

**UNIVERSIDADE FEDERAL DE MINAS GERAIS
DEPARTAMENTO DE ESTATÍSTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA**

Bruno Fernandes Estrella

**ANÁLISE REGIONAL DO VOLUME DE VENDAS NO VAREJO BRASILEIRO VIA
REGRESSÃO LINEAR**

Rio de Janeiro
2025

Bruno Fernandes Estrella

**ANÁLISE REGIONAL DO VOLUME DE VENDAS NO VAREJO BRASILEIRO VIA
REGRESSÃO LINEAR**

Monografia apresentada ao Departamento de Estatística da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Especialista em Estatística.

Orientador: Prof. Dr. Guilherme Lopes de Oliveira

Estrella, Bruno Fernandes.

E82a Análise regional do volume de vendas no varejo brasileiro via regressão linear [recurso eletrônico] / Bruno Fernandes Estrella – 2025.

1 recurso online (42 f. il., color.) : pdf.

Orientador: Guilherme Lopes de Oliveira.

Monografia (especialização) - Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Estatística.

Referências: f. 38-39.

1. Estatística. 2. Análise de regressão. 3. Comércio varejista – Vendas – Brasil. 4. Comércio varejista - Controle preditivo. 5. Indicadores econômicos - Métodos estatísticos – Brasil. 6. Políticas públicas – Brasil. I. Oliveira, Guilherme Lopes de. II. Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Estatística. III. Título.

CDU 519.2(043)



Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Estatística
Programa de Pós-Graduação

Caixa Postal 702
31270-901 Belo Horizonte- MG – Brasil

Telefone (31) 3409-5923

Fax (31) 3499-5924

E-mail: pgest@ufmg.br

WEB: <http://www.est.ufmg.br/posgrad/>

ATA DO 349ª. TRABALHO DE FIM DE CURSO DE ESPECIALIZAÇÃO EM ESTATÍSTICA DE BRUNO FERNANDES ESTRELLA.

Aos quinze dias do mês de abril de 2025, às 18:00 horas, com utilização de recursos de videoconferência a distância, reuniram-se os professores abaixo relacionados, formando a Comissão Examinadora homologada pela Comissão do Curso de Especialização em Estatística Computacional Aplicada, para julgar a apresentação do trabalho de fim de curso do aluno **Bruno Fernandes Estrella**, intitulado: “*Análise regional do volume de vendas no varejo brasileiro via regressão linear*”, como requisito para obtenção do Grau de Especialista em Estatística. Abrindo a sessão, o Presidente da Comissão, Professor Guilherme Lopes de Oliveira – Orientador, após dar conhecimento aos presentes do teor das normas regulamentares, passou a palavra ao candidato para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores com a respectiva defesa do candidato. Após a defesa, os membros da banca examinadora reuniram-se sem a presença do candidato e do público, para julgamento e expedição do resultado final. Foi atribuída a seguinte indicação: o candidato foi considerado Aprovado condicional às modificações sugeridas pela banca examinadora no prazo de 30 dias a partir da data de hoje por unanimidade. O resultado final foi comunicado publicamente ao candidato pelo Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a reunião e lavrou a presente Ata, que será assinada por todos os membros participantes da banca examinadora. Belo Horizonte, 15 de abril de 2025.

Documento assinado digitalmente



GUILHERME LOPES DE OLIVEIRA
Data: 16/04/2025 09:07:44-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Guilherme Lopes de Oliveira (orientador)
DECOM/CEFET-MG

Documento assinado digitalmente



SERGIO FELIPE ABREU DE BRITTO BASTOS
Data: 16/04/2025 13:20:14-0300
Verifique em <https://validar.iti.gov.br>

Prof. M.Sc. Sérgio Felipe Abreu de Britto Bastos
IFMG

RESUMO

A influência dos indicadores macroeconômicos no desempenho do setor varejista se refere à capacidade desses indicadores em explicar as variações nas vendas e atividades comerciais ao longo do tempo, sendo fundamental para a definição de estratégias comerciais e a formulação de políticas públicas regionais. Utilizando dados mensais coletados entre 2012 e 2022, este estudo desenvolveu um modelo preditivo usando regressão linear para relacionar variáveis econômicas ao Índice de Volume de Vendas no Varejo (IVV) nas cinco regiões do Brasil. Foram consideradas variáveis explicativas como o Índice de Atividade Econômica Regional (IBCR), taxa de desemprego, taxa de juros, além de indicadores sazonais e efeitos extraordinários decorrentes da pandemia de COVID-19 e defasagens do IVV. O modelo final apresentou elevado poder preditivo, com um R^2 ajustado de aproximadamente 0,79, e demonstrou consistência prática nos sinais dos coeficientes estimados. Os resultados indicaram que, controlado pelas demais variáveis, o desempenho do varejo é diferente entre as regiões do país e tende a melhorar com o aumento do IBCR e a redução dos índices de desemprego e juros, enquanto eventos sazonais e atípicos, exercem impacto significativo. A análise dos resíduos confirmou a adequação do modelo, com distribuição próxima à normalidade, homocedasticidade e ausência de autocorrelação. A avaliação preditiva revelou que o modelo possui maior eficiência preditiva nas regiões Nordeste, Sudeste e Centro-Oeste, enquanto as regiões Norte e Sul apresentam maiores desafios, sugerindo a necessidade de futuros refinamentos metodológicos. De forma geral, as análises realizadas fornecem subsídios importantes para a otimização de estratégias no setor varejista e para o desenvolvimento de políticas públicas que promovam o equilíbrio econômico regional.

Palavras-chave: Indicadores macroeconômicos; desempenho do varejo; diferenças regionais; predição

ABSTRACT

The influence of macroeconomic indicators on retail sector performance refers to their capacity to explain variations in sales and commercial activities over time, which is essential for defining business strategies and formulating regional public policies. Using monthly data collected between 2012 and 2022, this study developed a predictive linear regression model linking economic variables to the Retail Sales Volume Index (IVV) across the five regions of Brazil. The Regional Economic Activity Index (IBCR), unemployment rate, interest rate, seasonal indicators, extraordinary effects from the COVID-19 pandemic, and IVV lags were included in the model as covariates. The final model demonstrated high predictive power, with an adjusted R^2 of approximately 0.79, and showed practical consistency in the signs of estimated coefficients. The results indicated that, controlling for other variables, retail performance varies among regions and tends to improve with increases in IBCR and reductions in unemployment and interest rates, while seasonal and atypical events, significantly impact performance. Residual analysis confirmed the model's adequacy, revealing near-normal distribution, homoscedasticity, and absence of autocorrelation. Predictive evaluation demonstrated greater model efficiency in the Northeast, Southeast, and Central-West regions, while the North and South regions presented greater predictive challenges, suggesting the need for future methodological refinements. Overall, the analyses provide critical insights for optimizing retail sector strategies and developing public policies aimed at promoting regional.

