

WILLIAN SILVA GONÇALVES

Impacto do Clima nas Vendas do Varejo Físico de Moda na Cidade do Rio de Janeiro

PROJETO DE GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO
APRESENTADO AO DEPARTAMENTO DE ENGENHARIA INDUSTRIAL
DA PUC-RIO, COMO PARTE DOS REQUISITOS PARA OBTENÇÃO
DO TÍTULO DE ENGENHEIRO DE PRODUÇÃO

Orientadora: Soraida Aguilar

Departamento de Engenharia Industrial
Rio de Janeiro, 17 de novembro de 2023.

RESUMO

Este estudo visa investigar o impacto das variáveis climáticas nas vendas do varejo físico de uma marca de moda masculina localizada na cidade do Rio de Janeiro, com foco na quantidade de peças vendidas. A validação das variáveis foi conduzida por meio do modelo de regressão linear múltipla, empregando o método *stepwise* para a seleção das variáveis significativas. Os resultados revelaram que a temperatura exerce um impacto positivo nas vendas de peças, destacando assim a influência direta das condições climáticas sobre o desempenho do varejo físico dessa marca específica.

Palavras-chave: Varejo Físico; Variáveis Climáticas; Impacto Climático; Regressão Linear; Stepwise.

ABSTRACT

This study aims to investigate the impact of weather variables on brick-and-mortar retail sales of a men's fashion brand located in the city of Rio de Janeiro, focusing on the quantity of pieces sold. The variables validation was conducted using the multiple linear regression model, employing the stepwise method to select significant variables. The results revealed that temperature had a positive impact on parts sales, thus highlighting the direct influence of weather conditions on the physical retail performance of this specific brand.

Keywords: Brick-and-Mortar Retail; Weather Variables; Weather Impact; Linear Regression; Stepwise.

LISTA DE FIGURAS

Figura 1 – Fluxograma do método stepwise.....	14
Figura 2 – Boxplot da quantidade vendida ao longo do ano	18
Figura 3 – Boxplot da quantidade vendida ao longo da semana	20
Figura 4 – Temperatura média ao longo do ano (°C).....	21
Figura 5 – Radiação média ao longo do ano (KJ/m ²)	21
Figura 6 – Precipitação média ao longo do ano (mm).....	22
Figura 7 – Quantidade de Vendas × Valor.....	23
Figura 8 – Gráfico de Quantile-quantile plot (QQplot)	26
Figura 9 – Gráfico de Resíduos vs Valores Ajustados	27

LISTA DE TABELAS

Tabela 1 – Tipo das Variáveis de Venda.....	16
Tabela 2 – Tipo das Variáveis Climáticas.....	16
Tabela 3 – Descrição das Variáveis Categóricas.....	17
Tabela 4 – Análise descritiva das variáveis de venda.	19
Tabela 5 – Descrição das variáveis climáticas.....	23
Tabela 6 – Resultado do ajuste dos parâmetros da regressão	24
Tabela 7 – Resultado dos Testes de Estatísticos do Modelo	25
Tabela 8 – Resultado dos Testes de Normalidade.....	26
Tabela 9 – Resultado do Teste de Homoscedasticidade.....	27
Tabela 10 – Resultado do Teste de Multicolinearidade.....	28
Tabela 11 – Resultado da Análise de Valores Extremos	28

SUMÁRIO

1.	INTRODUÇÃO	7
2.	REFERENCIAL TEÓRICO	10
2.1	ESTRATÉGIAS DE ESTOQUE ADAPTATIVAS ÀS MUDANÇAS CLIMÁTICAS.....	10
2.2	ANÁLISE DE REGRESSÃO LINEAR MÚLTIPLA.....	10
2.3	PREMISSAS DA REGRESSÃO LINEAR MÚLTIPLA	11
2.4	SELEÇÃO AUTOMÁTICA DE VARIÁVEIS (STEPWISE)	13
3.	METODOLOGIA	15
3.1	DESCRIÇÃO DOS DADOS	15
3.2	DESCRIÇÃO DAS VARIÁVEIS SOB ANÁLISE.....	15
3.3	MODELAGEM	16
4.	ANÁLISE E DISCUSSÃO DOS RESULTADOS	18
4.1	ANÁLISE DESCRITIVA E GRÁFICA DOS DADOS	18
4.2	MODELAGEM.....	24
4.3	ANÁLISE DOS RESIDUAIS	25
4.3.1	<i>Normalidade</i>	25
4.3.2	<i>Homoscedasticidade</i>	26
4.3.3	<i>Multicolinearidade</i>	27
4.3.4	<i>Análise de valores extremos</i>	28
5.	CONCLUSÃO.....	30
6.	BIBLIOGRAFIA	31

1. INTRODUÇÃO

O varejo de moda no Brasil é um setor consolidado, mas que também está em contínua expansão. O setor têxtil tem grande importância na economia brasileira, por ser um forte gerador de empregos, com grande volume de produção e exportações crescentes (Fujita et al., 2013). De acordo com os dados da Associação Brasileira do Varejo Têxtil, o setor de vestuário brasileiro representou uma receita de mais de 229 bilhões de reais em 2021, sendo também responsável por mais de 775 mil postos de emprego.

A indústria da moda vem passando por um período de grandes transformações globalmente. A necessidade constante de reduzir custos e manter um nível de serviço elevado é crucial para as empresas aumentarem suas vantagens competitivas. Outro grande desafio é se adaptar ao gosto dinâmico e variado dos consumidores, elevando a complexidade da cadeia de operações.

As características particulares dos produtos dificultam a acurácia da previsão da demanda. As peças são marcadas por um ciclo de vida curto, geralmente desenhadas para atender a tendência da época, e possuem um período de vendas considerado pequeno, sendo medido em meses ou semanas. Outro ponto é a necessidade de o vendedor de sempre apresentar novidades aos consumidores, o que acarreta numa coleção com alto número de SKUs (Nenni et al., 2013a).

Um sistema de previsão de demanda certo pode gerar uma redução de custos com estoque, em específico o nível de estoque de segurança, que poderá ser otimizado. Isso faz com que a empresa entregue um nível de serviço melhor aos seus consumidores, aumentando sua competitividade em relação aos seus concorrentes (Nenni et al., 2013b). Imprecisões nas previsões levam a falta de produtos ou a excessos de inventários, resultando em perdas de vendas ou em capital de giro mais elevado e, em ambos os casos, em custos extras desnecessários (Verstraete et al., 2019).

Estudos recentes sugerem que as variáveis climáticas podem afetar as atividades de compras das pessoas (Busse et al., 2015). Os efeitos climáticos são amplamente pesquisados nos temas de energia, agricultura e biologia, porém, no varejo, em específico o setor têxtil, os estudos sobre os impactos ainda são limitados.

Inúmeros estudos têm destacado a sensibilidade do consumidor às condições climáticas, influenciando comportamentos de consumo específicos. De acordo com Smith et al. (2018), as variações climáticas afetam diretamente as preferências de vestuário, levando os consumidores a buscarem peças adequadas às condições do ambiente. Em períodos de temperaturas mais

elevadas, por exemplo, a demanda por roupas leves e frescas tende a aumentar, enquanto em climas mais frios, há uma inclinação para peças mais pesadas e acessórios de inverno. Com base neste tipo de evidência anedótica, pesquisas científicas descobriram que as variáveis meteorológicas podem afetar o comportamento de consumo (Murray et al., 2019).

