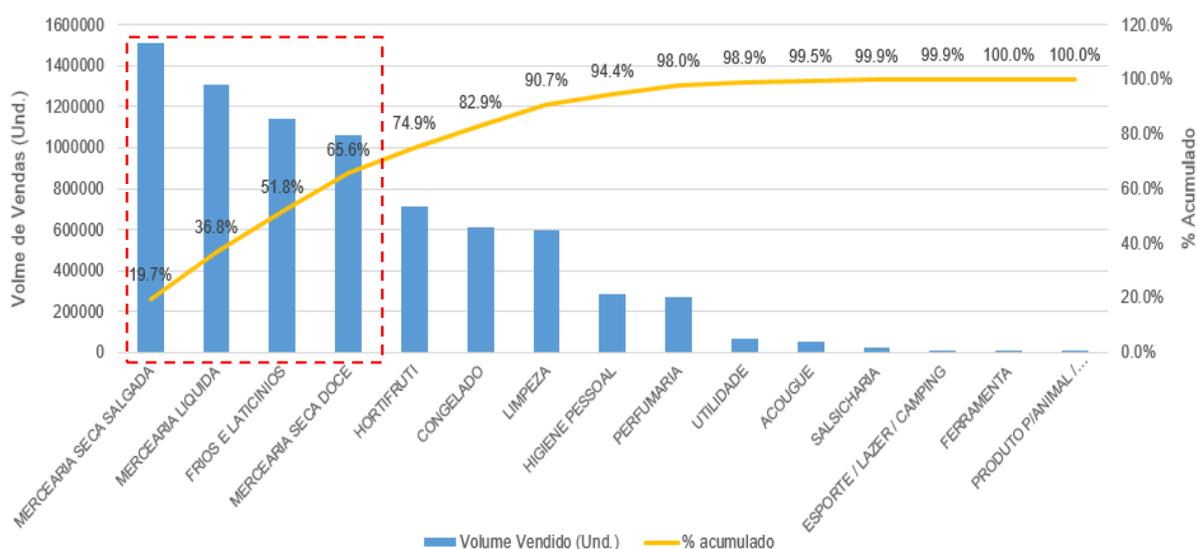


Gráfico 3 - Classificação das categorias pelo método de Pareto.
Fonte: Próprio autor.

Com o intuito de identificar qual seria a categoria mais apropriada para iniciar as análises, demonstra-se na Tabela 1 a quantidade de séries históricas em cada categoria. Nota-se que, a partir da análise das quatro categorias mais impactantes no volume de vendas, “Frios e Laticínios” se destaca com o menor número de séries para serem trabalhadas, o que facilita a análise e a demonstração de resultados. Portanto, deste ponto em diante, a análise segue utilizando apenas a categoria “Frios e Laticínios” no nível de gramatura. A composição específica de cada uma das séries que foram utilizadas está descrita no apêndice A.

CATEGORIA	SÉRIES NO NÍVEL GRAMATURA	SÉRIES NO NÍVEL SKU
Frios e laticínios	12	44
Mercearia seca salgada	111	209
Mercearia seca doce	104	233
Mercearia líquida	85	339
TOTAL	312	827

Tabela 1 - Quantidades de Séries no nível de gramatura e SKU.
Fonte: Próprio autor.

Em seguida, realizou-se a previsão para as séries selecionadas pelo método de Holt-Winters, encontrando resultados não satisfatórios com uma baixa assertividade. Observou-se a possibilidade de melhoria dos resultados encontrados, e iniciou-se a tentativa de inserir séries da Intra e Inter-categorias, para melhorar os resultados encontrados. Os valores encontrados nesta etapa estão descritos na primeira coluna da Tabela 4, mais adiante.

4.2 IDENTIFICAÇÃO DE COMPLEMENTARES E SUBSTITUTOS

Com o intuito de alcançar resultados mais concretos para validar a pergunta de pesquisa, ao invés de separar uma amostra dentro do universo da categoria na qual foi trabalhada, optou-se por utilizar todas as séries pertencentes à categoria “Frios e Laticínios” para realização do cálculo dos coeficientes de regressão. A entrada de dados para os cálculos foi uma matriz composta por 12 séries históricas no nível de gramatura, que estão dispostas nas colunas. Cada uma dessas colunas possui 158

observações, que representam a quantidade de vendas diárias entre os dias 01/07/2016 até 31/12/2016.

Realizou-se assim a seleção da primeira série da matriz que foi tratada como variável dependente e as outras 11 séries que assumem como variáveis independentes. Este processo foi repetido até que todas as séries fossem tratadas individualmente como dependentes, e cruzadas com as demais. Este procedimento resultou em uma matriz com todos os coeficientes de regressão indicando quais produtos complementam e substituem uns aos outros na Intra-categoria, descrito na Tabela 2. Com isso demonstra-se a relação entre as séries, sendo que os valores positivos, em verde, representam os complementares, já os negativos, em vermelho, representam os substitutos. Nota-se que em diversos pontos da matriz não existe valor, o que significa que não existe interação entre as séries.

A leitura correta da Tabela 2 deve ser realizada a partir das colunas. A coluna representa qual série é a focal para a análise, e as demais linhas representam os resultados dos coeficientes de regressão para respectivas séries, podendo-se identificar como as séries interagem entre si. A diagonal principal da matriz é o intercepto da regressão, que significa a interseção da reta quando x é igual a zero. Sendo assim, os achados indicam que as interações entre as séries são assimétricas, por não serem observados os mesmos valores entre as séries e vice-versa. Da mesma forma que algumas séries são mais relacionadas com as demais, algumas se demonstram mais isoladas.

Dentre as interações encontradas na Tabela 2, destacam-se algumas interações que sobressaíram em relação às demais, como a relação complementar entre a bebida láctea – 200ml (1.1.68) e a manteiga – 500g (1.6.54), além de sua substituição pela margarina especial – 1kg (1.8.61). O leite lv integral - 1 litro (1.3.88) se destacou por ser complementar do leite lv desnatado - 1 litro (1.2.88) e leite lv semidesnatado - 1 litro, (1.4.88). Além disso, foi encontrada relação substitutiva com a manteiga 200g (1.6.34), manteiga 500g (1.6.54) e margarina cremosa - 250g (1.7.39).

A fim de facilitar o entendimento, foi demonstrada no Quadro 3 a legenda dos códigos apresentados na Tabela 2. O primeiro dígito representa a categoria, o

segundo o subgrupo e por fim, a gramatura. A composição específica de cada código no nível de SKU está descrita no Apêndice A.

	1.1.68	1.1.88	1.2.88	1.3.88	1.4.88	1.6.34	1.6.54	1.7.39	1.7.54	1.7.61	1.8.54	1.8.61
1.1.68	46.05	0.04	0.12	-0.22	0.02	0.09	0.01	0.01	0.06	0.01	-0.01	-
1.1.88	1.04	23.83	0.46	-2.25	-	-0.22	-0.02	0.17	-	-	-	-
1.2.88	0.28	-	-14.85	3.85	-	0.32	-	-	0.06	-	0.02	-
1.3.88	-	-	0.09	-5.54	0.08	-0.05	-	-	0.01	-	-	-
1.4.88	-	-	-0.14	4.47	17.28	0.18	-	0.04	0.07	-	0.02	-
1.6.34	0.93	-	0.65	-3.53	0.13	16.03	-	-0.03	0.05	-	0.03	-
1.6.54	4.17	-	-	-3.57	-	-0.73	4.95	-0.19	1.02	-	0.40	-
1.7.39	2.00	1.11	0.14	-3.22	0.96	-0.90	-0.03	-0.98	-	0.03	0.15	-
1.7.54	1.10	-	0.18	0.77	0.30	0.19	0.05	-	8.60	-	-0.07	-
1.7.61	1.59	-	-	-	-	0.48	-	0.21	-	9.81	0.10	-
1.8.54	-	-	0.89	-	0.33	0.56	0.10	0.18	-	-	4.25	0.02
1.8.61	-2.70	-	2.24	3.30	-	0.74	-	-0.36	-	-	0.75	1.56

Tabela 2 – Matriz de coeficientes da regressão LASSO para demonstração de complementares e substitutos.

Fonte: Próprio autor.

CÓDIGO	CATEGORIA	SUBGRUPO	GRAMATURA
1.1.68	Frios e Laticínios	Bebida láctea	200 ml
1.1.88	Frios e Laticínios	Bebida láctea	1 litro
1.2.88	Frios e Laticínios	Leite lv desnatado	1 litro
1.3.88	Frios e Laticínios	Leite lv integral	1 litro
1.4.88	Frios e Laticínios	Leite lv semi-desnatado	1 litro
1.6.34	Frios e Laticínios	Manteiga	200 gramas
1.6.54	Frios e Laticínios	Manteiga	500 gramas
1.7.39	Frios e Laticínios	Margarina cremosa	250 gramas
1.7.54	Frios e Laticínios	Margarina cremosa	500 gramas
1.7.61	Frios e Laticínios	Margarina cremosa	1 Kg
1.8.54	Frios e Laticínios	Margarina especial	500 gramas
1.8.61	Frios e Laticínios	Margarina especial	1 Kg

Quadro 3 – Legenda de códigos da composição da categoria, subgrupo e gramatura de cada.

Fonte: Próprio autor.