Keywords: Macroeconomic indicators; retail performance; regional differences; forecasting

LISTA DE SIGLAS

BCB – Banco Central do Brasil

COVID-19 – *Corona Virus Disease 2019*

CreditoPF – Saldo das Operações de Crédito – Pessoas Físicas

EnergRes – Consumo de Energia Elétrica Residencial

FGV – Fundação Getúlio Vargas

IBCR – Índice de Atividade Econômica Regional

IBGE – Instituto Brasileiro de Geografia e Estatística

ICC – Índice de Confiança do Consumidor

INADIMP – Inadimplência das Operações de Crédito – Pessoas Físicas

IPCA – Índice Nacional de Preços ao Consumidor Amplo

IVV – Índice de Volume de Vendas no Varejo

MQO – Mínimos Quadrados Ordinários

PNAD – Pesquisa Nacional por Amostra de Domicílios

RM – Rendimento Médio Mensal Real

SBVC – Sociedade Brasileira de Varejo e Consumo

SUMÁRIO

1. INTRODUÇÃO	7
2. MATERIAIS E MÉTODOS	9
2.1 BASE DE DADOS E MATERIAIS.....	9
2.2 REGRESSÃO LINEAR MÚLTIPLA.....	10
2.2.1 Métricas de avaliação e validação de modelos de Regressão Linear Múltipla.....	12
3. RESULTADOS E DISCUSSÃO	15
3.1 ANÁLISE DESCRITIVA DOS DADOS.....	15
3.2 AJUSTE DA REGRESSÃO LINEAR MÚLTIPLA	28
3.3 ANÁLISE PREDITIVA.....	31
4. CONCLUSÕES E CONSIDERAÇÕES FINAIS.....	35
REFERÊNCIAS.....	37
ANEXO I - Função de Autocorrelação (ACF) versus Função de Autocorrelação Parcial (PACF) da série temporal do IVV em cada região do Brasil.....	39

1. INTRODUÇÃO

O setor varejista desempenha um papel crucial na economia brasileira, sendo responsável por uma parcela significativa do PIB, gerando milhões de empregos diretos e indiretos. Além disso, o varejo é um dos setores que mais refletem as mudanças no comportamento do consumidor e nas condições econômicas do país.

O setor varejista desempenha um papel crucial na economia brasileira, tanto em termos de contribuição para o Produto Interno Bruto (PIB) quanto na geração de empregos. De acordo com o estudo "O Papel do Varejo na Economia Brasileira", da Sociedade Brasileira de Varejo e Consumo (SBVC), o varejo restrito movimentou R\$ 1,99 trilhão em 2021, equivalente a 22,9% do PIB brasileiro, enquanto o varejo ampliado alcançou R\$ 2,41 trilhões, representando 27,7% do PIB nacional (SBVC, 2022).

O desempenho do varejo é frequentemente utilizado como um indicador para prever tendências econômicas mais amplas. Uma queda nas vendas pode sinalizar retração econômica, enquanto um aumento pode indicar expansão.

Uma ampla gama de produtos e serviços, desde alimentos e vestuário até eletrônicos e produtos de luxo são abrangidos pelo varejo. Isso permite uma análise abrangente dos padrões de consumo e das preferências dos consumidores. Dessa forma, o setor varejista está interligado a diversas cadeias produtivas, influenciando e sendo influenciado por setores como agricultura, indústria e serviços. Portanto, mudanças no varejo podem ter efeitos multiplicadores na economia.

De fato, o setor varejista tem sido amplamente discutido na literatura econômica devido à sua relevância macroeconômica e capacidade de indicar tendências econômicas. Conforme descrito por Kotler e Keller (2012), o varejo constitui um importante elo entre produtores e consumidores, refletindo diretamente as mudanças nas preferências do segundo grupo. Levy e Weitz (2018) reforçam que o varejo atua não apenas como ponto final das cadeias produtivas, mas também como impulsionador significativo de atividades econômicas, devido ao efeito multiplicador gerado pela sua interação com diversos setores, incluindo agricultura, indústria e serviços.

No contexto brasileiro, estudos como os de Rocha e Silva (2018) destacam a heterogeneidade regional do varejo, indicando que fatores culturais, econômicos e sociais desempenham papel determinante na variação do comportamento do consumidor. Por exemplo, produtos que são populares em uma região podem não ser em outra, devido a fatores como clima, tradições locais e poder aquisitivo. Além disso, a estrutura regional do varejo brasileiro implica desafios específicos para o desenvolvimento de estratégias comerciais eficazes e a interpretação adequada de indicadores econômicos (Figueiredo et al., 2002).

Adicionalmente, a análise dos serviços logísticos, conforme demonstrado por Figueiredo et al. (2002), revela que as diferenças regionais no Brasil se estendem às grandes macrorregiões, refletindo contextos marcadamente distintos. Na região Sudeste, por exemplo, onde se concentram centros econômicos robustos como São Paulo e Belo Horizonte, os varejistas demonstram expectativas elevadas e exigências mais rigorosas em termos de eficiência e agilidade logística. Em contraste, na região Nordeste, representada por mercados como o de Recife, as limitações estruturais e desafios de infraestrutura podem levar a percepções de desempenho inferiores, demandando estratégias adaptadas que considerem tais restrições. No Sul, embora os indicadores apontem para uma maior uniformidade e tradição em processos logísticos, as particularidades culturais e mercadológicas também impõem ajustes nas práticas de atendimento (Figueiredo et al., 2002).

A literatura identifica diversos indicadores econômicos que afetam diretamente o desempenho do varejo. Indicadores como o Índice de Atividade Econômica Regional (IBCR), Índice Nacional de Preços ao Consumidor Amplo (IPCA), taxa de juros, nível de emprego formal, inadimplência e confiança do consumidor têm sido frequentemente utilizados para prever oscilações no consumo e no volume de vendas varejistas (Mendonça e Sachsida, 2019; Carvalho et al., 2020). A literatura sugere ainda que esses indicadores têm impactos diferenciados nas diversas regiões brasileiras, dadas as desigualdades socioeconômicas e estruturais presentes no país (Sarti e Laplane, 2021).

Para investigar a relação entre indicadores econômicos e desempenho do setor varejista, a regressão linear tem sido amplamente empregada como ferramenta estatística eficaz. Segundo Montgomery, Peck e Vining (2015), a regressão linear possibilita identificar a magnitude e a significância estatística dos efeitos dos indicadores econômicos sobre uma variável dependente, por exemplo, o volume de vendas no varejo. Tal abordagem permite não só avaliar relações lineares diretas como também explorar o impacto simultâneo de múltiplos fatores sobre o comportamento das vendas.

Diante do cenário exposto, o objetivo central deste trabalho é analisar como os indicadores econômicos influenciam o volume de vendas no setor varejista nas diferentes regiões brasileiras, identificando os principais indicadores econômicos que afetam o varejo em cada região. Para tal, a regressão linear será utilizada como ferramenta estatística para modelar e compreender a relação entre indicadores econômicos e volume de vendas no varejo no Brasil entre 2012 e 2022. O estudo se justifica no sentido de que seus resultados poderão auxiliar empresas varejistas a entender melhor os mercados regionais, permitindo o desenvolvimento de estratégias de marketing e vendas adaptadas às especificidades locais e cenário macroeconômico.

2. MATERIAIS E MÉTODOS

2.1 Base de dados e variáveis

A base de dados considerada neste estudo contém informações mensais sobre indicadores econômicos selecionados, abrangendo o período de janeiro 2012 a dezembro 2022 para as cinco grandes regiões brasileiras: Norte, Nordeste, Centro-Oeste, Sudeste e Sul. Dentre tais indicadores, o Índice de Volume de Vendas no Varejo (IVV) representa a variável resposta de interesse do estudo, por medir diretamente o desempenho do setor varejista no país. Indicadores como a taxa média de juros das operações de crédito para pessoas físicas, taxa de desemprego e índice de confiança do consumidor serão testadas como variáveis preditoras. Uma breve descrição de todos os indicadores coletados e sua forma de medição estão apresentadas no Quadro 1. Dentre as fontes de dados, destacam-se os dados disponibilizados pelo Banco Central do Brasil (BCB) e pelo Instituto Brasileiro de Geografia e Estatística (IBGE).