A gestão eficaz do estoque e a criação de estratégias de promoção de vendas também são influenciadas pelas variáveis climáticas. Pesquisas indicam que varejistas de moda que adotam políticas flexíveis de estoque, ajustando suas coleções de acordo com as sazonalidades climáticas, experimentam melhores resultados financeiros (Jones, 2019). Além disso, campanhas de marketing sazonais podem ser elaboradas para atrair consumidores em sintonia com as mudanças climáticas, criando uma conexão entre a oferta do varejo e as necessidades do consumidor em diferentes momentos do ano.

Embora a influência das variáveis climáticas nas vendas do varejo de moda seja evidente, é importante destacar os desafios e oportunidades que surgem dessa interação complexa. Eventos climáticos extremos, como ondas de calor ou invernos rigorosos, podem levar a picos de demanda, mas também representam riscos operacionais. Nesse contexto, a implementação de tecnologias de previsão meteorológica e análise de dados pode ser uma ferramenta valiosa para o varejo, permitindo uma adaptação mais ágil às mudanças climáticas, ao compreender que as variáveis climáticas exercem uma influência significativa sobre as vendas no varejo, estas variáveis têm um papel determinante na tomada de decisões dos consumidores e nas estratégias comerciais adotadas pelas empresas. Dentre as variáveis climáticas, destacam-se aquelas que têm um impacto mais pronunciado nas vendas de varejo, moldando padrões de consumo e comportamento do consumidor tais como temperatura, precipitação, estação do ano (Smith et al., 2018).

Para ajustar os seus modelos de negócio para lidar com a incerteza causada pelo clima, as empresas de varejo online estão incorporando dados meteorológicos nos seus sistemas operacionais. Por exemplo, em 2014, a Alibaba fechou um acordo de cooperação estratégica com o serviço estatal de meteorologia chinês para explorar os dados climáticos fornecidos. Essa cooperação estratégica fornece aos varejistas serviços de tomada de decisão relacionados com o clima e previsões climáticas a curto, médio e longo prazo, ajudando assim os fabricantes e comerciantes de produtos sazonais a formular planos de negócios bem-sucedidos (Tian et al., 2019).

Intuitivamente muitos varejistas acreditam que os fenômenos climáticos impactam significativamente a quantidade de fluxo de clientes nas lojas físicas e, conseqüentemente, o

volume de vendas em suas lojas. Acreditam ainda que há efeito de substituição diante de um dia chuvoso entre lojas estabelecidas em rua com as lojas em shopping centers ou lojas virtuais (Dalmarco, 2015).

Para entender esse fenômeno, esse estudo visa realizar uma análise quantitativa através de uma regressão linear múltipla que usa a base de vendas de uma marca masculina de moda do segmento premium que tem suas lojas sediadas na capital carioca e alguns dados meteorológicos da cidade do Rio de Janeiro; com o intuito de validar a hipótese principal desta pesquisa é que as variáveis climáticas influenciam nas vendas do varejo físico do setor de moda.

O estudo é organizado em seis capítulos. O primeiro é a introdução, delineando a relevância e o escopo do trabalho. O segundo capítulo abrange o referencial teórico, destacando estudos anteriores relacionados e as definições necessárias para a compreensão do contexto. No terceiro capítulo, é apresentada a metodologia empregada, com uma descrição dos métodos e análise de dados. O quarto capítulo concentra-se na análise gráfica e descritiva dos resultados, fornecendo uma interpretação dos dados coletados. A quinta seção constitui a conclusão, consolidando os principais achados e destacando suas implicações e, por fim, o sexto capítulo abrange a bibliografia, detalhando as fontes consultadas ao longo do estudo, garantindo a integridade acadêmica do trabalho.

2. REFERENCIAL TEÓRICO

2.1 Estratégias de Estoque Adaptativas às Mudanças Climáticas

Segundo Kotler (2018, p. 582), varejo são todas as atividades relativas à comercialização de bens ou serviços diretamente ao consumidor final para uso pessoal e não comercial. Um varejista ou loja de varejo é qualquer empreendimento comercial cujo volume de vendas provenha principalmente do varejo.

A literatura mostra que há uma relação entre a variação climática e a atividade econômica (Lazo et al., 2011). O estudo realizado por Tian et al. (2019) em uma loja de conveniências, mostrou que nos dias mais chuvosos há um aumento do número de itens por pedido, pois as pessoas tendem a comprar mais produtos de uma só vez.

De acordo com a análise de Dalmarco (2015) sobre as vendas de clientes que usam o cartão de crédito como meio de pagamento, o consumo médio em dias chuvosos permanece o mesmo, mas há uma migração dos clientes do varejo físico de rua para os e-commerces ou lojas em shoppings centers.

As localizações das lojas são importantes no quesito do impacto climático. O artigo de Martínez-de-Albéniz e Belkaid (2021) mostra que no varejo físico de moda há uma redução de 7.4% no fluxo de clientes dentro das lojas de ruas em dias chuvosos, enquanto nas lojas em shopping centers houve um aumento de 5.2%. A temperatura também é uma variável climática importante, onde nos dias mais quentes há um aumento de venda de produtos de verão tais como vestidos e há uma queda na venda de produtos considerados de inverno, como casacos por exemplo.

2.2 Análise de Regressão Linear Múltipla

A análise de regressão linear múltipla é uma ferramenta estatística poderosa que fornece insights valiosos sobre as relações entre variáveis complexas. Sua aplicação adequada requer uma compreensão sólida dos fundamentos estatísticos subjacentes e uma cuidadosa consideração das pressuposições envolvidas. Ao superar esses desafios, os pesquisadores e profissionais podem utilizar essa técnica para realizar análises preditivas e explicativas que contribuem significativamente para o avanço do conhecimento em diversas áreas.

Formalmente, a regressão linear múltipla pode ser entendida como uma técnica multivariada cuja finalidade principal é obter uma relação matemática entre uma das variáveis estudadas (variável dependente ou resposta) e o restante das variáveis que descrevem o sistema

(variáveis independentes ou explicativas), e reduzir um grande número de variáveis para poucas dimensões com o mínimo de perda de informação, permitindo a detecção dos principais padrões de similaridade, associação e correlação entre as variáveis. Sua principal aplicação, após encontrar a relação matemática, é produzir valores para a variável dependente quando se têm as variáveis independentes (Sassi et al., 2022).

Em geral, a variável dependente ou resposta Y , pode ser relacionada a k variáveis independentes. O modelo de regressão linear múltipla se dá por:

$$Y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_kx_k + \varepsilon \quad (1)$$

Onde Y é a variável dependente ou de resposta, x_j ($j = 0, 1, \dots, k$) as variáveis independentes, os parâmetros β_j ($j = 0, 1, \dots, k$) são os coeficientes de regressão e ε é o erro. Os parâmetros β_j ($j = 0, 1, \dots, k$) são chamados de coeficientes de regressão parcial porque descrevem o efeito parcial de uma variável independente quando as outras variáveis interdependentes no modelo são mantidas constantes (Hines et al., 2003).