A fim de facilitar o entendimento das interações entre as séries apresentados na Tabela 2, optou-se por demonstrar de forma gráfica tais interações. As interações na

cor azul, representam os complementares, já na cor laranja, estão representados os substitutos. Com isso, é possível identificar que algumas séries possuem mais interações com as demais, e estão identificadas no centro do gráfico. Já as séries periféricas, possuem menos interações. Isso significa que qualquer alteração nestas séries periféricas, o impacto nas demais será reduzido, quando comparado as centrais.

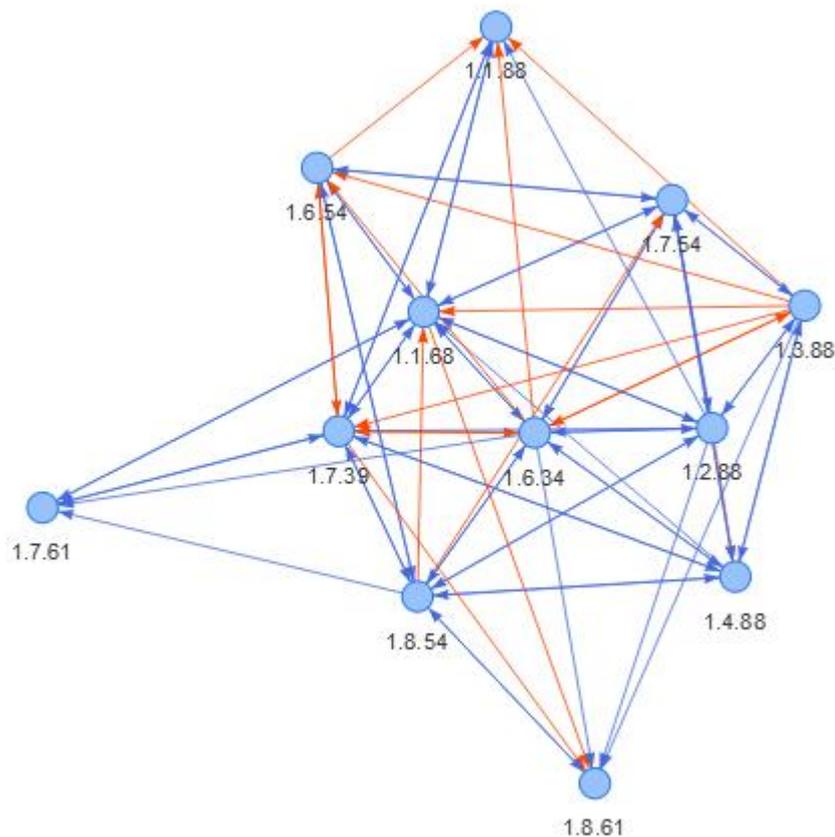


Figura 7 – Diagrama de relacionamento entre séries complementares e substitutas.
Fonte: Próprio autor.

Outra forma de demonstrar a relação entre as séries é por meio da matriz de correlação, demonstrada na Figura 8, a qual demonstra todas as correlações existentes entre as séries históricas. Este tipo de análise varia de zero até um, onde zero significa correlação nula, e um é a correlação perfeita, a qual é encontrada na diagonal principal, por ser a correlação da própria série com ela mesma. Com isso, nota-se que os resultados encontrados na matriz corroboram com os achados demonstrados anteriormente pelo algoritmo LASSO. As séries com as correlações

mais elevadas, o modelo LASSO foi capaz de identifica-las e seleciona-las para compor o modelo. Exemplificando, no caso da série 1.1.88, as séries com maiores correlações são 1.7.39 e 1.1.68, e nos resultados da Tabela 2, demonstrada anteriormente, somente elas foram selecionadas para o modelo.

	1.1.68	1.1.88	1.2.88	1.3.88	1.4.88	1.6.34	1.6.54	1.7.39	1.7.54	1.7.61	1.8.54	1.8.61
1.1.68	1	0.58	0.78	0.54	0.65	0.72	0.52	0.57	0.77	0.46	0.36	0.19
1.1.88	0.58	1	0.47	0.23	0.31	0.32	0.1	0.66	0.41	0.36	0.21	0.01
1.2.88	0.78	0.47	1	0.8	0.75	0.72	0.46	0.49	0.79	0.37	0.48	0.37
1.3.88	0.54	0.23	0.8	1	0.86	0.43	0.35	0.38	0.69	0.23	0.38	0.38
1.4.88	0.65	0.31	0.75	0.86	1	0.54	0.42	0.51	0.73	0.27	0.44	0.35
1.6.34	0.72	0.32	0.72	0.43	0.54	1	0.44	0.28	0.65	0.36	0.41	0.29
1.6.54	0.52	0.1	0.46	0.35	0.42	0.44	1	0.14	0.57	0.27	0.37	0.22
1.7.39	0.57	0.66	0.49	0.38	0.51	0.28	0.14	1	0.42	0.38	0.34	0.06
1.7.54	0.77	0.41	0.79	0.69	0.73	0.65	0.57	0.42	1	0.36	0.33	0.33
1.7.61	0.46	0.36	0.37	0.23	0.27	0.36	0.27	0.38	0.36	1	0.28	0.17
1.8.54	0.36	0.21	0.48	0.38	0.44	0.41	0.37	0.34	0.33	0.28	1	0.33
1.8.61	0.19	0.01	0.37	0.38	0.35	0.29	0.22	0.06	0.33	0.17	0.33	1

Figura 8 – Matriz de correlação das séries históricas da categoria Frios e Laticínios.
Fonte: Próprio autor.

A discussão conceitual acerca dos motivos das interações, assim como a explicação sobre cada uma delas não faz parte do escopo da pesquisa, e foi considerada uma possibilidade de exploração em pesquisas futuras de marketing ou áreas relacionadas com o varejo. O entendimento dessas relações pode servir como base para a tomada de decisões na organização no que diz respeito à escolha de produtos para realização de promoções ou mudanças de layout da loja.

Pesquisas apontam a existência de relações complementares e substitutas na Intra e Inter-categorias (GELPER; WILMS; CROUX, 2016; MA; FILDES; HUANG, 2016). Com isso, a fim de confirmar tal afirmação, decidiu-se testar a existência de tais relações entre as categorias individualmente, podendo-se confirmar a existência de

relações complementares e substitutas na Intra-categoria, conforme a Tabela 3. Os resultados da Tabela 3 demonstram, dentre os coeficientes calculados, quantos foram complementares e substitutos, encontrando-se assim, os percentuais para cada cenário, Intra-categoria.

Além de confirmar os achados de Gelper; Wilms; Croux (2016) e Ma; Fildes; Huang (2016), destaca-se nos resultados da Tabela 3 que há maior predominância da relação complementar, em relação ao substituto, na Intra-categoria. A presente pesquisa não buscou o aprofundamento do tema, pois não faz parte do escopo encontrar os motivos das relações, e sim destacar sua existência. Com isso, demonstra-se a existência de uma das limitações da pesquisa, assim como sugestões para estudos futuros.

TIPOS DE RELAÇÃO	INTRA-CATEGORIA				MÉDIA
	FRIOS	SALG.	DOCE	LIQ.	
Complementar	76.1%	93.7%	93.6%	91.9%	88.8%
Substituto	23.9%	6.3%	6.4%	8.1%	11.2%

Tabela 3 – Relação complementar e substituta Intra e Inter-categorias.
Fonte: Próprio autor.

Por fim, com os coeficientes de regressão gerados, pode-se identificar que o modelo se ajusta de forma satisfatória aos dados, conforme demonstrado no Gráfico 4. Exemplificando com a série 1.1.68, está demonstrado na cor vermelha como o modelo de regressão LASSO se ajusta as vendas reais, demonstrados na cor preta. É possível identificar que o ajuste do modelo é capaz de acompanhar a distribuição dos dados reais, acompanhando as variações diárias.

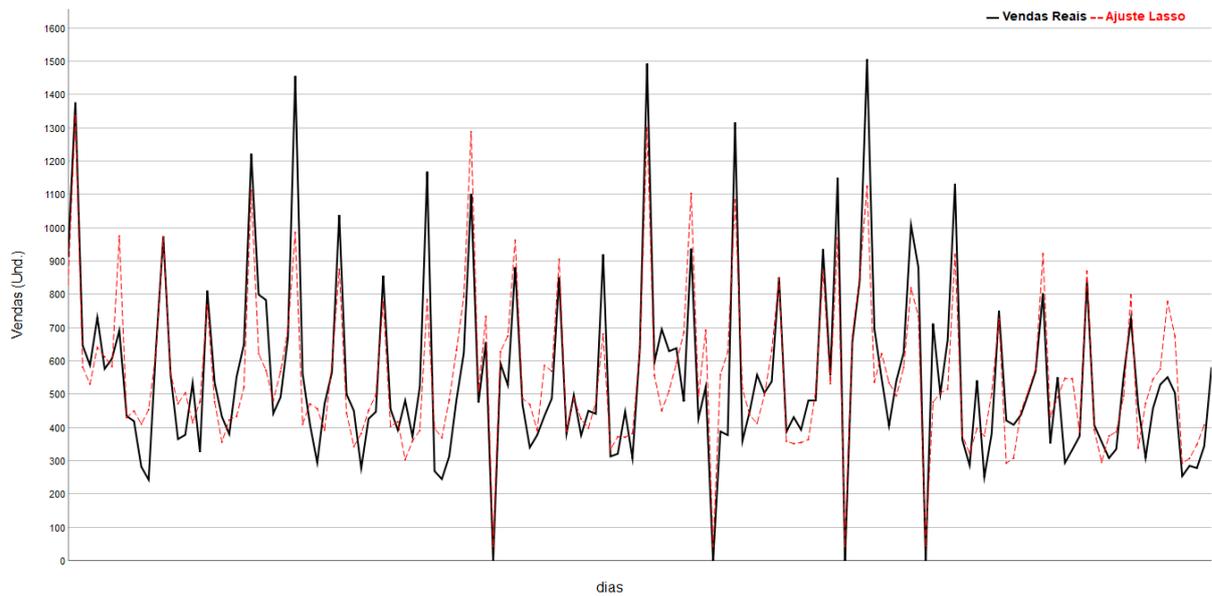


Gráfico 4 – Demonstração do ajuste do modelo de regressão Lasso aos dados reais.
Fonte: Próprio autor.