Quadro 1 – Definição do banco de dados do estudo

Variável	Descrição	Características e Detalhes	Fonte
IVV	Índice de Volume de Vendas no Varejo – Ampliado (sem ajuste sazonal)	Calculado a partir dos índices estaduais fornecidos pela Pesquisa Mensal de Comércio (IBGE), ponderados pela receita bruta de revenda (Pesquisa Anual de Comércio, IBGE). Índice em escala regional e periodicidade mensal.	BCB-IBGE
IBCR	Índice de Atividade Econômica Regional	Indicador mensal que mede o nível da atividade econômica em cada região do país. Índice em periodicidade mensal.	BCB
IPCA	Índice Nacional de Preços ao Consumidor Amplo – Variação mensal	Representa a variação mensal dos preços de bens e serviços consumidos por famílias com renda mensal entre 1 e 40 salários mínimos. Índice regional em periodicidade mensal.	IBGE
Desemprego	Taxa de Desemprego	Taxa de desocupação, na semana de referência, das pessoas de 14 anos ou mais de idade. Os dados apresentados em escala trimestral foram mensalisados por intermédio de divisão simples por 3.	PNAD
INADIMP	Inadimplência das Operações de Crédito – Pessoas físicas	Mede o percentual das operações de crédito em atraso superior a 90 dias, considerando tanto o segmento de crédito livre quanto o direcionado. Índice regional em periodicidade mensal.	BCB
CreditoPF	Saldo das Operações de Crédito – Pessoas físicas	Valor total das operações ativas de crédito (empréstimos, financiamentos e arrendamento mercantil) concedidas às pessoas físicas pelas instituições financeiras que compõem o Sistema Financeiro Nacional. Valor em bilhões de reais, regionalizado e em periodicidade mensal.	BCB
EnergRes	Consumo de Energia Elétrica Residencial	Volume total de energia elétrica em milhares de Gigawatts-hora consumida mensalmente pelas residências por região.	Eletrobras
Juros	Taxa Média de Juros das Operações de Crédito – Pessoas físicas	Representa a média ponderada das taxas de juros das operações de crédito contratadas por pessoas físicas, considerando créditos livres e direcionados. Valores medidos em % ao ano, em escala nacional e mensalisados.	BCB

Variável	Descrição	Características e Detalhes	Fonte
ICC	Índice de Confiança do Consumidor	Avalia mensalmente o grau de confiança dos consumidores das cidades de Belo Horizonte, Brasília, Porto Alegre, Recife, Salvador, Rio de Janeiro e São Paulo, com pesos proporcionais à representatividade dessas cidades no consumo nacional.	FGV
RM	Rendimento Médio Mensal Real	Representa o rendimento médio real regional em milhares de reais dos trabalhadores com 14 anos ou mais, referente ao rendimento habitual obtido no trabalho principal na semana pesquisada, conforme dados da PNAD Contínua trimestral. Os dados apresentados em escala trimestral foram mensalisados por intermédio de divisão simples por 3.	IBGE – PNAD Contínua trimestral

Fonte: elaborado pelo autor

A maioria dos indicadores estava disponível em escala mensal, exceto Desemprego e Rendimento Médio Mensal Real (RM). Nos dois casos, a escala original dos dados é trimestral e, para fins desse estudo, os valores mensais foram obtidos através de uma distribuição uniforme do respectivo valor trimestral, ou seja, pela simples divisão entre os três meses do trimestre associado. Por se tratar de um período de análise de 12 anos, janeiro de 2012 a dezembro de 2022, tem-se um conjunto de dados de 132 observações em cada uma das cinco regiões do Brasil, totalizando 660 observações no banco de dados.

Com relação à escala geográfica dos dados, apenas Juros e ICC não estavam disponíveis ao nível regional. Por se tratarem de indicadores macroeconômicos importantes e que podem afetar a variável de desfecho, tais indicadores foram utilizados nas análises e para mês o valor nacional foi considerado constante para todas as regiões.

Por fim, vale destacar que variáveis que representam a dependência e a estrutura temporal dos dados foram criadas após realização de análises descritivas e modelagem. A saber, os modelos preditivos aplicados à série de dados mensais levaram em conta a indicação do trimestre, além de considerar uma defasagem da variável resposta IVV que permitiu considerar o valor do índice no mês anterior como preditor do índice de um dado mês.

Os dados referentes ao período de janeiro de 2012 a dezembro de 2021 foram utilizados para o treinamento dos modelos, constituindo o conjunto de dados de treino. Já os dados de janeiro de 2022 a dezembro de 2022 foram reservados para avaliar a capacidade preditiva dos modelos, formando o conjunto de dados de teste.

2.2 Regressão Linear Múltipla

A regressão linear múltipla é uma técnica estatística robusta que modela a relação entre uma variável dependente e duas ou mais variáveis independentes ao mesmo tempo. Ao contrário da regressão linear simples — que considera apenas um preditor — esse método incorpora diversos fatores, permitindo uma análise mais completa. Amplamente aplicada em campos como economia, ciências sociais, saúde, engenharia e *marketing*, a regressão linear múltipla não apenas quantifica a contribuição individual de cada variável, mas também fornece *insights* valiosos para compreender fenômenos complexos e embasar decisões estratégicas.

A forma geral do modelo de regressão linear múltipla é

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon,$$

onde

- y é a *variável dependente*;
- β_0 é o *intercepto*;
- $\beta_1, \beta_2, \dots, \beta_k$ são os *coeficientes associados às variáveis independentes* x_1, x_2, \dots, x_k ;
- ε representa o *termo de erro aleatório*.

Segundo a abordagem de mínimos quadrados, o objetivo ao construir um modelo de regressão linear é estimar os coeficientes $\beta_0, \beta_1, \dots, \beta_k$ de modo a minimizar a soma dos quadrados dos erros ($\sum \varepsilon_i^2$), produzindo o modelo linear que melhor descreve a relação entre a variável dependente e as variáveis independentes nos dados observados. Detalhes sobre os estimadores de mínimos quadrados para os coeficientes do modelo e outros métodos de estimação podem ser vistos, por exemplo, em Montgomery, Peck e Vining (2021) e Cordeiro, Demétrio e Moral (2024).

A regressão linear múltipla oferece vantagens significativas, destacando-se como uma ferramenta poderosa e versátil na modelagem estatística. Entre os principais benefícios desse método, destacam-se:

- **Controle de variáveis de confusão (*confounding factors*):** Na regressão linear múltipla é possível incluir simultaneamente diversas variáveis no mesmo modelo. Isso permite identificar o efeito específico de cada variável independente sobre a variável dependente, já considerando e ajustando o impacto das demais variáveis presentes no modelo. Dessa forma, é possível minimizar o risco de conclusões incorretas decorrentes da omissão de variáveis importantes, garantindo maior precisão e validade dos resultados.

- Análise eficaz dos dados: ao analisar conjuntos de dados complexos que envolvem múltiplas variáveis, a regressão linear múltipla oferece uma estrutura analítica robusta que facilita a identificação de padrões e promovendo uma compreensão abrangente e aprofundada dos fenômenos estudados.
- Avaliação da importância relativa dos preditores: permite determinar quais variáveis têm maior impacto ou influência na variável dependente, ajudando a identificar fatores-chave.
- Detecção de interações: é possível investigar como diferentes variáveis independentes interagem entre si, identificando efeitos combinados que não poderiam ser percebidos na análise simples.
- Maior poder explicativo: ao incorporar múltiplas variáveis independentes, a regressão múltipla explica melhor a variabilidade da variável dependente, proporcionando um entendimento mais completo do fenômeno estudado.