A estimação dos parâmetros β_j na regressão linear múltipla é geralmente realizada por meio do método dos mínimos quadrados. O objetivo é minimizar a soma dos quadrados dos resíduos, ou seja, a diferença entre os valores observados e os valores previstos. A solução para os coeficientes é obtida através de cálculos que envolvem matrizes e álgebra linear.

A avaliação do modelo de regressão linear múltipla é crucial para determinar sua eficácia e utilidade na explicação do fenômeno em estudo. Medidas como o coeficiente de determinação (R^2), testes de significância dos parâmetros e análise dos resíduos são comumente empregados para avaliar a qualidade do ajuste do modelo.

2.3 Premissas da Regressão Linear Múltipla

A aplicação efetiva da análise de regressão linear múltipla pressupõe algumas condições fundamentais. Estas incluem a linearidade, independência dos erros, homocedasticidade (variância constante dos erros), normalidade dos resíduos e ausência de multicolinearidade significativa entre as variáveis independentes. A violação dessas pressuposições pode comprometer a validade dos resultados obtidos.

Tanto na Regressão Linear Simples quanto na Regressão Múltipla, as suposições do modelo ajustado precisam ser validadas para que os resultados sejam confiáveis. Análise de

Resíduos são um conjunto de técnicas utilizadas para investigar a adequabilidade de um modelo de regressão com base nos resíduos.

O resíduo ε_j é dado pela diferença entre a variável resposta observada Y e a variável resposta estimada \hat{Y} . A ideia básica da análise dos resíduos é que, se o modelo for apropriado, os resíduos devem refletir as propriedades impostas pelo termo de erro do modelo. Tais suposições são:

1. ε_j e ε_i são independentes;
2. $Var(\varepsilon_i) = \sigma^2$ (constante);
3. $\varepsilon_i \sim N(0, \sigma^2)$ (normalidade);
4. Modelo é linear;
5. Não existir outliers (pontos atípicos) influentes.

Na Regressão Múltipla, além das suposições listadas acima, é necessário diagnosticar colinearidade e multicolinearidade entre as variáveis de entrada para que a relação existente entre elas não interfira nos resultados, causando inferências errôneas ou pouco confiáveis. As técnicas utilizadas para verificar as suposições descritas acima podem ser informais (como gráficos) ou formais (como testes estatísticos) (Maçaira et al., 2013).

Algumas técnicas gráficas para análise dos resíduos são:

- **Quantile-quantile plot (QQplot):** verifica a normalidade dos dados;
- **Gráfico dos resíduos versus a ordem de coleta dos dados:** avalia a hipótese de independência dos dados;
- **Gráfico dos resíduos versus valores ajustados:** verifica a homoscedasticidade do modelo, isto é, σ^2 constante;
- **Gráfico dos Resíduos Padronizados versus valores ajustados:** verifica se existem outliers em Y ;

Para a análise formal dos resíduos, pode-se realizar os seguintes testes:

- **Testes de Normalidade:** Shapiro-Wilks, Kolmogorov-Smirnov, Cramer-von Mises e Anderson-Darling.

As hipóteses estatísticas a serem testadas são:

H_0 : Os resíduos são normalmente distribuídos

H_a : Os resíduos não são normalmente distribuídos

- **Teste para independência dos resíduos:** Durbin-Watson.

As hipóteses estatísticas a serem testadas são:

H_0 : Os resíduos são independentes

H_a : Os resíduos não são independentes

- **Teste de homoscedasticidade dos resíduos:** White para modelos de regressão.

As hipóteses estatísticas a serem testadas são:

H_0 : As variâncias dos resíduos são iguais

H_a : As variâncias dos resíduos não são iguais

2.4 Seleção automática de variáveis (Stepwise)

Ao realizar o modelo de regressão, no conjunto de variáveis independentes pode haver variáveis que pouco influenciam o conjunto de variáveis dependentes (saída). O método Stepwise é usado para selecionar quais variáveis mais influenciam o conjunto de saída podendo, assim, diminuir o número de variáveis a compor a equação de regressão. A Figura 1 ilustra o processo, em notação BPMN, da seleção automática de variáveis utilizando o método Stepwise. O método é feito de forma iterativa, adicionando (passo Forward) e removendo variáveis (passo Backward), a partir de um critério de seleção, um dos critérios de seleção mais usados é o teste F, mas também pode ser feito com o coeficiente correlação linear múltipla ou erro quadrático total (Alves et al., 2013).

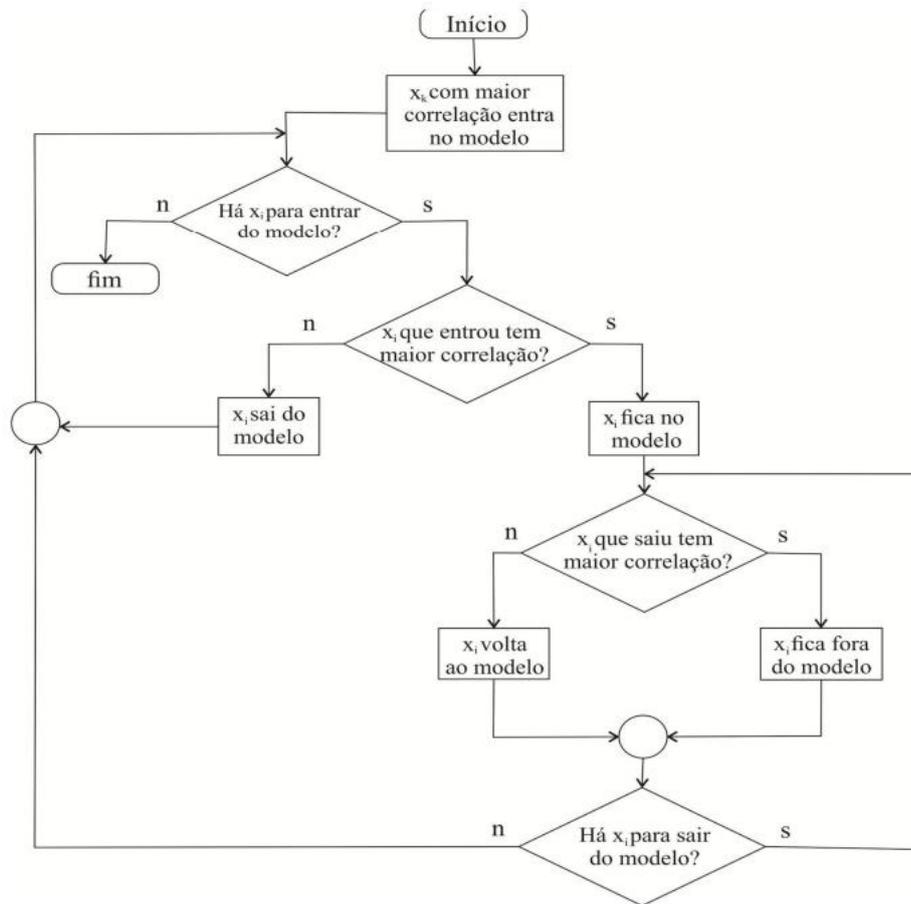


Figura 1 – Fluxograma do método stepwise. Fonte: Alves et al. (2013).