4.3 PROPOSTA DO MODELO DE PREVISÃO

Para desenvolver a proposta do modelo de previsão, foram realizados seis cenários, a fim de identificar a existência de influência de itens Intra e Inter categóricos. Todos os cenários calculados foram então comparados com o Holt-Winters, como forma de identificar se houve melhoria na acurácia. A composição de cada cenário foi descrita no Quadro 4. O primeiro cenário foi composto apenas pelas séries de Frios e Laticínios, representando a Intra-categoria. Para representar a Inter-Categoria, foram realizados mais quatro cenários, sendo compostos pela categoria focal, frios e laticínios, com cada uma das demais e, por fim, todas juntas.

	INTRA CATEGORIA	INTER CATEGORIA			
	FRIOS	FRIOS + SALGADA	FRIOS + DOCE	FRIOS + LIQUIDA	FRIOS + TODOS
	Frios e laticínios	Frios e laticínios	Frios e laticínios	Frios e laticínios	Frios e laticínios
CATEGORIAS UTILIZADAS NOS CÁLCULOS	-	Mercearia seca salgada	Mercearia seca doce	Mercearia liquida	Mercearia seca salgada
	-	-	-	-	Mercearia seca doce
	-	-	-	-	Mercearia liquida

Quadro 4 – Cenários das categorias inseridas nos respectivos cálculos.
Fonte: Próprio autor.

Na sequência, os resultados demonstrados na Tabela 4 são os valores da acurácia das previsões para um período à frente $t+1$, medidos a partir do RMSE. Na primeira coluna estão as respectivas séries no nível de gramatura, nas quais as previsões foram realizadas. A segunda coluna apresenta os resultados da acurácia para o cenário utilizando o método Holt-Winters, que serviu de base de comparação com as demais. Na sequência, a terceira coluna demonstra os resultados para o cenário com dados Intra-categoria, que pertencem somente a “Frios e Laticínios”. Os demais cenários são os resultados das relações Inter-categorias. Na sequência, apresentaram-se os percentuais de redução do RMSE, quando calculados a partir da Intra e Inter-categorias, respectivamente. Por fim, as últimas duas colunas destacam quais foram os percentuais de contribuição para a redução RMSE, a fim de identificar quem foi o principal fator para melhoria da acurácia – Intra ou Inter-categoria.

Os resultados da Tabela 4 apontam que as previsões que utilizam séries da Intra e Inter-categorias conseguem reduzir, na maioria dos casos, o erro da previsão Holt-Winters. Os valores destacados na cor verde são os menores resultados encontrados para todos os cenários, na sua respectiva linha. Demonstra-se assim, a melhora da acurácia quando séries da Intra e Inter-categorias foram inseridas. Nota-se também que existem casos nos quais a utilização de dados somente da Intra-categoria resulta no melhor resultado, porém informações Inter-categorias também contribuem para a melhora da acurácia.

Seguindo com a análise da influência das informações Intra e Inter-categorias no resultado final da acurácia da previsão, apresentam-se na Tabela 4 os resultados do percentual de RMSE que foi reduzido nas colunas “% de Redução”. Suas duas colunas demonstram, respectivamente, os percentuais de redução alcançados, quando comparados com o resultado base do Holt-Winters para ambos os casos, Intra e Inter-categoria. Exemplificando para a primeira linha, o “Intra-categoria” é calculado pela fórmula $1 - (\text{INTRA CAT.} / \text{Holt-Winters})$, resultando em 75% de redução do valor base de 1218,65. O caso da coluna “Inter-categoria” é calculado pela fórmula $1 - (\text{INTER CAT.} / \text{Holt-Winters})$, resultando em 87% de redução do valor de 1218,65. Note-se que é utilizado o menor valor encontrado em Inter-categoria.

De forma resumida destaca-se que, em média, a Intra-categoria foi capaz de reduzir o RMSE em 43,3%, ao passo que a Inter-categoria foi capaz de reduzi-lo em 57,5%.

Por fim, as duas últimas colunas, denominadas “Contribuição”, representam o quanto cada um dos casos foi capaz de contribuir para a redução do modelo Holt-Winters. Buscou-se identificar, entre os resultados obtidos, qual foi a origem da melhoria da acurácia da previsão. Assim sendo, destaca-se que em média, para a categoria “Frios e Laticínios”, 85,8% da contribuição para reduzir o RMSE foi proveniente da Intra-categoria, e a Inter-categoria contribuiu, em média, com somente 14,2%. Nos casos em que a Inter-categoria foi capaz de reduzir o valor do RMSE, a redução foi discreta, o que nos leva a concluir que as informações mais relevantes para a redução do erro são provenientes da Intra-categoria, e o Inter-categoria somente auxilia, contribuindo com uma parcela menor no resultado final.

Em determinados casos isolados, a maior parcela de redução foi proveniente de itens Inter categóricos, como é o caso do “1.7.61”. Ademais, existem situações em que a inserção de informações referentes a Intra ou Inter-categorias não foi capaz de melhorar a acurácia, obtendo-se o resultado oposto, piorando a acurácia, como é o caso de “1.8.61”. Por fim, vale ressaltar que o teste acf para identificar autocorrelação, foi realizado, e não foi encontrado autocorrelação nas séries. Sendo assim, um dos pressupostos para esse tipo de análise foi atendido, o de não haver autocorrelação nas séries para realização da regressão.

SÉRIES	HOLT-WINTERS	INTRA-CAT.	INTER-CATEGORIA				% DE REDUÇÃO		CONTRIBUIÇÃO
		FRIOS	FRIOS + SALG.	FRIOS + DOCE	FRIOS + LIQ.	FRIOS + TODOS	INTRA-CATEGORIA	INTER-CATEGORIA	INTER-CATEGORIA
1.3.88	1218.65	307.16	158.12	408.76	310.88	178.07	75%	87%	12%
1.8.54	9.18	2.56	0.19	2.54	8.26	2.51	72%	98%	26%
1.6.34	149.97	46.75	42.32	66.57	78.7	59.2	69%	72%	3%
1.7.39	20.06	6.41	8.7	2.88	12.38	7.08	68%	86%	18%
1.2.88	326.85	115.7	90.12	91.92	143.69	111.8	65%	72%	8%
1.4.88	237.11	100.91	118.09	98.02	148.96	125.5	57%	59%	1%
1.1.68	379.04	171.75	210.33	244.36	219.29	215.82	55%	45%	-10%
1.7.54	90.04	45.46	34.22	38.45	53.19	45.67	50%	62%	12%
1.6.54	12.11	7.59	7.88	6.68	8.96	7.36	37%	45%	8%
1.1.88	38.22	31.41	24.17	18.85	32.85	20.08	18%	51%	33%
1.7.61	9.58	10.64	8.38	9.24	10.58	7.33	-11%	23%	35%
1.8.61	1.05	1.41	2.01	1.85	1.14	2.14	-34%	-9%	26%
MÉDIA							43.3%	57.5%	14.2%

Tabela 4 - Comparação entre resultados para previsão Holt-Winters e LASSO.
Fonte: Próprio autor.

Como forma de comprovar a capacidade de redução do erro de previsão quando informações referentes a Intra-categoria são comparadas ao método Holt-Winters, realizou-se a mesma operação de Frios e Laticínios, utilizando-se apenas a Intra-categoria das outras três identificadas como mais representativas para a organização. Destaca-se então que, entre as 312 séries no nível de gramatura previstas, as informações referentes a Intra-categoria foram capazes de reduzir o erro de 83,7% das séries. Além do mais, nos casos em que houve diminuição, o RMSE foi reduzido em média 56,3% quando comparado com o RMSE do Holt-Winters, conforme a Tabela 5.

CATEGORIA	TOTAL DE VARIÁVEIS	% DE SÉRIES MELHORADAS INTRA-CATEGORIA	MÉDIA DE REDUÇÃO RMSE
Frios e laticínios	12	83.3	56.5
Mercearia seca salgada	111	90.1	61.5
Mercearia seca doce	104	88.5	54.4
Mercearia líquida	85	72.9	52.9
MÉDIA	312	83.7	56.3

Tabela 5 - Comparação entre resultados da Intra-categorias.
Fonte: Próprio autor.