Por fim, é essencial verificar cuidadosamente os pressupostos da regressão linear múltipla, garantindo a validade, precisão e confiabilidade das análises realizadas. A violação desses pressupostos pode comprometer a qualidade das estimativas obtidas, conduzindo a conclusões incorretas ou enviesadas. Entre os principais pressupostos e análises diagnósticas recomendadas estão: avaliação da multicolinearidade, análise da normalidade dos resíduos, verificação da homoscedasticidade e independência dos resíduos.

2.2.1 Métricas de avaliação e validação de modelos de Regressão Linear Múltipla

Os testes de hipóteses são ferramentas essenciais para avaliar a significância estatística dos coeficientes em um modelo. O teste t examina individualmente cada coeficiente, verificando se este difere significativamente de zero, enquanto o teste F avalia a relevância global do modelo, analisando conjuntamente todos os coeficientes. Resultados significativos indicam que as variáveis independentes têm um impacto estatisticamente significativo sobre a variável dependente, reforçando sua importância para o modelo (Montgomery, Peck e Vining, 2021). A significância neste estudo foi considerada quando o p-valor do teste t associado ao coeficiente foi inferior ao nível de significância de 10%.

O R^2 ajustado é uma métrica que permite avaliar o grau de adequação do modelo aos dados observados, levando em consideração o número de variáveis independentes utilizadas. Um valor próximo a 1 indica que grande parte da variabilidade da variável dependente é explicada pelas variáveis independentes presentes no modelo, já corrigindo pelo número de preditores e evitando

distorções causadas por variáveis pouco relevantes. Essa é uma das métricas mais utilizadas na comparação entre modelos e na avaliação da capacidade preditiva (Morettin e Singer, 2022).

O Erro Padrão da Estimativa (SEE, do inglês *Standard Error of Estimate*), também chamado de desvio-padrão residual, mede a dispersão dos resíduos em relação à linha de regressão. Valores baixos de SEE indicam que o modelo tem um ajuste mais preciso aos dados observados. Interpretar adequadamente o SEE auxilia na avaliação da qualidade e da confiabilidade das estimativas feitas pelo modelo, permitindo identificar possíveis discrepâncias entre os valores observados e aqueles previstos pela regressão.

A presença de multicolinearidade pode prejudicar significativamente a interpretação dos coeficientes do modelo, tornando-os instáveis e imprecisos. Para diagnosticar esse problema, utiliza-se o Fator de Inflação da Variância (VIF). Valores elevados de VIF indicam forte correlação entre as variáveis independentes, exigindo uma análise mais detalhada, podendo levar à exclusão ou combinação das variáveis altamente correlacionadas, garantindo assim maior robustez e clareza na interpretação dos resultados.

A realização de testes para verificar a normalidade dos resíduos é uma etapa fundamental na validação do modelo. Caso os resíduos não apresentem uma distribuição aproximadamente normal, isso pode indicar a existência de padrões ou relações não capturadas adequadamente pelo modelo, sugerindo a necessidade de ajustes, inclusão de variáveis adicionais ou outras medidas corretivas. A não normalidade dos resíduos compromete ainda a validade dos testes de hipóteses e dos intervalos de confiança que podem ser construídos para os coeficientes do modelo. Neste trabalho, adotaremos o teste de Shapiro-Wilk, cuja hipótese nula postula que os dados seguem uma distribuição normal, enquanto a hipótese alternativa indica a não normalidade, como ferramenta para avaliar estatisticamente a normalidade dos resíduos. A transformação da variável resposta é um mecanismo que pode ser utilizado na tentativa de se aproximar da condição de normalidade dos erros quando isso não acontece de maneira direta (Cordeiro, Demétrio e Moral, 2024).

Utilizaremos o teste de Breusch-Pagan modificado¹ para verificar se a variância dos erros dos ajustes depende das variáveis explicativas, a hipótese nula assume homoscedasticidade, variância constante, e a alternativa indica heterocedasticidade. Quando há a indicação de heterocedasticidade, a aplicação de transformações na variável resposta ou a utilização de métodos mais robustos de estimação podem ser utilizados a fim de contornar o problema. Por exemplo, a estimação dos parâmetros via métodos de mínimos quadrados ponderados (Cordeiro, Demétrio e Moral, 2024).

¹ A versão modificada do teste de Breusch-Pagan é mais robusta por não exigir a suposição de normalidade dos resíduos, o que a torna mais confiável em contextos práticos onde essa condição nem sempre é atendida (Fox & Weisberg).

A presença de autocorrelação nos resíduos viola a suposição de independência das observações, essencial para garantir a validade das inferências realizadas pela regressão linear. A autocorrelação residual pode causar estimativas imprecisas dos erros padrões, comprometendo a confiabilidade dos testes de hipóteses e dos intervalos de confiança obtidos pelo modelo. Para identificar a presença desse problema, será aplicado o teste de Breusch-Godfrey (BG) (cuja hipótese nula assume ausência de autocorrelação serial nos resíduos, enquanto a hipótese alternativa indica a presença de autocorrelação de ordem superior). Resultados significativos quanto à autocorrelação indicam a necessidade de ajustes adicionais, como a inclusão de termos com defasagem temporal ou o uso de modelos alternativos que levem em consideração a estrutura temporal dos dados.

Para realizar comparações entre diferentes ajustes, será utilizado o Critério de Informação de Akaike (AIC). O AIC fornece uma medida relativa da qualidade dos modelos considerando tanto o ajuste aos dados quanto a complexidade do modelo. Valores menores de AIC indicam modelos mais adequados, permitindo selecionar aquele que oferece o melhor equilíbrio entre a precisão do ajuste e a simplicidade estrutural. Em outras palavras, o AIC auxilia na escolha do modelo mais parcimonioso e eficaz para explicar os dados observados.

No processo de seleção de variáveis, será utilizado o procedimento de eliminação regressiva, também denominado método *backward*. Essa abordagem consiste em iniciar a análise com a especificação completa do modelo e, em seguida, proceder à exclusão iterativa das variáveis menos significativas do ponto de vista estatístico. A cada iteração, reavalia-se o modelo com base em critérios como o AIC, até que não se verifique ganho adicional na qualidade do ajuste com a retirada de preditores.

Por fim, para avaliar a capacidade preditiva do ajuste, será realizada uma validação por meio da divisão do conjunto de dados em duas partes: um conjunto de treino, utilizado para a estimação dos parâmetros, e um conjunto de teste, reservado exclusivamente para avaliação das previsões. A qualidade das previsões será aferida utilizando-se métricas como o Erro Médio Absoluto (MAE) e a Raiz do Erro Quadrático Médio (RMSE), possibilitando identificar de maneira objetiva o grau de precisão das estimativas do modelo em situações práticas.

Todas as análises conduzidas nesse estudo foram feitas no *software* R (R CORE TEAM, 2024).

3. RESULTADOS E DISCUSSÃO

Conforme detalhado na Seção 3.1, a base de dados é composta por 660 observações referentes a indicadores econômicos selecionados, abrangendo o período de janeiro de 2012 a dezembro de 2022 para as cinco grandes regiões brasileiras: Norte, Nordeste, Centro-Oeste, Sudeste e Sul, incluindo as variáveis descritas na Tabela 1. Nesta seção, será feita uma análise descritiva e da correlação dos dados, seguida da aplicação dos modelos de regressão linear múltipla.

3.1 Análise descritiva dos dados

A análise descritiva² inicial apresentada na Tabela 1 revela particularidades importantes sobre o comportamento econômico regional ao longo do período analisado. A região Centro-Oeste apresentou um comportamento econômico relativamente estável, refletido pela baixa variabilidade do IVV, com coeficiente de variação (CV) de 11,45%, média de 88,23 e ausência de *outliers*. No entanto, a inflação regional (IPCA) mostrou-se bastante volátil (CV de 97,21%) com curtose alta (4,56), sugerindo ocorrência esporádica de eventos inflacionários extremos. A inadimplência média ficou em torno de 3,02%, com uma variabilidade moderada (CV de 20,79%) e presença de poucos eventos atípicos (1,52% de *outliers*).