No final do processo iterativo será fornecido o ajuste do modelo para as variáveis mais significativas da modelagem.

3. METODOLOGIA

3.1 Descrição dos dados

A marca sob análise atua no setor de varejo de moda masculina, posicionando-se como uma empresa premium, voltada para um público de maior poder aquisitivo, pertencente às classes A e B, e oferecendo produtos com um ticket médio substancialmente mais elevado. A organização mantém uma presença geográfica abrangente no território brasileiro, no entanto, para efeitos deste estudo, foram selecionadas especificamente seis de suas lojas. Dessas, cinco estão localizadas em centros comerciais, enquanto a sexta é uma loja de rua. Todas essas lojas estão localizadas na cidade do Rio de Janeiro e apresentaram registros diários de vendas nos anos de 2017, 2018, 2019 e 2022. Além da base de dados de vendas, que foi adquirida diretamente junto à marca objeto de estudo, também foram incorporadas variáveis como a mediana de preço das peças vendidas e se o período de vendas é de liquidação ou não. Também foi incluída a variável econômica da Selic para analisar o cenário macroeconômico do período.

Inicialmente, essa base estava consolidada, considerando a agregação por filial e dia. Contudo, posteriormente, optou-se por organizá-la, realizando uma agregação exclusivamente por dia. Para esta pesquisa, as informações relativas à cidade do Rio de Janeiro nos anos de 2017, 2018, 2019 e 2022 foram selecionadas como fonte de análise. Os anos de 2020 e 2021 foram deliberadamente excluídos desta análise devido aos efeitos limitativos do lockdown e seu impacto no comércio varejista físico.

Por outro lado, os dados climáticos históricos são de domínio público e foram adquiridos por meio do portal do Instituto Nacional de Meteorologia (INMET, 2023). Foram selecionadas as variáveis: Precipitação Diária Acumulada, Temperatura Média Diária e Radiação Solar Diária Média. Essa base de dados compreende observações das variáveis climáticas em intervalos horários, e no âmbito deste estudo, contudo, procedeu-se à agregação dos dados em formato diário. Nesse processo, calculou-se a média de todas as variáveis, além da inclusão de duas novas dimensões referentes à precipitação e à radiação global máximas registradas no decorrer do dia.

3.2 Descrição das Variáveis sob Análise

As Tabelas abaixo apresentam todas as variáveis subjacentes à análise realizada nas bases de dados. Cada variável é acompanhada pela sua unidade de medida e classificação, considerando sua natureza categórica ou numérica.

Tabela 1 – Tipo das Variáveis de Venda.

Variável	Unidade	Tipo de Variável
Dia da Semana	-	Catégorica
Mês	-	Catégorica
Quantidade	Peças	Numérica
Valor	R\$	Numérica
Desconto Médio	%	Numérica
Mediana do Preço	R\$	Numérica
Período de Vendas	-	Catégorica
Selic	%	Numérica
Coleção	-	Catégorica
Fim de Semana	-	Catégorica

Fonte: Elaborado pelo autor.

Tabela 2 – Tipo das Variáveis Climáticas.

Variável	Unidade	Tipo de Variável
Radiação Global	KJ/m ²	Numérica
Temperatura Média	°C	Numérica
Precipitação	mm	Numérica

Fonte: Elaborado pelo autor.

Estas variáveis incluem a coleção à qual a venda foi realizada como "*Inverno*" ou "*Verão*", a quantidade total de peças adquiridas em um determinado dia, o valor total das transações diárias, a média de descontos aplicados e o período de vendas, que pode ser classificado como "*liquidação*" ou "*normal*". Para a definição do período de liquidação, considerou-se que um mês é caracterizado como tal se a média dos descontos aplicados pelas lojas durante esse período for superior a 15%. Além disso, foi incorporada uma variável adicional que indica se o dia corresponde a um fim de semana ou não, o que pode influenciar o comportamento dos consumidores e, por conseguinte, as métricas de vendas observadas.

3.3 Modelagem

Para este estudo, no processo de modelagem, a variável dependente, Y , está apresentada pelo volume de venda, isto é, a variável "*quantidade*"; além disso, foram consideradas as variáveis climáticas denotadas x_j ($j = 0, 1, \dots, k$) abrangendo todas as variáveis de vendas

listadas na Tabela 1, bem como as variáveis climáticas apresentadas no Tabela 2. Para a condução da regressão, as variáveis categóricas relacionadas às vendas foram transformadas em variáveis Dummy. A dimensão "Mês" foi mantida, levando em consideração a influência sazonal nas vendas do varejo. Os parâmetros estimáveis, β_j , representam os coeficientes associados a cada variável das Tabelas 1, 2 e 3, enquanto ε denota o erro do modelo.

$$Y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_{10}x_{10} + \varepsilon \quad (2)$$

Subsequentemente à modelagem inicial, empregou-se o método de seleção de variáveis stepwise com o intuito de otimizar o modelo, identificando as variáveis significativa quando seus p-valores são menores ao nível de significância $\alpha = 5\%$. Este procedimento visa refinar a eficiência do modelo, promovendo uma abordagem mais precisa e ajustada aos padrões observados nos dados.

Os níveis das variáveis categóricas, para a realização da regressão linear múltipla, são como apresentados na Tabela 3.

Tabela 3 – Descrição das Variáveis Categóricas.

VARIÁVEL	CATEGORIAS
Dia da semana	Segunda – Terca – Quarta – Quinta – Sexta – Sabado – Domingo
Período vendas	Normal – Liquidação
Mês	Jan – Fev – Mar – Abr – Mai – Jun – Jul – Ago – Set – Out – Nov – Dez
Coleção	Inv - Ver
Fim de semana	Sim – Não

Fonte: Elaborado pelo autor.

As dimensões “Segunda”, “Normal”, “Jan” e “Sim” foram utilizadas como variáveis de referência.

4. ANÁLISE E DISCUSSÃO DOS RESULTADOS

4.1 Análise descritiva e gráfica dos dados

Ao analisar o histórico do volume de vendas ao longo dos diferentes meses, como visto na Figura 2, destaca-se a importância de agosto, novembro e dezembro. O mês de agosto assume relevância significativa no contexto do varejo masculino devido à comemoração do Dia dos Pais, que impulsiona as transações comerciais. Por sua vez, o mês de novembro se destaca com a Black Friday, caracterizada por promoções e descontos expressivos, estimulando o consumo. Além disso, dezembro é notável devido ao volume expressivo de vendas associado às compras de Natal, representando o ápice do varejo de moda, uma vez que a data é central para a indústria.