Além dos achados da contribuição expressiva da Intra-categoria para redução do RMSE quando comparado com o método Holt-Winters, a presente dissertação buscou confirmações mais concretas para confirmar os achados. Com isso, o mesmo cálculo constante da Tabela 4 foi realizado para as demais categorias. Então, 20 cenários diferentes foram calculados, a fim de reforçar os achados anteriores. Demonstrou-se, no Quadro 5, a composição de cada cenário calculado. Na coluna identificada como Intra-categoria, foram utilizadas somente as séries da categoria focal, e nas colunas identificadas como Inter-categorias, foram identificadas a categoria focal, além de outra categoria; e por fim todas em conjunto, conforme o Quadro 5.

	INTRA-CAT.	INTER-CATEGORIA			
		CATEGORIAS UTILIZADAS NOS CÁLCULOS	Frios	Frios + Salgada	Frios + Doce
Salgada	Salgada + Frios		Salgada + Doce	Salgada + Líquida	Salgada + Todos
Doce	Doce + Frios		Doce + Salgada	Doce + Líquida	Doce + Todos
Líquida	Líquida + Frios		Líquida + Doce	Líquida + Salgada	Líquida + Todos

Quadro 5 – Cenários para cálculos Intra e Inter-categorias.

Fonte: Próprio autor.

Assim pode-se concluir, a partir de 20 cenários de previsão, incluindo Intra e Inter-categorias, os quais foram compostos por aproximadamente duas mil previsões distintas, que a Intra-categoria é capaz de contribuir em média com 83,8% das melhorias do RMSE, ao passo que a Inter-categoria contribui discretamente com 16,2%, conforme a Tabela 6.

CATEGORIAS	% DE REDUÇÃO		CONTRIBUIÇÃO	
	INTRA-CATEGORIA	INTER-CATEGORIA	INTER-CATEGORIA	INTRA-CATEGORIA
Frios e laticínios	43.3%	57.5%	14.2%	85.8%
Mercearia seca salgada	47.6%	56.9%	9.3%	90.7%
Mercearia seca doce	31.3%	43.8%	12.5%	87.5%
Mercearia líquida	12.6%	41.5%	28.9%	71.1%
MÉDIA	33.7%	49.9%	16.2%	83.8%

Tabela 6 – Contribuição da Intra e Inter-categorias para redução do RMSE.

Fonte: Próprio autor.

Esses resultados se aproximam consideravelmente dos achados de Ma; Fildes; Huang (2016), os quais encontraram uma contribuição de cerca de 95% da Intra-

categoria. No entanto, para alcançar tais resultados, foram utilizadas diversas outras variáveis, as quais não foram utilizadas na presente pesquisa. Este fato pode ser entendido como uma limitação de pesquisa e sugestão para pesquisas futuras.

Destaca-se também que, mesmo não levando em consideração outras variáveis, como preço ou localização do produto, a presente pesquisa foi capaz de alcançar resultados satisfatórios e semelhantes. Isso comprova a eficácia da utilização de modelos de regressão como forma de melhoria de previsão de vendas no varejo supermercadista.

Pode-se concluir que a ferramenta apresentada para seleção de coeficientes para realização de previsões demonstra resultados satisfatórios, podendo melhorar a acurácia da maioria das séries, quando comparada a métodos vastamente difundidos na literatura.

5 CONSIDERAÇÕES FINAIS

A presente dissertação teve como objetivo geral verificar se séries referentes a Intra e Inter-categorias são capazes de melhorar o modelo de previsão de vendas de curto prazo para o varejo. E ainda, pretende confirmar empiricamente se o comportamento entre as séries afeta o resultado final da previsão, quando comparado ao realizado somente por séries históricas.

Desta forma, o primeiro resultado destacado é a comprovação da existência de itens complementares e substitutos em Intra e Inter-categorias. Além disso, os resultados demonstraram maior prevalência de itens complementares no nível de gramatura, o que representa em média 88,8% das interações, sendo os demais 11,2% substitutos.

Na sequência, para alcançar o resultado final, foram realizadas aproximadamente duas mil previsões por meio de modelagem computacional, com o objetivo de confirmar a proposta de pesquisa. Os resultados apontam que 83,8% da contribuição para redução dos resultados do RMSE são provenientes das séries Intra-categoria, o que representa a maioria expressiva da contribuição. Entretanto, o estudo destaca que séries Inter-categorias também são capazes de contribuir com 16,2% para melhorar a acurácia, demonstrando assim uma redução do erro e comprovando a existência de interação entre séries ao longo das categorias. No entanto, para se alcançar tal redução discreta, proveniente da Inter-categoria, é necessário utilizar um volume significativamente maior de dados, o que aumenta o tempo de processamento do modelo.

Ademais, destaca-se que o modelo de previsão composto apenas por séries da Intra-categoria mostrou-se capaz de reduzir o RMSE de 83,7% das 312 séries no nível de gramatura previstas. Dentro deste percentual, a redução média do RMSE foi de 56,3%.

Conclui-se que a utilização de séries pertencentes apenas a Intra-categoria para compor o modelo de previsão consegue melhorar a acurácia na maioria dos casos, e que a redução alcançada atinge resultados satisfatórios. Isso confirma a proposta de que a interação Intra-categoria é capaz de melhorar o modelo de previsão Holt-

Winters, corroborando os achados de Ma, Fildes e Huang (2016). A inclusão de muitas variáveis que não pertencem à categoria pode auxiliar no resultado final, mas não de forma expressiva. Assim, os resultados encontrados a partir da pesquisa empírica confirmam as propostas apresentadas.

5.1 CONTRIBUIÇÕES DA PESQUISA

Baseando-se nos resultados obtidos e analisados nesta dissertação, evidenciam-se contribuições para o campo da Administração de forma teórica e empírica. Sob a ótica da contribuição teórica, observa-se a evolução da literatura, no que diz respeito à compreensão da relação dos produtos complementares e substitutos na Intra e Inter-categoria, para utilização em modelos de previsões. E ainda, a possibilidade de incluir tais variáveis em modelos de previsão demonstrou que, quando inseridas nos modelos, elas são capazes de melhorar a acurácia.

Ademais, foi demonstrado, por meio de diversos testes, que a relação Intra-categoria foi mais impactante para a redução dos erros de previsão do que as Inter-categorias. A utilização de um alto número de previsões para comprovação do resultado final representou a contribuição teórica desta investigação para auxiliar futuras pesquisas acerca do tema. A literatura sobre a relação entre categorias se limita a testes diretos entre marcas ou produtos específicos, e o presente estudo vai além, ao testar diversas possibilidades de interação.

Outra constatação de contribuição teórica foi a colaboração da pesquisa para incremento da literatura, uma vez que se identificou uma lacuna teórica: há poucos estudos que levam em consideração a relação de itens complementares e substitutos na Intra e Inter-categoria, visando a melhoria de previsões no varejo alimentício.

Observa-se necessidade crescente de extração de informações relevantes a partir de vastos bancos de dados. Com isso, emerge a necessidade de se escolher bons preditores em meio a muitas variáveis, muitas vezes irrelevantes, o que traz dificuldades ao modelo para selecionar as variáveis “corretas” na realização da previsão. Desta forma, destaca-se outra contribuição teórica – a utilização do

algoritmo LASSO, que é vastamente utilizado no campo da estatística e matemática, porém pouco utilizado na área de estudos organizacionais e *marketing*. Esse algoritmo se mostrou um bom método de análise, com potencial para incrementar ainda mais a literatura. Emerge então outra contribuição teórica: o desenvolvimento de um método para utilizar os produtos complementares e substitutos da Intra e Inter-categoria, para alcançar resultados mais acurados na previsão de vendas no varejo alimentício, e que possivelmente podem ser utilizados em outros segmentos.

No que tange à parte empírica, as contribuições se deram no âmbito das possibilidades de usos e interpretações dos resultados obtidos. O estudo de itens complementares e substitutos da Intra e Inter-categoria afeta a eficiência de decisões gerenciais sobre o *marketing* e a gestão de estoque.

No que se refere à influência sobre o *marketing*, decisões mais assertivas podem ser tomadas com o objetivo de promover um produto e influenciar outro, sabendo-se da complementaridade sobre outro, promovendo-o assim de forma indireta. Outra possibilidade é o reconhecimento dos complementares e substitutos, a fim de identificar concorrentes e parceiros, além de possíveis vulnerabilidades entre produtos, para que ações proativas de *marketing* sejam realizadas.

Em geral, redes varejistas podem não estar interessadas na relação entre um produto ou marca específica, e sim como todas elas se relacionam entre si, para que ações mais assertivas sejam realizadas e se alcance maior lucratividade entre as categorias comercializadas pela loja. Desta forma, os achados da pesquisa possibilitam a avaliação de possíveis concorrentes, o que pode se tornar uma informação importante para a tomada de decisões estratégicas. Para fins gerenciais, decisões mais assertivas para gestão de categorias podem ser tomadas a partir do modelo proposto, uma vez que ficou demonstrada a possibilidade de automatização do processo para um grande volume de dados, podendo-se identificar relações entre categorias e séries específicas.

No que tange à gestão de estoque, a melhoria da previsão de vendas pode influenciar positivamente de diversas maneiras. Com a previsão acurada das vendas, a gestão de estoque pode se planejar para alocar recursos necessários, para que não ocorram *stock-out* ou *overstocking*. *Stock-out* é a falta de produtos

para serem comercializados, que acarreta perda de potenciais vendas, resultando na insatisfação do cliente. Casos como *overstocking*, para produtos alimentícios com vencimentos curtos, são ainda mais impactantes, pois acarretam necessidade de descarte de produtos e resultam em perdas financeiras. Sendo assim, resolver tal questão, tornando esse tipo de problema cada vez menos recorrente, passa por uma boa previsão de vendas.