A região Norte destacou-se por uma forte dependência do crédito pessoal, refletida por uma alta variabilidade nessa variável (CV de 40,57%) e significativa presença de eventos extremos (6,06% de *outliers*). A inadimplência apresentou média elevada (4,16%) e variabilidade moderada (CV de 20,24%), porém sem *outliers* significativos, indicando uma alta inadimplência estrutural. O consumo de energia residencial (CV de 15,27%) e rendimento médio (CV de apenas 3,36%) apresentaram comportamento bastante estável, sugerindo menor volatilidade nesses indicadores regionais.

A região Sudeste caracterizou-se por relativa estabilidade econômica geral, evidenciada pelo IVV (média de 95,36 e CV baixo de 9,88%) e atividade econômica regional (IBCR com CV de 4,95%). Contudo, o desemprego permaneceu alto (média de 10,75%), embora constante, com variabilidade de 26,89%. Inadimplência (3,72%) e crédito pessoal (791,38 milhões) apresentaram momentos pontuais de elevação extrema, com *outliers* representando 3,03% e 3,79% dos casos, respectivamente. A inflação mostrou elevada volatilidade (CV de 78,06%), com episódios inflacionários pontuais (1,52% de *outliers*).

² Adotou-se a seguinte escala interpretativa: $CV < 15\%$ (baixa variabilidade), $15\% \leq CV \leq 30\%$ (variabilidade moderada) e $CV > 30\%$ (alta variabilidade), conforme Mingoti (2005) e Montgomery e Runger (2012). Além disso, curtose superior a 3 foi interpretada como indicativo de caudas pesadas, associadas a eventos extremos (Wilks, 2011), e a presença de *outliers* foi avaliada com base em critérios interquartis (IQR).

Tabela 1: Estatísticas descritivas para as variáveis quantitativas por região do Brasil

Região	Variável	Média	Mediana	Desv. Pad.	Coef Var.	Skew	Curtose	Mín	Máx	Outliers	Outliers (%)
Centro-Oeste	IVV	88,2264	87,7400	10,1036	11,4520	0,3067	2,7585	61,4800	116,5100	0,0000	0,0000
Centro-Oeste	IBCR	169,2117	165,7250	15,4553	9,1337	1,1748	4,7996	140,4600	227,9800	3,0000	2,2727
Centro-Oeste	IPCA	0,4748	0,4800	0,4615	97,2051	-0,2384	4,5623	-1,4600	1,7500	2,0000	1,5152
Centro-Oeste	Desemprego	8,6636	8,7500	2,3293	26,8861	0,1158	1,8392	4,9000	12,9000	0,0000	0,0000
Centro-Oeste	INADIMP	3,0173	2,9800	0,6274	20,7933	0,6814	3,0572	2,0200	4,6400	2,0000	1,5152
Centro-Oeste	CreditoPF	223,9900	202,0060	84,5591	37,7513	0,8273	2,9774	105,7180	445,2060	3,0000	2,2727
Centro-Oeste	EnergRes	0,9725	0,9595	0,1394	14,3306	0,2716	2,2913	0,6920	1,3070	0,0000	0,0000
Centro-Oeste	RM	3,4315	3,4060	0,1313	3,8253	0,7165	2,9982	3,2160	3,7590	6,0000	4,5455
Nordeste	IVV	89,0588	87,7150	10,4620	11,7473	0,4667	4,5945	55,1300	124,0900	4,0000	3,0303
Nordeste	IBCR	140,5371	140,7450	5,3284	3,7915	-0,4801	4,8971	119,8400	155,0300	4,0000	3,0303
Nordeste	IPCA	0,4971	0,4550	0,4357	87,6478	-0,1180	3,1953	-0,7800	1,6100	2,0000	1,5152
Nordeste	Desemprego	13,0886	13,9000	3,0778	23,5150	-0,0006	1,8161	8,0000	18,9000	0,0000	0,0000
Nordeste	INADIMP	4,4271	4,3300	0,6810	15,3826	0,4965	2,6899	3,3000	6,1300	1,0000	0,7576
Nordeste	CreditoPF	270,1823	249,2560	92,6048	34,2750	0,7477	2,9194	132,3100	507,0730	2,0000	1,5152
Nordeste	EnergRes	2,2869	2,2790	0,2747	12,0134	-0,0572	2,4247	1,6980	2,8390	0,0000	0,0000
Nordeste	RM	2,0236	2,0130	0,0894	4,4165	0,8293	3,5413	1,8770	2,2710	6,0000	4,5455
Norte	IVV	96,4677	94,2450	13,5781	14,0753	0,4862	2,8309	65,0300	133,1000	1,0000	0,7576
Norte	IBCR	162,1465	162,7300	7,5678	4,6673	0,0507	3,9615	142,7900	190,5300	3,0000	2,2727
Norte	IPCA	0,4814	0,5000	0,4609	95,7283	-0,0469	3,5007	-1,2500	1,6700	1,0000	0,7576
Norte	Desemprego	10,4864	11,3000	2,2991	21,9245	-0,0271	1,7673	6,6000	15,0000	0,0000	0,0000
Norte	INADIMP	4,1642	4,2500	0,8430	20,2427	0,0418	2,0862	2,7300	5,7900	0,0000	0,0000
Norte	CreditoPF	87,2683	76,0215	35,4014	40,5661	1,1281	3,5723	42,3840	187,4170	8,0000	6,0606
Norte	EnergRes	0,7707	0,7820	0,1177	15,2691	-0,2015	2,6728	0,5030	1,0440	0,0000	0,0000
Norte	RM	2,2765	2,2815	0,0766	3,3647	0,1713	2,2900	2,1270	2,4440	0,0000	0,0000
Sudeste	IVV	95,3569	94,3900	9,4200	9,8787	0,4305	3,9082	65,8700	123,8600	5,0000	3,7879
Sudeste	IBCR	138,7405	139,3500	6,8712	4,9525	-0,3166	2,8359	117,4800	153,0900	1,0000	0,7576
Sudeste	IPCA	0,4864	0,4600	0,3797	78,0555	0,3129	3,1228	-0,3600	1,4900	2,0000	1,5152
Sudeste	Desemprego	10,7523	11,5500	2,8909	26,8862	-0,0728	1,5805	6,3000	15,7000	0,0000	0,0000
Sudeste	INADIMP	3,7227	3,7100	0,6134	16,4779	0,5159	2,8081	2,6700	5,1700	4,0000	3,0303
Sudeste	CreditoPF	791,3848	727,4715	235,2619	29,7279	0,8306	3,0089	441,9550	1394,1870	5,0000	3,7879
Sudeste	EnergRes	5,5596	5,4945	0,4016	7,2241	0,5377	2,7304	4,8290	6,6410	2,0000	1,5152
Sudeste	RM	3,4905	3,4595	0,1637	4,6891	0,7302	2,9135	3,2440	3,9000	3,0000	2,2727
Sul	IVV	96,8016	97,0150	10,9661	11,3284	0,4311	3,1828	73,8000	126,4800	1,0000	0,7576
Sul	IBCR	141,9509	139,9550	10,2837	7,2445	1,5124	5,3688	125,8300	179,0200	11,0000	8,3333
Sul	IPCA	0,4891	0,4400	0,4744	96,9878	0,1899	3,9492	-0,9900	1,9900	3,0000	2,2727
Sul	Desemprego	6,6159	7,1000	1,7791	26,8918	-0,1640	1,5692	3,8000	9,6000	0,0000	0,0000
Sul	INADIMP	2,7326	2,6450	0,4990	18,2603	1,0178	3,7425	1,9200	4,1100	11,0000	8,3333
Sul	CreditoPF	344,5647	318,5210	111,8097	32,4496	0,7273	2,9054	175,7040	634,3750	4,0000	3,0303
Sul	EnergRes	1,8164	1,7795	0,2080	11,4527	0,6889	2,9391	1,4610	2,3920	0,0000	0,0000
Sul	RM	3,3101	3,2760	0,1510	4,5619	0,8300	3,1174	3,0480	3,6730	0,0000	0,0000