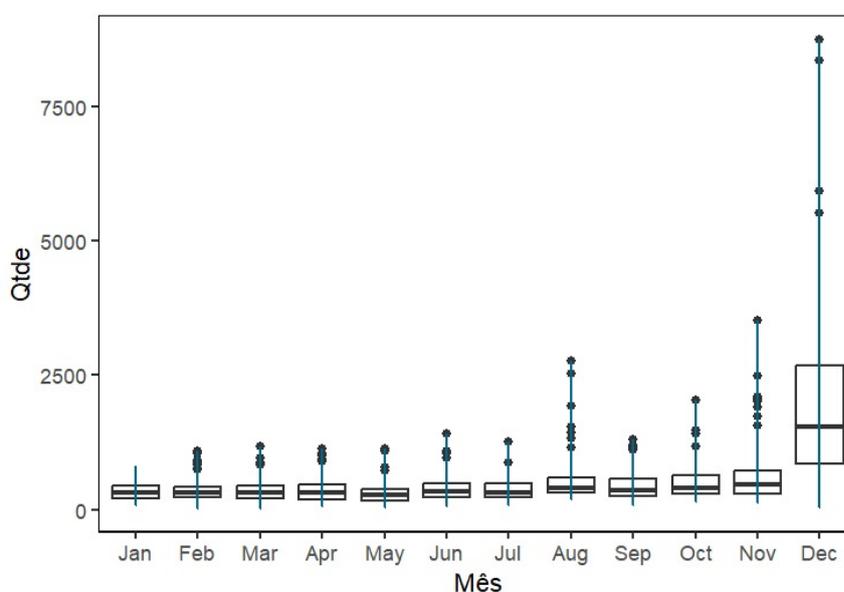


Figura 2 – Boxplot da quantidade vendida ao longo do ano. Fonte: Elaborado pelo autor.

Na Tabela 4, abaixo pode-se analisar as medidas de resumo de tendência central e dispersão de cada variável numérica:

Tabela 4 – Análise descritiva das variáveis de venda.

	VALOR DE COMPRA (R\$)	QTDE (un)	MEDIANA DE PREÇO (R\$)	SELIC (%)
Min	1.961,00	15	139,00	4,40
Max	3.402.899,00	8751	498,00	13,65
Amplitude	3.400.938,00	8700	359,00	9,25
Média	152.722,00	543	267,00	8,74
Desv Pad	239.477,00	680	70,47	2,96
25%	50.336,00	244	198,00	6,40
50%	83,386,00	362	249,00	6,90
75%	165,143,00	575	298,00	11,65

Fonte: Elaborado pelo autor.

A Tabela 4 revela uma ampla variação nos valores de compra, com mínimo de R\$ 1.961,00 e máximo de R\$ 3.402.899,00, resultando em uma amplitude de R\$ 3.400.938,00. A quantidade varia de 15 a 8.751, com média de 543 e desvio padrão de 680, indicando significativa dispersão. A taxa SELIC varia de 4,40% a 13,65%, com média de 8,74%. Quartis destacam pontos-chave, como o primeiro quartil de R\$ 50.336,00 e o terceiro quartil de 11,65% para a taxa SELIC

Verificando graficamente o comportamento das vendas ao longo da semana, Figura 3, torna-se evidente que sexta-feira e sábado se destacam como os dias de maior volume de vendas. Em contrapartida, o domingo se revela como o dia da semana com o menor volume de vendas, indicando uma diminuição notável na atividade comercial nesse dia.

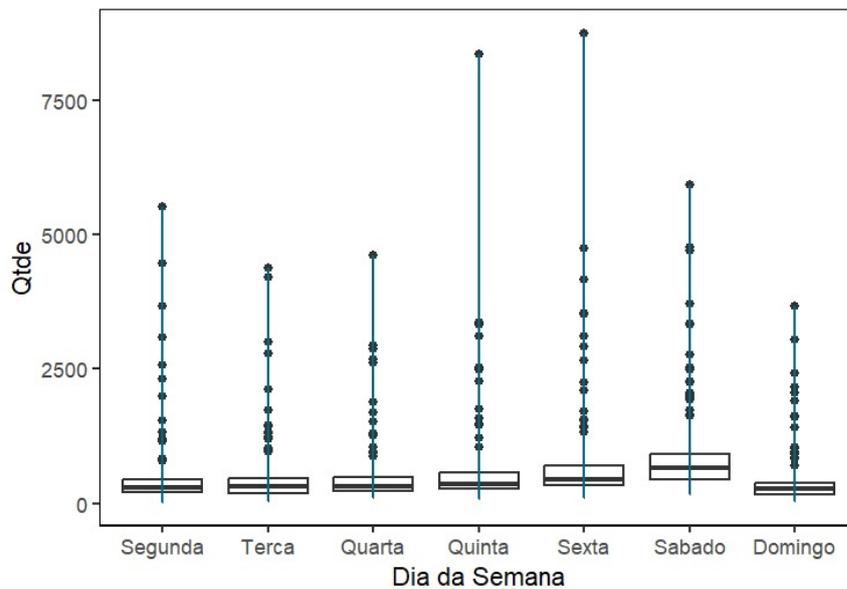


Figura 3 – Boxplot da quantidade vendida ao longo da semana. Fonte: Elaborado pelo autor.

Por outro lado, ao observar as variáveis climáticas, e especificamente a temperatura, na Figura 4, é apresentada o comportamento da temperatura média mensal de todos os anos. Observa-se uma clara tendência de queda nas temperaturas durante o outono (21 de março a 21 de junho), atingindo os valores mais baixos nos meses de julho e agosto, que correspondem ao período do inverno (21 de junho a 23 de setembro). Posteriormente, evidencia-se uma acentuada ascensão nas temperaturas durante a primavera (23 de setembro a 21 de dezembro), atingindo seu pico em janeiro, marcando o início do verão (21 de dezembro a 21 de março).

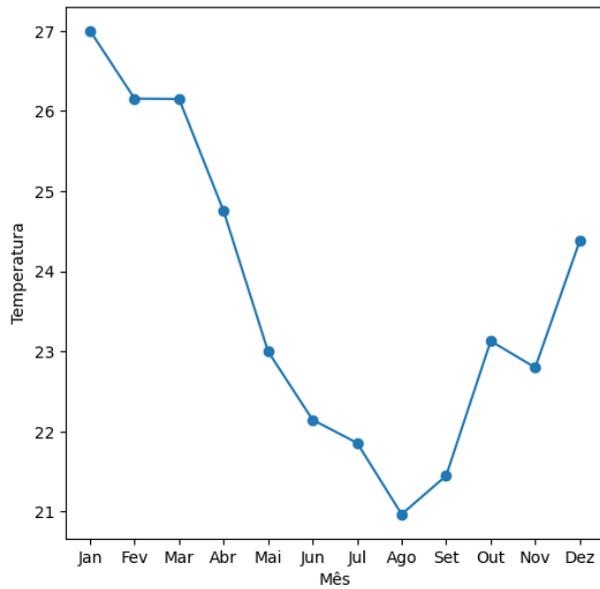


Figura 4 – Temperatura média ao longo do ano (°C). Fonte: Elaborado pelo autor.

A radiação, conforme observado na Figura 5, exibe um padrão de comportamento semelhante ao da temperatura, revelando picos durante os meses de verão, notadamente em dezembro e janeiro, enquanto atinge seus valores mínimos no período de inverno, destacando-se especialmente no mês de junho.