Estudos apontam que clientes insatisfeitos com a loja por conta de *stock-outs* estão mais propensos a mudar seus hábitos de consumo, transferindo suas compras para outras lojas (HUANG; FILDES; SOOPRAMANIEN, 2014). Isso comprova a importância do controle preciso das previsões, não só para controle de estoque no curto prazo, mas também para o posicionamento estratégico da organização no mercado de atuação para médio e longo prazo. Ademais, a previsão de vendas exerce papel central na gestão do estoque. A identificação das vendas futuras é a informação base para a realização de reabastecimentos, investimento em armazenamento e alocação de recursos para distribuição.

5.2 LIMITAÇÕES E SUGESTÕES DE FUTURAS PESQUISAS

No desenvolvimento da presente pesquisa, limitações para sua conclusão foram encontradas. Uma delas foi a utilização de dados de uma única loja, o que permite conclusões somente para a loja estudada, sem expandi-las às demais da rede, uma vez que estas se encontram em bairros e até mesmo cidades distintas.

Assim, uma das sugestões para pesquisas futuras é expandir a base de dados para mais de uma loja, e empregar o mesmo método da presente dissertação nas demais, para que se possam concluir com mais robustez os achados. Tal questão não desqualifica a pesquisa, uma vez que, para o estabelecimento estudado, os itens que representam aproximadamente 70% do volume de vendas foram analisados.

Pesquisas futuras que utilizarem dados de diversas organizações, até mesmo de diferentes estados ou regiões do país, podem realizar inferências estatísticas que proporcionem uma visão mais apurada acerca do tema. Este tipo de análise mais

ampla permite ao pesquisador determinadas estratificações dos resultados para identificar semelhanças e diferenças nas diversas regiões ou quanto ao porte da organização estudada. Esse tipo de análise pode ser interessante para uma única organização que esteja presente em diversas regiões, como o caso de grandes redes varejistas, como forma de posicionamento estratégico no mercado e controle geral de estoque.

Quanto ao caráter metodológico, a pesquisa também apresentou limitações na análise qualitativa. Se este tipo de análise tivesse sido realizado juntamente com os gestores da organização, resultados explicativos no que diz respeito à relação dos complementares e substitutos poderiam ter sido alcançados. Tal análise não foi realizada por questões de prazo de pesquisa e localização geográfica do estabelecimento, o qual se encontra fora da região metropolitana de Vitória. Mesmo com a presença de tais restrições, o estudo seguiu com contatos por meio eletrônico, com a organização enviando todos os dados necessários para realização do estudo.

A pesquisa não levou em consideração a relação de complementares e substitutos existentes entre as marcas e até mesmo sabores, o que pode ser interessante do ponto de vista gerencial, considerando a possibilidade de existir margens de lucro diferentes para as diversas marcas comercializadas. Ademais, o modelo apresentado não levou em consideração outros fatores passíveis de influenciar nas vendas de determinado produto, como preço do produto focal, preço de seus complementares e substitutos, presença de promoções, tipos de clientes, clima, entre outros diversos fatores que podem influenciar na distribuição final das vendas. A inclusão desses fatores se mostra uma possível forma de melhorar a acurácia da previsão de vendas.

Tendo em vista que a presente pesquisa se trata de uma dissertação acadêmica, não fez parte do escopo do trabalho apresentar uma implementação prática do projeto. Entretanto, o *script* utilizado para realização de todos os cálculos no *software R* está descrito no apêndice B.

REFERÊNCIAS

- ABURTO, L.; WEBER, R. Improved supply chain management based on hybrid demand forecasts. **Applied Soft Computing Journal**, v. 7, n. 1, p. 136–144, 2005.
- ACHABAL, D. et al. A decision support system for vendor managed inventory. **Journal of Retailing**, v. 76, n. 4, p. 430–454, 2000.
- ALFTAN, A.; KAIPIA, R.; LOIKKANEN, L. Centralised grocery supply chain planning: Improved exception management. **International Journal of Physical Distribution & Logistics Management**, v. 45, n. 3, p. 237–259, 2015.
- ALI, Ö. G. et al. SKU demand forecasting in the presence of promotions. **Expert Systems with Applications**, v. 36, n. 10, p. 12340–12348, 2009.
- ARNOLD, T. **Administração de materiais: uma introdução**. 1. ed. São Paulo, SP: Atlas, 1999.
- ARUNRAJ, N. S.; AHRENS, D. A hybrid seasonal autoregressive integrated moving average and quantile regression for daily food sales forecasting. **International Journal of Production Economics**, v. 170, p. 321–335, 2015.
- ARUNRAJ, N. S.; AHRENS, D. Improving Food Supply Chain using Hybrid Semiparametric Regression Model. **Supply Management Research**, n. October, p. 213–238, 2017.
- AYE, G. C. et al. Forecasting aggregate retail sales: The case of South Africa. **International Journal of Production Economics**, v. 160, p. 66–79, 2015.
- BABAI, M. Z. et al. Forecasting and inventory performance in a two-stage supply chain with ARIMA(0,1,1) demand: Theory and empirical analysis. **International Journal of Production Economics**, v. 143, n. 2, p. 463–471, 2013.
- BALLOU, R. H. Necessidades de Previsão da cadeia de suprimentos. In: BALLOU, R. H. (Ed.). **Gerenciamento da cadeia de suprimentos/logística empresarial**. 5. ed. Porto Alegre: Bookman, 2008. p. 241–267.
- BANDYOPADHYAY, S. A dynamic model of cross-category competition: theory, tests and applications. **Journal of Retailing**, v. 85, n. 4, p. 468–479, 2009.
- BERNARDI, A. C. et al. Análise e melhoria do processo de avaliação dos impactos econômicos, sociais e ambientais de tecnologias da Embrapa pecuária sudeste. **Gestão da Produção**, v. 17, n. 2, p. 297–316, 2010.
- BOWERSOX, D.; CLOSS, D. **Logística empresarial: O processo de interação da cadeia de suprimento**. São Paulo, SP: Atlas, 2010.
- BOYLAN, J. E.; SYNTETOS, A. A.; KARAKOSTAS, G. C. Classification for Forecasting and Stock Control: A Case Study. **The Journal of the Operational Research Society**, v. 59, n. 4, p. 473–481, 2008.

BRADLOW, E. T. et al. The Role of Big Data and Predictive Analytics in Retailing. **Journal of Retailing**, v. 93, n. 1, p. 0–53, 2017.

CAMPBELL, C. Eu compro, logo sei que existo: as bases metafísicas do consumo moderno. In: BARBOSA, L.; CAMPBELL, C. (Eds.). . **Cultura, consumo e identidade**. Rio de Janeiro: FGV, 2006.

CATELA, E. Y. S.; BÊRNI, D. A.; FERNANDEZ, B. P. M. Como fazer levantamento de dados. In: BÊRNI, D. DE A.; FERNANDEZ, B. P. M. (Eds.). . **Métodos e técnicas de pesquisa: modelando as ciências empresariais**. São Paulo, SP: Saraiva, 2012. p. 173–205.

CHEN, H.; CHIANG, R.; STOREY, V. C. Business Intelligence And Analytics : From Big Data To Big Impact. **Management Information Systems Quarterly**, v. 36, n. 4, p. 1165–1188, 2012.

CHOI, T. M.; YU, Y.; AU, K. F. A hybrid SARIMA wavelet transform method for sales forecasting. **Decision Support Systems**, v. 59, n. 1, p. 130–140, 2011.

CHOPRA, S.; MEINDL, P. **Gerenciamento da cadeia de suprimentos: estratégia, planejamento e operação**. São Paulo, SP: Pearson Prentice Hall, 2006.

CHU, C.-W.; ZHANG, G. P. A comparative study of linear and nonlinear models for aggregate retail sales forecasting. **International Journal of Production Economics**, v. 86, n. 3, p. 217–231, 2003.

COSTANTINO, F. et al. SPC forecasting system to mitigate the bullwhip effect and inventory variance in supply chains. **Expert Systems with Applications**, v. 42, n. 3, p. 1773–1787, 2015.

COVINGTON, D. **Analytics: Data Science, Data Analysis and Predictive Analytics for Business**. 5. ed. [s.l.] CreateSpace, 2016.

FABIANOVA, J. et al. Using a software tool in forecasting: A case study of sales forecasting taking into account data uncertainty. **Open Engineering**, v. 6, n. 1, p. 270–279, 2016.

FINCH, V. **Big Data For Business: Your Comprehensive Guide To Understand Data Science, Data Analytics and Data Mining To Boost More Growth and Improve Business**. [s.l.] Auva Press, 2017.

GARCIA, M. G. P.; MEDEIROS, M. C.; VASCONCELOS, G. F. R. Real-time inflation forecasting with high-dimensional models: The case of Brazil. **International Journal of Forecasting**, v. 33, n. 3, p. 679–693, 2017.

GELPER, S.; WILMS, I.; CROUX, C. Identifying Demand Effects in a Large Network of Product Categories. **Journal of Retailing**, v. 92, n. 1, p. 25–39, 2016.