Por fim, a região Sul apresentou o mercado de trabalho mais sólido entre as regiões, com a menor média de desemprego (6,62%), embora com variabilidade alta (CV de 26,89%). Apesar da estabilidade do IVV (média de 96,80, CV de 11,33%), destacou-se negativamente pela elevada ocorrência de episódios extremos de inadimplência, com 8,33% dos casos considerados *outliers*, além de alta volatilidade inflacionária (CV de 96,99%) e presença de eventos inflacionários significativos (2,27% de *outliers*).

Estes resultados sugerem a existência de diferenças econômicas regionais relevantes. O Norte demonstra forte sensibilidade à expansão do crédito pessoal, enquanto o Sul, apesar de economicamente equilibrado, enfrenta desafios específicos na inadimplência e inflação. Centro-Oeste e Sudeste exibem estabilidade em indicadores como IVV e atividade econômica, mas têm desafios pontuais como inflação elevada e desemprego persistente, respectivamente.

Com relação ao comportamento da variável resposta ao longo do tempo, por intermédio da Figura 1 é possível observar uma forte sazonalidade no comportamento do IVV em todas as regiões analisadas, caracterizada especialmente por picos acentuados em dezembro. É possível supor que esses aumentos estejam diretamente relacionados ao crescimento das vendas e do consumo decorrentes das festas de fim de ano, estes contrastam com as quedas significativas observadas no início dos anos seguintes.

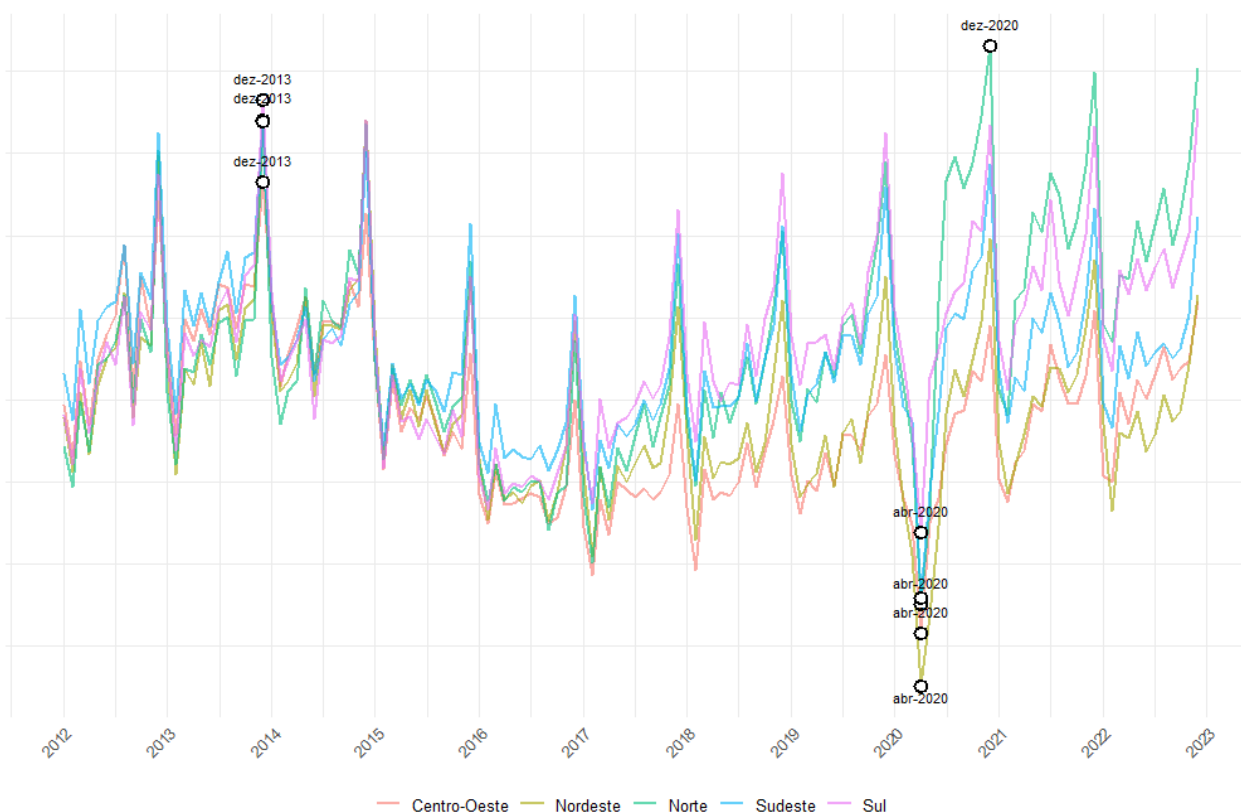
Também é possível verificar uma queda acentuada em todas as regiões para o mês de abril de 2020, refletindo claramente os impactos iniciais da pandemia de COVID-19, que reduziu fortemente tanto o consumo quanto a atividade econômica em geral. No entanto, poucos meses depois, todas as regiões exibiram uma recuperação bastante expressiva.

Observa-se que Norte e Nordeste apresentam variações mais intensas, com oscilações mais destacadas tanto nos picos quanto nas retrações. Por outro lado, Centro-Oeste, Sudeste e Sul mostram comportamentos mais sincronizados, embora com flutuações relativamente menos pronunciadas.

Considerando esses padrões, é possível indicar que os dados possam sugerir a utilização de variáveis de controle para capturar tanto os efeitos específicos da pandemia quanto efeitos sazonais. Essas variáveis poderiam ajudar a isolar e quantificar melhor os impactos desses eventos no comportamento do IVV.

As funções de Autocorrelação (ACF) e Autocorrelação Parcial (PACF) apresentadas no Anexo I reforçam visualmente essa análise, indicando padrões significativos de correlação sazonal nos períodos anuais (12 meses). Além disso, considerando o comportamento autocorrelacionado do IVV demonstrado nos gráficos ACF e PACF, pode-se avaliar também a inclusão de uma variável defasada do próprio IVV como preditora no modelo, uma vez que valores passados do índice podem ser significativos na explicação das observações futuras.