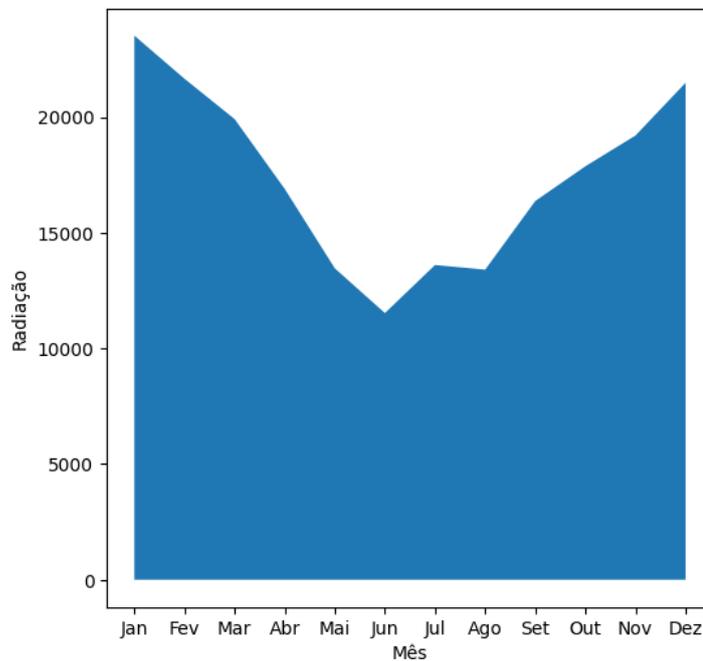


Figura 5 – Radiação média ao longo do ano (KJ/m²). Fonte: Elaborado pelo autor.

Examinando a Figura 6, que contém a distribuição da precipitação média ao longo dos meses, torna-se evidente a ocorrência do pico em abril, enquanto a mínima é registrada em julho. A estação do inverno se destaca como o período de menor incidência de chuvas, refletindo em sua caracterização como uma estação mais seca. Por outro lado, o verão se apresenta como a estação de maior precipitação, revelando um padrão sazonal marcado por variações significativas ao longo do ano.

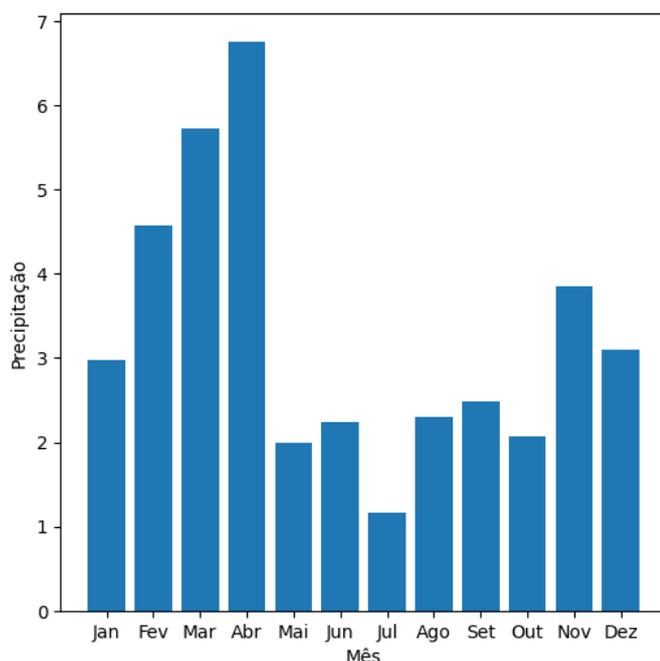


Figura 6 – Precipitação média ao longo do ano (mm). Fonte: Elaborado pelo autor.

Para entender melhor a relação das variáveis climáticas, foram calculadas as medidas de resumo, como pode ser visto na Tabela 5. Dentre as variáveis avaliadas, destaca-se a precipitação, cujos resultados revelam uma incidência relativamente baixa de dias chuvosos. Tanto a média quanto a mediana indicam valores próximos a 0mm, sugerindo que a ocorrência de precipitação significativa é uma ocorrência rara no contexto investigado.

Tabela 5 – Descrição das variáveis climáticas.

	TEMPERATURA	PRECIPITAÇÃO	RADICAÇÃO
Min	16,88	0,00	1.233,00
Max	30,78	241,40	32.536,00
Amplitude	13,90	241,40	31.303,00
Média	23,62	3,25	17.349,00
Desv Pad	2,61	11,51	7.396,95
25%	21,70	0,00	12.335,00
50%	23,41	0,00	16.965,00
75%	25,50	1,20	22.903,00

Fonte: Elaborado pelo autor.

A análise da Figura 7, do gráfico de dispersão, revela uma associação notável direta entre a quantidade de produtos vendidos e o valor total da compra.

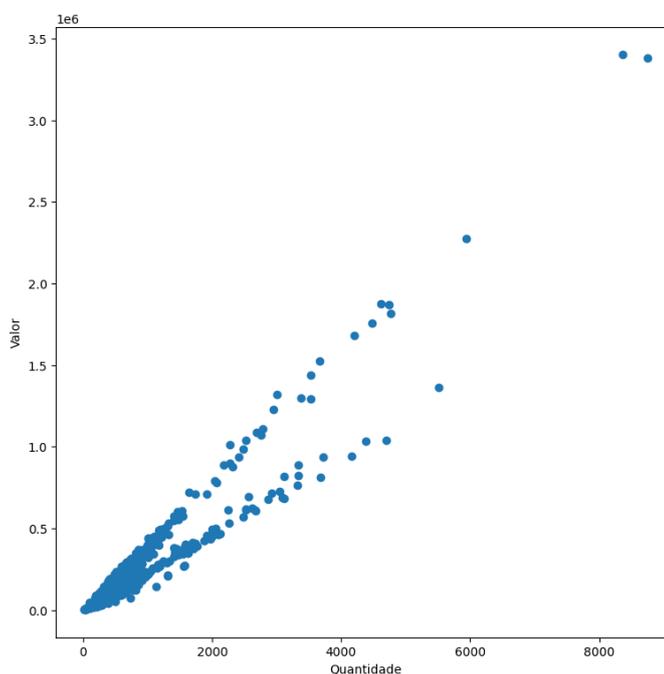


Figura 7 – Quantidade de Vendas × Valor. Fonte: Elaborado pelo autor.

A identificação de uma correlação positiva sugere que, à medida que a quantidade vendida aumenta, o valor total da compra também aumenta. Essa relação robusta entre as duas

variáveis evidencia a interdependência substancial entre a quantidade adquirida pelos clientes e o montante financeiro correspondente.

4.2 Modelagem

No processo de modelagem, ao aplicar a regressão linear múltipla em conjunto com o processo de seleção de variáveis Stepwise, o melhor resultado para a regressão foram as variáveis expostas na Tabela 6, que obtiveram um p-valor menor que um $\alpha = 0,05$, indicando que todas as variáveis são estatisticamente significativas.

Tabela 6 – Resultado do ajuste dos parâmetros da regressão

Variável	Coefficiente	Erro padrão	Estatística T	Pr(> t)
Mediana de Preço	-0,5375	0,0699	-7,690	0,0000
Valor	0,0026	0,0000	114,2880	0,0000
Dia da Semana – Sexta	62,5200	14,6100	4,2790	0,0000
Dia da Semana – Sabado	112,7000	14,7400	7,6450	0,0000
Mês – Jul	54,0300	22,2400	2,4300	0,0152
Mês – Ago	107,4000	23,3400	4,6030	0,0000
Mês – Set	71,9200	23,0200	3,1240	0,0018
Mês – Out	41,9600	21,2900	1,9710	0,0489
Mês – Nov	53,9200	21,6500	2,4900	0,0129
Mês – Dez	297,9000	23,1300	12,8810	0,0000
Temperatura	8,6270	2,2080	3,9070	0,0001
Selic	-22,8100	1,5460	-14,7500	0,0000

Fonte: Elaborado pelo autor.