GODOY, A. S. Estudo de caso qualitativo. In: SILVA, A. B. DA; GODOI, C. K.; BANDEIRA-DE-MELLO, R. (Eds.). . **Pesquisa qualitativa em estudos organizacionais: paradigmas, estratégias e métodos**. São Paulo, SP: Saraiva, 2006. p. 115–143.

GREWAL, D.; ROGGEVEEN, A. L.; NORDFALT, J. The Future of Retailing. **Journal of Retailing**, v. 93, n. 1, p. 1–6, 2017.

HAMILTON, J. D. Difference equation. In: HAMILTON, J. D. (Ed.). . **Time series analysis**. Princeton: Princeton university press, 1994. p. 1–25.

HANSUN, S. New estimation rules for unknown parameters on holt-winters multiplicative method. **Journal of Mathematical and Fundamental Sciences**, v. 49, n. 2, p. 127–135, 2017.

HILL, C.; JUDGE, G.; GRIFFITHS, W. O modelo de regressão múltipla. In: HILL, C.; JUDGE, G.; GRIFFITHS, W. (Eds.). . **Econometria**. 3. ed. São Paulo, SP: Saraiva, 2010. p. 167–190.

HOOPER, D.; COUGHLAN, J.; MULLEN, M. R. Structural Equation Modelling: Guidelines for Determining Model Fit. **The Electronic Journal of Business Research Methods**, v. 6, n. 1, p. 53–60, 2008.

HUANG, T.; FILDES, R.; SOOPRAMANIEN, D. The value of competitive information in forecasting FMCG retail product sales and the variable selection problem. **European Journal of Operational Research**, v. 237, n. 2, p. 738–748, 2014.

HYNDMAN, R. J.; KOEHLER, A. B. Another Look at Measures of Forecast Accuracy. **International journal of forecasting**, v. 22, n. November, p. 679–688, 2005.

JAIPURIA, S.; MAHAPATRA, S. S. An improved demand forecasting method to reduce bullwhip effect in supply chains. **EXPERT SYSTEMS WITH APPLICATIONS**, v. 41, n. 5, p. 2395–2408, 2014.

JAMES, G. et al. **An Introduction to Statistical Learning**. New York: Springer Texts in Statistics, 2013.

JISANA, T. K. Consumer Behaviour Models: An Overview. **Sai Om Journal of Commerce & Management**, v. 1, n. 5, p. 34–43, 2014.

KESTEN C. GREEN, J. S. A. Demand forecasting: evidence-based methods. **International Journal of Forecasting**, n. October, p. 1–27, 2012.

KOTLER, P.; ARMSTRONG, G. **Introdução ao marketing**. 4. ed. Rio de Janeiro: LTC, 2000.

LANTZ, B. **Machine learning with R**. 2. ed. Birmingham, UK: Packt Publishing Ltd, 2015.

LAVILLE, C.; DIONNE, J.; SIMAN, L. M. As estratégias de verificação. In: LAVILLE, C.; DIONNE, J.; SIMAN, L. M. (Eds.). . **A construção do saber: manual de metodologia da pesquisa em ciências humanas**. Belo Horizonte, MG: UFMG, 1999. p. 131–161.

LEDOLTER, J. **Data Mining And Business Analytics With R**. Hoboken, New Jersey: University of Iowa, 2013.

LEITE, F. T. **Metodologia científica: métodos e técnicas de pesquisa: monografias, dissertações, teses e livros**. 2. ed. Aparecida, SP: Idéias & Letras, 2008.

LO, S.; WANG, F.; LIN, J. T. Forecasting for the LCD monitor market Forecasting for the LCD Monitor Market. **Journal of Forecasting**, v. 27, n. 4, p. 341–356, 2008.

MA, S.; FILDES, R.; HUANG, T. Demand forecasting with high dimensional data: The case of SKU retail sales forecasting with intra- and inter-category promotional information. **European Journal of Operational Research**, v. 249, n. 1, p. 245–257, 2016.

MARCONI, M. DE A.; LAKATOS, M. E. **Técnicas de pesquisa**. 6. ed. São Paulo, SP: Atlas, 2007.

MELO, D. D. C.; ALCÂNTARA, R. L. C. A gestão da demanda em cadeias de suprimentos : uma abordagem além da previsão de vendas. **Gestão & Produção**, p. 809–824, 2011.

MERSEREAU, A. J. Demand Estimation from Censored Observations with Inventory Record Inaccuracy. **Manufacturing & Service Operations Management**, v. 17, n. 3, p. 335–349, 2013.

MORETTIN, P. A.; TOLOI, C. M. C. **Análise de séries temporais**. 2. ed. São Paulo, SP: Edgard Blucher, 2006.

NAGASHIMA, M. et al. Impacts of adaptive collaboration on demand forecasting accuracy of different product categories throughout the product life cycle. **Supply Chain Management**, v. 20, n. 4, p. 415–433, 2015.

NG, W. L. A simple classifier for multiple criteria ABC analysis. **European Journal of Operational Research**, v. 177, n. 1, p. 344–353, 2007.

PEREIRA, L. **Geometria Dos Métodos De Regressão Lars, Lasso E Elastic Net Com Uma Aplicação Em Seleção Genômica**. [s.l.] Universidade Federal de Lavras, 2017.

PETER, J. P.; OLSON, J. C. **Comportamento do consumidor e estratégia de marketing**. 8. ed. São Paulo, SP: McGraw-Hill, 2009.

PETERS, J. **Improving the promotional forecasting accuracy for perishable items at Sligro Food Group B . V .** [s.l.] TUE. School of Industrial Engineering, 2012.

PROVOST, F.; FAWCETT, T. **Data Science for Business: What you need to know about data mining and data-analytic thinking**. Sebastopol: O'Reilly Media, Inc., 2013.

R-PROJECT. **What is R?** Disponível em: <<https://www.r-project.org/about.html>>. Acesso em: 1 nov. 2016.

RAMOS, P.; SANTOS, N.; REBELO, R. Performance of state space and ARIMA

models for consumer retail sales forecasting. **Robotics and Computer Integrated Manufacturing**, v. 34, p. 151–163, 2015.

REIS, F. D. **AVALIAÇÃO DE MODELOS DE PREVISÃO DE VENDAS A PARTIR DA EXPLORAÇÃO DE TÉCNICAS DE ANÁLISE DE SÉRIES TEMPORAIS, MÉTODOS CAUSAIS E DE REDES NEURAIS ARTIFICIAIS**. Belo Horizonte, MG: Universidade FUMEC, 2014.

REISEN, V. A.; SILVA, A. N. Introdução. In: REISEN, V. A.; SILVA, A. N. (Eds.). . **O uso da linguagem R para cálculos de estatística básica**. Vitória - ES: EDUFES, 2011. p. 2–11.

RIBEIRO, C. S. P. **Gestão da Atividade Promocional Baseada em Previsões Suportadas por Informação Competitiva Diversa**. [s.l.] Universidade do Porto, 2015.

SBVC. **Raio X: Os impactos da crise econômica no varejo Brasileiro**. Disponível em: <http://www.sbvc.com.br/wp-content/uploads/2016/05/Raio-X-da-Crise-no-Varejo_Completo-v2.pdf>. Acesso em: 1 jan. 2016.

SBVC. **O papel do varejo na economia brasileira**. Disponível em: <http://sbvc.com.br/wp-content/uploads/2017/03/O-Papel-do-Varejo-na-Economia-dados-2016-Versão-Completa_reduzido.pdf>. Acesso em: 23 mar. 2017.

SCHLÄFKE, M.; SILVI, R.; MÖLLER, K. A framework for business analytics in performance management. **International Journal of Productivity and Performance Management**, v. 62, n. 1, p. 110–122, 2013.

SCHWARTZ, J. D.; RIVERA, D. E. A process control approach to tactical inventory management in production-inventory systems. **International Journal of Production Economics**, v. 125, n. 1, p. 111–124, 2010.

SEREL, D. A. Optimal ordering and pricing in a quick response system. **International Journal of Production Economics**, v. 121, n. 2, p. 700–714, 2009.

SHOCKER, A. D.; BAYUS, B. L.; KIM, N. Product Complements and Substitutes in the Real World: The Relevance of “Other Products”. **Journal of Marketing**, v. 68, n. 1, p. 28–40, 2004.

SIMCHI-LEVI, D. **Cadeia de suprimentos: projetos e gestão**. Porto Alegre: Bookman, 2003.

SIMON, N. et al. Regularization paths for Cox’s proportional hazards model via coordinate descent. **Journal of Statistical Software**, v. 39, n. 5, p. 1–13, 2011.

SLACK, N.; CHAMBERS, S.; JOHNSTON, R. Natureza de planejamento e controle. In: SLACK, N.; CHAMBERS, S.; JOHNSTON, R. (Eds.). . **Administração da produção**. 2. ed. São Paulo, SP: Atlas, 2007. p. 313–342.

SONG, I.; CHINTAGUNTA, P. K. Measuring Cross-Category Price Effects with Aggregate Store Data. **Management Science**, v. 52, n. 10, p. 1594–1609, 2006.

STAKE, R. E. Qualitative case studies. In: STAKE, R. E. (Ed.). . **The Sage Handbook of Qualitative Research**. Thousand Oaks: Sage, 2011. p. 443–466.