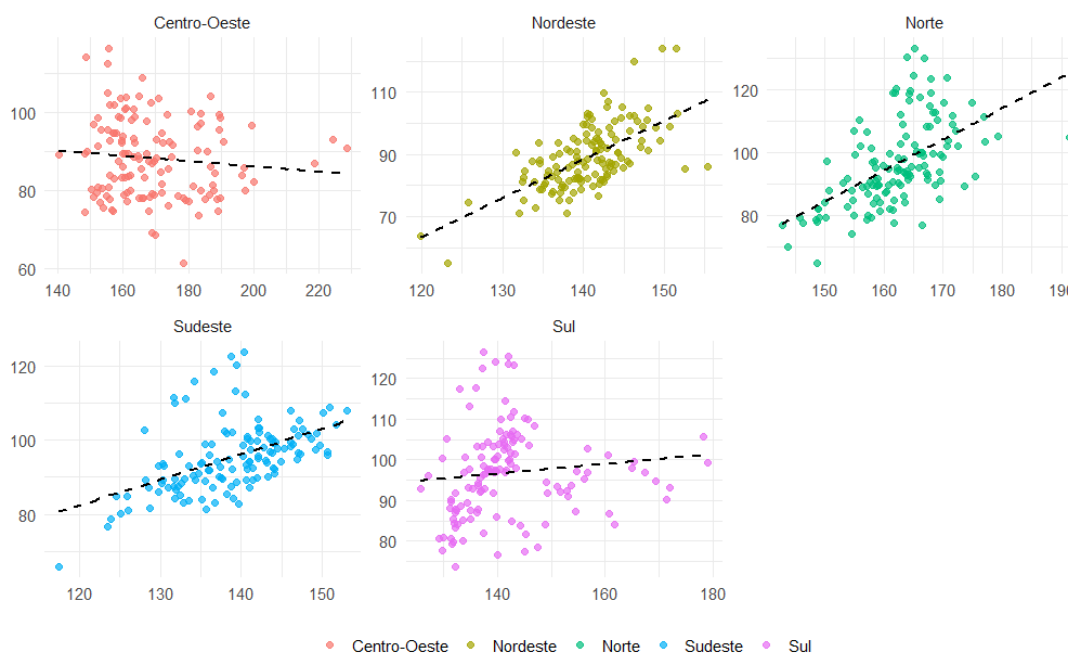
Figura 1 – Série temporal mensal do IVV por região do Brasil



Na sequência, as Figuras 2 a 8 trazem uma análise da relação entre a variável resposta IVV e os indicadores macroeconômicos que são calculados separadamente para cada região do país.

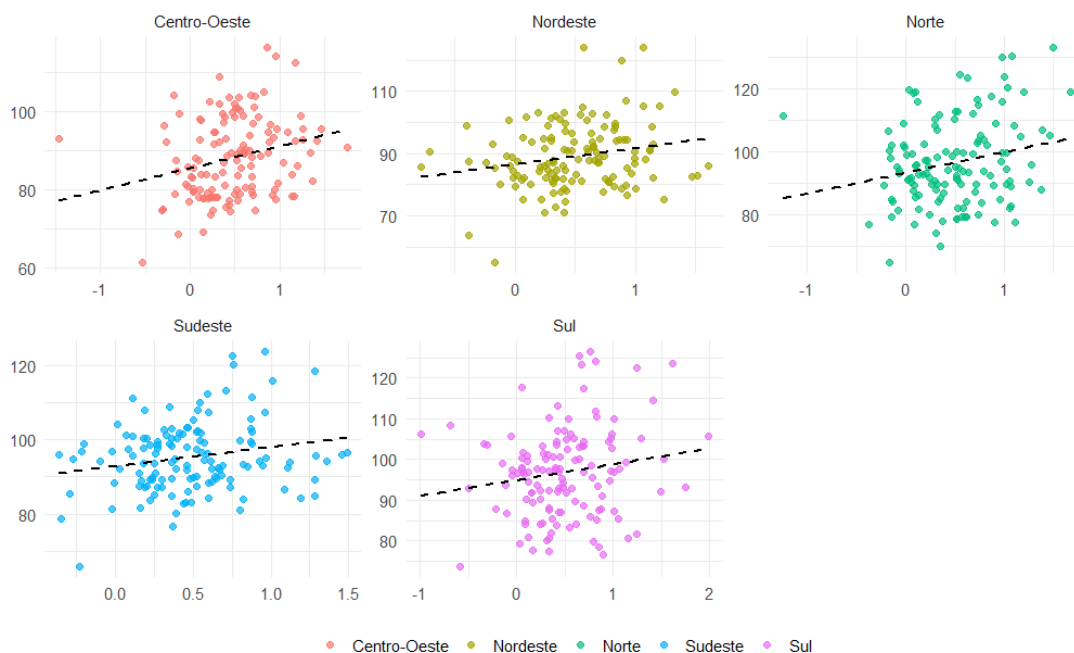
Por intermédio da Figura 2 é possível observar que as regiões Nordeste, Norte, Sudeste exibem linhas de tendência positivas, em graus variados de inclinação. Isso condiz com a hipótese de que a expansão da atividade econômica costuma aquecer o varejo. As regiões Centro-Oeste e Sul, por outro lado, sugere uma relação neutra, ligeiramente negativa para o Centro-Oeste, sinalizando que fatores específicos podem desconectar o crescimento do IBCR do consumo local.

Figura 2 – Dispersão do IVV *versus* IBCR calculado em cada região do Brasil



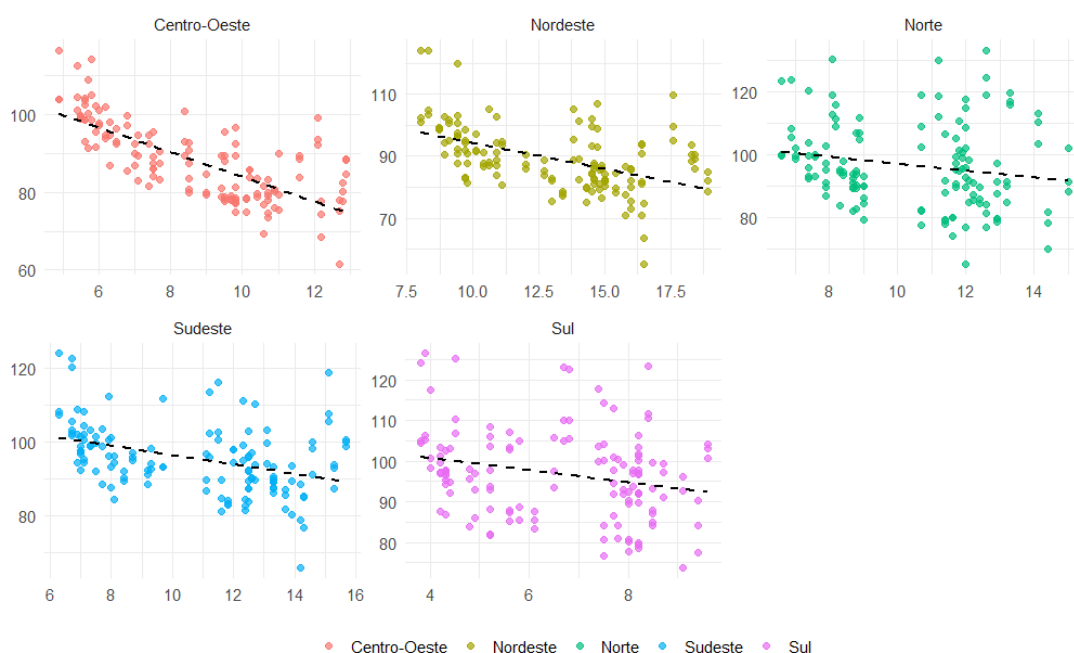
Ao observar a Figura 3 é possível verificar uma relação positiva entre IPCA e IVV. Conforme apontado na literatura, tal comportamento é esperado. Isso ocorre porque, em um cenário de alta inflação, os preços dos bens e serviços aumentam, fazendo com que o valor nominal das vendas seja inflacionado, mesmo que o volume físico de vendas não se altere significativamente (Kotler & Keller, 2012; Carvalho et al., 2020). Além disso, as variações regionais observadas podem ser atribuídas a múltiplos fatores que influenciam essa relação. Por exemplo, em determinadas regiões podem ocorrer políticas pontuais de redução de impostos que, ao diminuir o custo final dos produtos, podem não ser imediatamente refletidas no IVV. De forma similar, choques de oferta ou variações sazonais específicas podem ocasionar discrepâncias na relação entre o IPCA e o IVV (Rocha & Silva, 2018; Mendonça & Sachsida, 2019). Portanto, as conclusões obtidas a partir da análise dos dados estão alinhadas com a literatura, a qual evidencia que a relação entre inflação e vendas nominais no varejo é complexa e sujeita à influência de diversos fatores econômicos, fiscais e setoriais.