Adicionalmente, na Tabela 7 são apresentados outros resultados importantes da regressão como o Teste F, que permite analisar se a regressão é estatisticamente significativa como um todo, e para o qual se obteve um p-valor menor que 5%, o que leva à rejeição da hipótese nula que estabelece que todos os coeficientes da regressão são iguais a zero, indicando que a regressão é significativa, isto é, que existe pelo menos um parâmetro diferente de zero. Além disso coeficiente de determinação (R^2) ajustado explicar 95,53% da variabilidade da variável de resposta.

Tabela 7 – Resultado dos Testes de Estatísticos do Modelo

Teste	Estatística
Multiple R-squared	0,9540
Adjusted R-squared	0,9533
Estatística F	1402 (p-valor: 0,0000)

Fonte: Elaborado pelo autor.

A análise dos resultados apresentados na Tabela 6 demonstra o impacto das variáveis no modelo em estudo. Destaca-se que o aumento no preço das peças está associado a uma diminuição nas quantidades adquiridas, conforme revelado pela dimensão da mediana de preço. As variáveis de dia da semana e mês validam a existência de padrões sazonais, identificando sexta-feira e sábado como os principais dias em termos de volume de compras, com ênfase nos meses do segundo semestre, especialmente em agosto, influenciado pelo Dia dos Pais, e dezembro, marcado pelo Natal. A variável macroeconômica, Selic, evidencia que o aumento da taxa Selic prejudica as vendas. Destaca-se também, que, entre as variáveis climáticas analisadas, apenas a temperatura demonstra significância estatística, apresentando uma correlação positiva, indicando que um aumento na temperatura está associado a um aumento no volume de peças vendidas.

4.3 Análise dos Residuais

4.3.1 Normalidade

Na análise da normalidade dos resíduos, empregou-se o método gráfico do QQPlot como instrumento de avaliação (Figura 8). A inspeção visual revela que os dados não se aderem a uma distribuição normal, uma vez que os pontos extremos do gráfico demonstram uma significativa disparidade em relação à reta teórica. Esse desvio é evidente devido a presença de assimetrias ou caudas mais pronunciadas, indicando uma possível distribuição que difere da esperada distribuição normal.

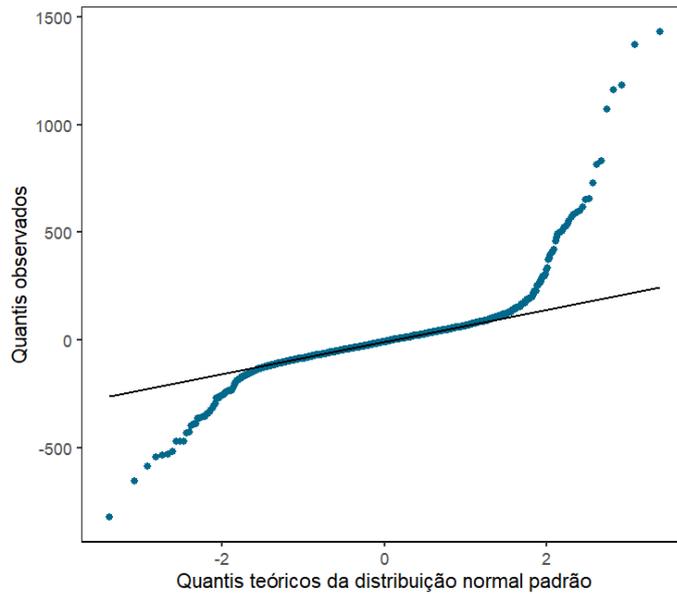


Figura 8 – Gráfico de Quantile-quantile plot (QQplot) Fonte: Elaborado pelo autor.

Uma análise mais formal foi empregada ao realizar os testes estatísticos conforme apresentado no Tabela 8. Os resultados obtidos constataam que a distribuição dos resíduos não se ajusta a uma distribuição normal, como evidenciado na Figura 8. Esta conclusão é fundamentada no p-valor obtido, que se posiciona abaixo do nível de significância de 5%, conduzindo à rejeição da hipótese nula.

Tabela 8 – Resultado dos Testes de Normalidade

Teste	Estatística	p-valor
Shapiro-Wilk	0,7241	0,000
Kolmogorov-Smirnov	0,1664	0,000
Cramer-von Mises	121,4262	0,000
Anderson-Darling	85,3603	0,000

Fonte: Elaborado pelo autor.

4.3.2 Homoscedasticidade

Para a análise de homoscedasticidade será realizada inicialmente uma inspeção visual dos resíduos vs os valores ajustados; e posteriormente, será empregado o teste estatístico de White para uma validação formal dos achados gráficos. Ao analisar a Figura 9, pode-se perceber que a distribuição dos dados está concentrada em torno do zero, no eixo x . A identificação deste

cluster de pontos sugere a presença de uma estrutura sistemática na dispersão dos resíduos. Este fenômeno indica que a variabilidade dos erros não é constante em toda a faixa de valores ajustados, violando a premissa de homocedasticidade.

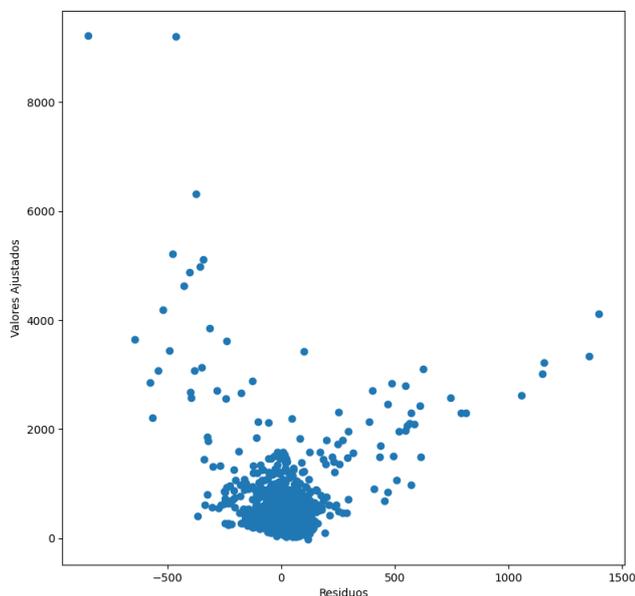


Figura 9 – Gráfico de Resíduos vs Valores Ajustados. Fonte: Elaborado pelo autor.

Com o intuito de validar de forma estatística que a variabilidade dos erros (ou resíduos) seja constante, utilizou-se o teste estatístico de White, conforme evidenciado no Tabela 9. A partir dos resultados, pode-se rejeitar a hipótese nula de igualdade de variâncias dos erros, isto é, H_0 : A variância dos resíduos é constante. Nesse contexto, a constatação de heteroscedasticidade nos erros implica que a variabilidade destes não é constante em relação às variáveis explicativas do modelo.