SUSTROVA, T. An Artificial Neural Network Model for a Wholesale Company ' s Order-cycle Management. **International Journal of Engineering Business Management**, p. 1–6, 2016.

TANAKA, K. A sales forecasting model for new-released and nonlinear sales trend products. **Expert Systems with Applications**, v. 37, n. 11, p. 7387–7393, 2010.

TIACCI, L.; SAETTA, S. An approach to evaluate the impact of interaction between demand forecasting method and stock control policy on the inventory system performances. **International Journal of Production Economics**, v. 118, n. 1, p. 63–71, 2009.

TIBSHIRANI, R. Regression Shrinkage and Selection via the Lasso Robert Tibshirani. **Journal of the Royal Statistical Society**, v. 58, n. 1, p. 267–288, 1996.

VAN DONSELAAR, K. et al. Inventory control of perishables in supermarkets. **International Journal of Production Economics**, v. 104, n. 2, p. 462–472, 2006.

VAN ZELST, S. et al. Logistics drivers for shelf stacking in grocery retail stores: Potential for efficiency improvement. **International Journal of Production Economics**, v. 121, n. 2, p. 620–632, 2009.

VASCONCELLOS, M. A. S. Demanda, oferta e equilíbrio de mercado. In: VASCONCELLOS, M. A. S. (Ed.). . **Economia: micro e macro**. 5. ed. São Paulo, SP: Atlas, 2011. p. 29–64.

WERKEMA, C. **Ferramentas estatísticas básicas para o gerenciamento de processo**. 1. ed. São Paulo, SP: Werkema, 2006.

WERKEMA, M. C. C.; AGUIAR, S. **Análise de regressão: como entender o relacionamento entre as variáveis de um processo**. Belo Horizonte, MG: Fundação Christiano Ottoni, 1996.

WHICKER, L. et al. Understanding the relationships between time and cost to improve supply chain performance. **International Journal of Production Economics**, v. 121, n. 2, p. 641–650, 2009.

WHITTEN, D. G.; GREEN, K. W.; ZELBST, P. J. Triple-A supply chain performance. **International Journal of Operations & Production Management**, v. 32, n. 1, p. 28–48, 2012.

XU, X. et al. Forecasting tourism demand by extracting fuzzy Takagi–Sugeno rules from trained SVMs. **CAAI Transactions on Intelligence Technology**, v. 1, n. 1, p. 30–42, 2016.

YIN, R. K. **Case Study Reserch: Design and Methods**. 4. ed. [s.l.] Sage, 2009. v. 5

YIN, R. K. **Estudo de caso: planejamento e métodos**. 4. ed. Porto Alegre: Bookman, 2010.

ZHAO, Q.; ZHANG, Y. Multi-Product Utility Maximization for Economic Recommendation. **Proceedings of the Tenth ACM International Conference on Web Search and Data Mining**, p. 435–443, 2017.

ZLIOBAITE, I.; BAKKER, J.; PECHENIZKIY, M. Beating the baseline prediction in food sales: How intelligent an intelligent predictor is? **Expert Systems with Applications**, v. 39, n. 1, p. 806–815, 2012.

ZOTTERI, G.; KALCHSCHMIDT, M.; CANIATO, F. The impact of aggregation level on forecasting performance. **International Journal of Production Economics**, v. 93, p. 479–491, 2005.

APÊNDICE A

CATEG.	CÓDIGO	SUBGRUPO	GRAMATURA	DESCRIÇÃO SKU
FRIOS E LATICÍNIOS	1.1.68	BEBIDA LACTEA	200 ml	BEB LACTEA DANETTE CHOC TP 200ML
				BEB LACTEA ENERGIA NAT CHOC VIT TP 200ML
				BEB LACTEA NESCAU CHOC TP 200ML
				BEB LACTEA PIRACANJUBA ZERO CHOC 200ML
				BEB LACTEA SELITA CHOC 200ML
				BEB LACTEA SELITA LIGHT CHOC TP 200ML
				BEB LACTEA SELITA MORANGO TP 200ML
				BEB LACTEA SELITA VIT FRUTAS TP 200ML
				BEB LACTEA TODDYNHO CHOC TP 200ML
	1.1.88	BEBIDA LACTEA	1 LITRO	BEB LACTEA CAPIXABA TP 1L
				BEB LACTEA ENERGIA NAT CHOC VIT TP 1L
				BEB LACTEA SELITA CHOCOLATE TP 1L
				BEB LACTEA SELITA MORANGO TP 1L
				BEB LACTEA SELITA VIT FRUTA TP 1L
	1.2.88	LEITE LV DESNATADO	1 LITRO	LEITE L VIDA DAMARE DESN TP 1L
				LEITE L VIDA ITAMBE DESN ZERO LACT TP 1L
				LEITE L VIDA SELITA C/TAMPA LIGHT TP 1L
				LEITE L VIDA SELITA ZERO LACTOSE TP 1L
	1.3.88	LEITE LV INTEGRAL	1 LITRO	LEITE L VIDA DAMARE INTG TP 1L
				LEITE L VIDA DAMARE TP 1L
				LEITE L VIDA ITAMBE INTG ZERO LACT TP 1L
				LEITE L VIDA PIRACANJUBA 0% LACT TP 1L
				LEITE L VIDA PIRACANJUBA C/TAMPA INTG 1L
				LEITE L VIDA SELITA C/TAMPA INTG TP 1L
	1.4.88	LEITE LV SEMI-DESNATADO	1 LITRO	LEITE L VIDA PIRACANJUBA C/TAMPA SDESN 1L
				LEITE L VIDA PIRACANJUBA C/TAMPA SEMI 1L
				LEITE L VIDA SELITA C/TAMPA SEMI TP 1L
	1.6.34	MANTEIGA	200 Gramas	MANTEIGA DAMARE C/SAL PT 200G
				MANTEIGA DOURADA C/SAL PT 200G
				MANTEIGA EXTRA SELITA C/SAL PT 200G
				MANTEIGA EXTRA SELITA C/SAL TB 200G
				MANTEIGA VENEZA C/SAL PT 200G
	1.6.54	MANTEIGA	500 Gramas	MANTEIGA DAMARE C/SAL PT 500G
				MANTEIGA EXTRA SELITA C/SAL PT 500G
	1.7.39	MARGARINA CREMOSA	250 Gramas	MARG CREM QUALY C/SAL PT 250G
	1.7.54	MARGARINA CREMOSA	500 Gramas	MARG CREM DORIANA C/SAL PT 500G
				MARG CREM QUALY C/SAL PT 500G
				MARG CREM QUALY S/SAL PT 500G
				MARG QUALY AERADA C/SAL PT 500G
	1.7.61	MARGARINA CREMOSA	1 KG	MARG CREM QUALY C/SAL PT 1KG

	1.8.54	MARGARINA ESPECIAL	500 Gramas	MARG CLAYBOM C/SAL PT 500G
				MARG CREM DELICIA C/SAL PT 500G
				MARG QUALY LIGHT C/SAL PT 500G
	1.8.61	MARGARINA ESPECIAL	1 KG	MARG CREM DELICIA C/SAL PT 1KG

APÊNDICE B – SUMÁRIO EXECUTIVO

- **Leitura inicial dos pacotes necessários**

É necessário identificar os pacotes que serão utilizados para os cálculos. Como forma de facilitar, todos os pacotes necessários são solicitados no início.

```
library(readxl)
library(HDeconometrics)
library(forecast)
library(data.table)
library(DT)
library(dplyr)
```

- **Entrada de dados**

Neste passo é realizada a entrada das séries temporais, as séries devem estar organizadas de forma que as colunas sejam as produtos e as linhas os dias.

```
Data.all <- read_excel("C:/Users/j__p/OneDrive/Backup
arquivos/novo/master/dissertação/R/Base/em utilização/matriz comp-
sub/complementares - base.xlsx", sheet = "diario - todos", range = "A2:L160")
```

```
Data.all <- as.matrix(Data.all)
names <- as.data.frame(colnames(Data.all))
```

- **Corte principal**

Identificação de qual parte da série sera realizada o cálculo e qual será utilizada para comparação dos resultados da previsão

```
Data.cut <- Data.all[1:157,] #para fazer o lasso e previsão
Data.tail <- as.data.frame(Data.all[-c(1:157),]) #para comparar a previsão com os
dados reais
Data.tail <- as.data.frame(t(Data.tail))
```

- **Holt winters**

Previsão de forma automática de todas as séries pelo método de Holt-winters

```

Result.holt = matrix(NA,nrow=1,ncol=0)           #Criação de data frame em
branco para ser preenchido
Result.holt =data.frame(Result.holt)

j=0
repeat{
  y <- Data.cut[(j+1)]                          #Entrada da primeira
coluna(produto) para prever
  demand <- ts(y, start = c(1, 1), frequency = 6) #transforma em ts
  hw <- HoltWinters(demand)                       #ajusta o training
set
  forecast <- predict(hw, n.ahead = 1, prediction.interval = T, level = 0.95)
#previsão
  f <- as.data.frame(forecast)                   #Transformar em data
frame para pegar só a previsão
  f <- as.ts(f$fit)                              ##tira só a previsão e exclui o
upper e lower limits
  f <- as.data.frame(f)                          #transforma em data
frame

  colnames(f) <- names[j+1,1]                   #da o nome da variável
que fez a conta

  Result.holt = cbind(Result.holt,f)             #cria a matriz resposta

  j=j+1
  if (j >= ncol(Data.all)) break()
}

```

- **LASSO**

Previsão de forma automática de todas as séries pelo método LASSO

```

Result.lasso <- matrix(NA,nrow=1,ncol=0)
Result.lasso <- as.data.frame(Result.lasso)

```

```
#####loop do Lasso
```

```
i=0 #início do loop em 0
```

```

repeat{

  #refresh do y e x
  y = as.matrix(Data.cut[,i+1]) ##variável dependente primeira coluna
  x = (Data.cut) #cópia da base toda
  x <- x[,c(-(i+1))] #retira a variável y e fica com todas as outras variaveis
  x = as.matrix(x) # transforma em matriz

  #refresh do training e test set de x e y
  y.in=y[1:156]      #training set
  y.out=y[-c(1:156)] #test set
  x.in=x[1:156,]    #training set
  x.out=x[-c(1:156),] #test set

  ## ajuste do modelo e previsão LASSO
  lasso=ic.glmnet(x.in,y.in,crit = "bic") #ajuste do modelo com o training set
  previsao.lasso=predict(lasso,newdata=x.out) #previsão com o test set

  colnames(previsao.lasso) <- names[i+1,] #nomeando a linha que serviu como
dependente

  Result.lasso <- cbind(Result.lasso,previsao.lasso) #

  i <- i+1
  if (i >= ncol(Data.all)) break()

}

```

- **Acuracia LASSO**

Medição da acurácia da previsão realizada pelo método LASSO

```

acuracia.lasso <- matrix(NA,nrow=0,ncol=0)
acuracia.lasso <- as.data.frame(acuracia.lasso)
i = 0
repeat{

  #transformação para ts para conseguir medir a acurácia
  y.real <- as.ts(Data.tail[,i+1])
  f.lasso <- as.ts(Result.lasso[,i+1])

  results = as.data.frame(accuracy(f.lasso,y.real)) #medida da acuracia

  rownames(results) <- names[i+1,] #nomeando a linha que serviu como dependente

```

```

acuracia.lasso = rbind(acuracia.lasso,results)

i <- i+1
if (i >= ncol(Data.tail)) break()
}

```

- **Acuracia Holt winters**

Medição da acurácia da previsão realizada pelo método Holt-winters

```

acuracia.holt <- matrix(NA,nrow=0,ncol=0)
acuracia.holt <- as.data.frame(acuracia.holt)
i = 0
repeat{

  #transformação para ts para conseguir medir a acurácia
  y.real <- as.ts(Data.tail[,i+1])
  f.holt <- as.ts(Result.holt[,i+1])

  results = as.data.frame(accuracy(f.holt,y.real)) #medida da acuracia

  rownames(results) <- names[i+1,] #nomeando a linha que serviu como
dependente

  acuracia.holt = rbind(acuracia.holt,results)

  i <- i+1
  if (i >= ncol(Data.tail)) break()
}

```

- **Comparação total**

Comparação da acurácia da previsão realizada pelo método Holt-winters e pelo Metodo LASSO.

```

acuracia.final <- data.frame()
acuracia.final <- cbind(acuracia.lasso$RMSE,acuracia.holt$RMSE)
rownames(acuracia.final) <- names[,]
colnames(acuracia.final) <- c("RMSE - lasso","RMSE - holt")

redu <- as.data.frame(round((((acuracia.lasso$RMSE/acuracia.holt$RMSE)-1)*100,
digits = 2))
colnames(redu) <- "Redução % lasso/holt"

```

```

compara <- round(cbind(acuracia.final, redu), digits = 2)

datatable((compara),
  options = list(pageLength = 12))%>%
  formatStyle('Redução % lasso/holt',
    fontWeight = styleInterval(0, c('normal', 'bold')),
    backgroundColor = styleInterval(0, c('NULL', 'lightblue')))

write.table(compara, "PARAMETROS.xls", sep="\t")

```

- **Quantidade reduzida**

Calculo da quantidade de séries que a acurácia reduziu quando comparado o LASSO com o Holt-winters.

```

c <- table(compara$`Redução % lasso/holt`)
resumo <- as.data.frame(table(compara$`Redução % lasso/holt` >="0"))
resumo$Var1 <- NULL
resumo <- transpose(resumo)
colnames(resumo) <- c("negativo", "positivo")
rownames(resumo) <- "liq" ## alterar o h
resumo <- cbind(totvar <- sum(resumo$negativo, resumo$positivo), resumo)
colnames(resumo) <- c("Total de variaveis", "negativo", "positivo")
resumo <- cbind(resumo, reduc <- round((1-(resumo$positivo/resumo$Total))*100,
digits = 2))
colnames(resumo) <- c("Total de variaveis", "negativo", "positivo", "Redução %")
resumo$negativo <- NULL
resumo$positivo <- NULL

media.reduzida <- mean(compara$`Redução % lasso/holt`)
resumo <- cbind(resumo, media.reduzida)

```

- **Resumo**

Calculo da quantidade de séries que a acurácia reduziu quando comparado o LASSO com o Holt-winters.

- **complementares e substitutos**

Identificação das séries que são complementares e substitutas

```
#Entrada de dados normal
Data <- read_excel("C:/Users/j__p/OneDrive/Backup
arquivos/novo/master/dissertação/R/Base/em utilização/matriz comp-
sub/complementares - base.xlsx",
sheet = "diario - todos", range = "a2:l160")
```

```
Data <- as.data.frame(Data)
names <- as.data.frame(colnames(Data))
```

- **Loop dos coeficientes lasso**

Calculo para geração dos coeficientes da matriz de complementares e substitutos

```
i=0 #inicio do loop em 0
```

```
Result.lasso <- matrix(NA,nrow=(ncol(Data)),ncol=0) #cria matriz para ser
preenchida
```

```
Result.lasso <- as.data.frame(Result.lasso) # transforma em dataframe
```

```
repeat{
```

```
  #refresh do y e x
```

```
  y = as.matrix(Data[,i+1]) ##variável dependente primeira coluna
```

```
  x = (Data) #cópia da base toda
```

```
  x[,i + 1] <- NULL #retira a variável y e fica com todas as outras variaveis
```

```
  x = as.matrix(x) # transforma em matriz
```

```
  ## ajuste do modelo e previsão LASSO
```

```
  lasso=ic.glmnet(x,y,crit = "bic") #ajuste do modelo com o training set
```

```
  b <- as.data.frame(lasso$coefficients)
```

```
  #c <- as.data.frame(cbind(rownames(b),b[,1]))
```

```
  Result.lasso <- cbind(Result.lasso,b) #preenche a matriz cada vez que o loop
passa
```

```
  i <- i+1
```

```
  if (i >= ncol(Data)) break()
```

```
}
```

- **MATRIZ**

Organiza em forma de matriz o resultado do loop anterior

```

matriz.lasso <- matrix(NA,nrow=(ncol(Data)),ncol=0) #criação de matriz em branco
apra ser preenchida
matriz.lasso <- as.data.frame(matriz.lasso)

i =0

repeat{

  df <- as.data.frame(Result.lasso[,i+2]) #pega a partir da segunda coluna (a
primeira coluna ja vem certa)
  intercept <- as.data.frame(df[1,]) #pega o intercepto que sempre esta na
primeira linha
  colnames(intercept) <- "x" #da o nome para a coluna para conseguir
juntar tudo depois

  aux <- as.data.frame(df[2:ncol(Data),]) #cria matrix auxiliar para ser cortada
  top <- as.data.frame(aux[0:i+1,]) #retira a parte que vai ficar acima do
intercepto
  colnames(top) <- "x" #da o nome para a coluna para conseguir juntar
tudo depois

  down <- as.data.frame(aux[-c(0:i+1),]) #retira a parte que vai ficar abaixo do
intercepto
  colnames(down) <- "x" #da o nome para a coluna para conseguir juntar
tudo depois

  new <- rbind(top,intercept,down) #vetor organizado

  matriz.lasso <- cbind(matriz.lasso,new) #preenche a matriz final organizada

  i= i+1

  if (i >= ncol(Data)-1) break()
}

matriz.lasso <- cbind(as.data.frame(Result.lasso[,1]),matriz.lasso) #Coloca a
primeira coluna que foi retirada porque ja veio certa do primeiro loop
rownames(matriz.lasso) <- names[,] #Coloca o nome das
variaveis nas linhas
colnames(matriz.lasso) <- t(names) #Coloca o nome das
variaveis nas colunas
matriz.lasso <- round(matriz.lasso, 2) #Arredonda para 2 casas
decimais

#excel
# write.table(matriz.lasso, "PARAMETROS.xls", sep="\t") #se quiser salvar
como excel

```

- **Tabela formatada DT**

Para fins de demonstração no próprio R, a matriz anterior pode ser demonstrada de forma formatada a partir do script abaixo

```
matriz.plot <- matriz.lasso
matriz.plot[matriz.plot==0] <- "-" #coloca "-" para todos valores
iguais a zero
```

```
datatable((matriz.plot),
  class = 'cell-border stripe',
  options = list(pageLength = 120)) %>%
```

```
  formatStyle(names(matriz.lasso),
    Color = styleInterval(0, c('red', 'darkgreen')),
    backgroundColor = styleEqual(0, 'lightgrey'))
```

```
write.table(matriz.plot, "PARAMETROS.xls", sep="\t")
```