Tabela 9 – Resultado do Teste de Homoscedasticidade

Teste	Estatística	p-valor
Teste White	521	0,0000

Fonte: Elaborado pelo autor.

4.3.3 Multicolinearidade

A análise de multicolinearidade é essencial para compreender as inter-relações entre as variáveis independentes do modelo. Utilizando o Fator de Inflação da Variância (VIF), cujos

resultados foram sintetizados no Tabela 10, revela-se não há multicolinearidade uma vez que para isto o VIF deve ser maior do que 10.

Tabela 10 – Resultado do Teste de Multicolinearidade

Variável	GVIF	Graus de Liberdade
Mediana de Preço	1,7693	1
Valor	2,0708	1
Dia da Semana	1,0732	6
Mês	9,1245	11
Temperatura	2,8321	1
Selic	1,6062	1
Umidade	1,3793	1
Período de Vendas	2,2462	1
Desconto	2,0683	1

Fonte: Elaborado pelo autor.

4.3.4 Análise de valores extremos

A análise de valores extremos foi utilizada para avaliar se há observações que se destacam significativamente do restante do conjunto de dados, podendo ter um impacto significativo nas análises estatísticas e nos resultados do modelo. O teste realizado foi o Teste de Bonferroni, e seus resultados podem ser vistos na Tabela 11. Como temos p-valores abaixo de 0,05, podemos rejeitar a hipótese nula (H_0 : A observação é um outlier).

Tabela 11 – Resultado da Análise de Valores Extremos

Nº Observação	Rstudent	Unadjusted p-value	Bonferroni p
1072	10,2452	0,0000	0,0000
711	9,8365	0,0000	0,0000
1073	8,3162	0,0000	0,0000
710	8,2518	0,0000	0,0000
712	7,5600	0,0000	0,0000
1433	-6,6752	0,0000	0,0000
709	5,7457	0,0000	0,0000
713	5,5980	0,0000	0,0000
353	5,2670	0,0002	0,0002
1437	-4,5688	0,0002	0,0077

Fonte: Produzido pelo autor.

Os resíduos não satisfazem as premissas de normalidade, incorrelação e homocedasticidade, logo, há indícios que a relação seja não linear. Isto pode afetar a validade das inferências feitas a partir do modelo ou nos dados para abordar as violações das premissas, ou seja, é necessário considerar métodos estatísticos alternativos no intuito de reavaliar a formulação do modelo. No entanto, apesar de não serem satisfeitos os pressupostos da regressão linear, é possível verificar a existência da relação das vendas do varejo com as variáveis climáticas e a influência nos dias da semana e nos meses.

5. CONCLUSÃO

Ao considerar tanto os elementos visuais quanto os indicadores estatísticos, constatou-se que as variáveis climáticas exercem uma influência substancial na performance de vendas ao longo do ano, apesar de algumas premissas da regressão linear não terem sido plenamente satisfeitos. No âmbito específico do varejo de moda masculina, com lojas localizadas na cidade do Rio de Janeiro, das três variáveis climáticas investigadas – radiação, precipitação e temperatura –, somente a última revelou uma correlação estatisticamente significativa. Especificamente, observou-se que o aumento da temperatura esteve associado a um incremento no volume de peças vendidas no dia.

Além disso, a análise revela a marcante importância das datas comemorativas no cenário do varejo físico, destacando especialmente o Dia dos Pais, a Black Friday e o Natal como momentos de expressivo impacto nas vendas. Tais informações devem ser levadas em conta na realização de modelos preditivos do volume de vendas, e podem ajudar em um melhor fluxo de estoque.

No contexto da pesquisa sobre as interações entre variáveis climatológicas e o desempenho de vendas no setor varejista, é pertinente destacar a importância de fortalecer a compreensão já estabelecida por outras análises de modelos, além do modelo linear apresentado nesse trabalho, como modelos de regressão polinomial ou Modelos de Regressão Simples. Além disso, ressalta-se a necessidade de explorar recomendações adicionais, tais como a performance de vendas por tipo de produto, como por exemplo uma análise pormenorizada de itens distintos tais como bermudas e calças, que possam servir como orientações para pesquisas futuras.

6. BIBLIOGRAFIA

ALVES, Marleide F.; LOTUFO, Anna Diva P.; LOPES, Mara Lúcia M. Seleção de variáveis stepwise aplicadas em redes neurais artificiais para previsão de demanda de cargas elétricas. **Proceeding Series of the Brazilian Society of Computational and Applied Mathematics**, v. 1, n. 1, 2013.

DALMARCO, Leandro Silva. **Impacto dos fatores climáticos nas vendas do varejo no Brasil. Uma abordagem com dados de painel com efeito fixo**. 2013. Tese de Doutorado.

INSTITUTO NACIONAL DE METEOROLOGIA (INMET). Dados históricos. Disponível em: <<https://portal.inmet.gov.br/dadoshistoricos>>. Acesso em: 15 ago. 2023.

JONES, R. (2019). Weathering the Storm: The Impact of Weather on Retail Sales. **Journal of Retailing**, 95(1), 22-41.

KOTLER, Philip. Administração de marketing: análise, planejamento, implementação e controle. In: **Administração de marketing: análise, planejamento, implementação e controle**. 1994. p. 582

LAZO, Jeffrey K. et al. US economic sensitivity to weather variability. **Bulletin of the American Meteorological Society**, v. 92, n. 6, p. 709-720, 2011.

MAÇAIRA, P.; Bastos, L.; AGUILAR, S. & PERES, I. (2022) **Inferência Estatística com R**, 1ª edição, Rio de Janeiro, Brasil. https://paulamacaira.github.io/livro_inferencia_com_R/. Acessado em 2022-03-24.

MARTÍNEZ-DE-ALBÉNIZ, Victor; BELKAID, Abdel. Here comes the sun: Fashion goods retailing under weather fluctuations. **European Journal of Operational Research**, v. 294, n. 3, p. 820-830, 2021.

MURRAY, Kyle B. et al. The effect of weather on consumer spending. **Journal of Retailing and Consumer Services**, v. 17, n. 6, p. 512-520, 2010.

NENNI, Maria Elena; GIUSTINIANO, Luca; PIROLO, Luca. Demand forecasting in the fashion industry: a review. **International Journal of Engineering Business Management**, v. 5, p. 37, 2013b.

NENNI, Maria Elena; SCHIRALDI, Massimiliano M. Validating virtual safety stock effectiveness through simulation. **International Journal of Engineering Business Management**, v. 5, p. 41, 2013a.

SMITH, J., et al. (2018). Climate and Clothing: Understanding the Relationship between Weather and Consumer Behavior in the Fashion Industry. **Weather, Climate, and Society**, 10(2), 387-399.

TIAN, Xin; CAO, Shasha; SONG, Yan. The impact of weather on consumer behavior and retail performance: Evidence from a convenience store chain in China. **Journal of Retailing and Consumer Services**, v. 62, p. 102583, 2021.

VERSTRAETE, Gylan; AGHEZZAF, El-Houssaine; DESMET, Bram. A data-driven framework for predicting weather impact on high-volume low-margin retail products. **Journal of Retailing and Consumer Services**, v. 48, p. 169-177, 2019.