Figura 3 - Dispersão do IVV versus IPCA calculado em cada região do Brasil



A Figura 4 apresenta a relação entre o IVV e a taxa de desemprego, em todas as regiões observadas, é perceptível uma relação negativa entre o IVV as variáveis, ou seja, quanto maior a taxa de desemprego, menor tende a ser o índice de vendas no varejo. Entretanto, a intensidade desta relação é variável de acordo com a região, o Centro-Oeste destaca-se como uma das regiões mais sensíveis à variação no desemprego. Enquanto as regiões Norte e Sul mostram maior dispersão dos dados, sugerindo que outros fatores regionais podem ter uma influência mais relevante sobre o IVV.

Figura 4 - Dispersão do IVV versus Desemprego calculado em cada região do Brasil



A Figura 5 exibe como o Índice de Vendas do Varejo (IVV) se relaciona com Inadimplência (INADIMP) em cada região. Em teoria, poderíamos esperar que níveis mais altos de inadimplência pudessem restringir a capacidade de consumo, pois consumidores endividados ou em atraso têm menor acesso a crédito e maior comprometimento de renda, entretanto, a análise dos gráficos sugere relações pouco claras ou até inversas. Em muitas regiões, um aumento da inadimplência não se traduz de forma imediata nem proporcional em queda no varejo. Em alguns casos (Norte, Sul), há leve inclinação negativa na tendência, sugerindo que a alta da inadimplência pode, sim, pressionar o consumo para baixo. Em outros (Centro-Oeste, Nordeste, Sudeste), a linha é praticamente horizontal ou até levemente positiva, sugerindo que outros fatores (renda, transferências, crédito ampliado, sazonalidades) podem estar compensando o efeito da inadimplência.

Por intermédio da Figura 6 podemos verificar que as regiões apresentam comportamentos distintos na relação entre crédito pessoal e o volume de vendas no varejo. Não há um padrão único em todas as regiões, sugerindo especificidades regionais. As regiões Norte e Sul mostram maior sensibilidade positiva do IVV ao crédito pessoal, sugerindo uma economia varejista possivelmente impulsionada por crédito. As regiões Centro-Oeste e Nordeste não mostram grande sensibilidade ao crédito pessoal, com dispersões elevadas, destacando que outros fatores econômicos possivelmente têm papel mais relevante. Por fim, o Sudeste apresenta tendência levemente positiva, mas não conclusiva.

Figura 5 - Dispersão do IVV versus Inadimp calculado em cada região do Brasil

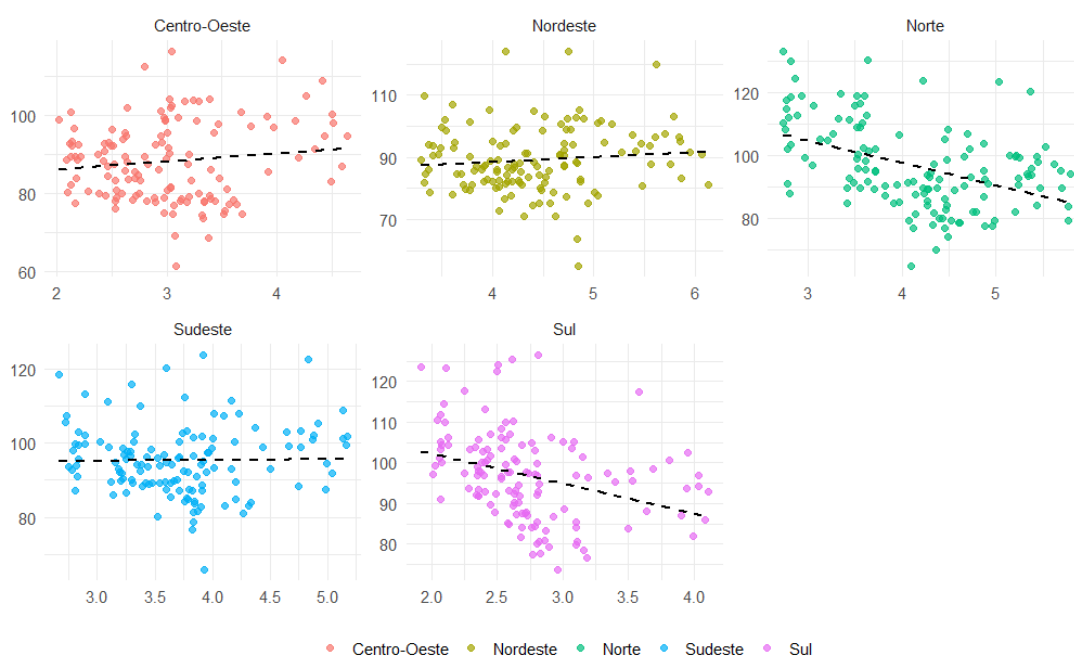
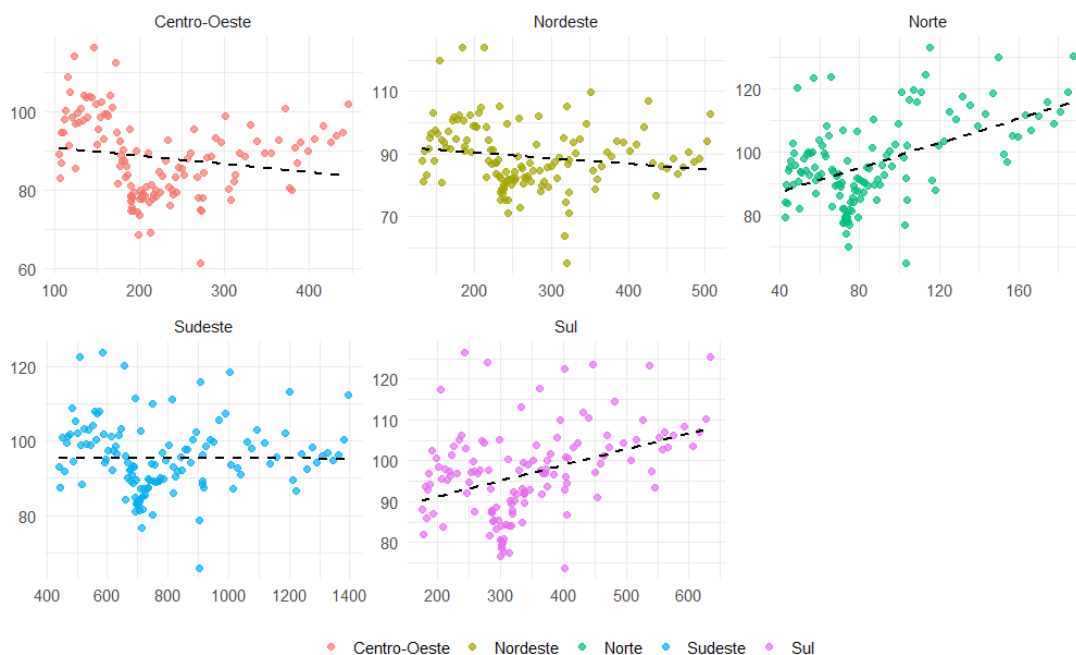
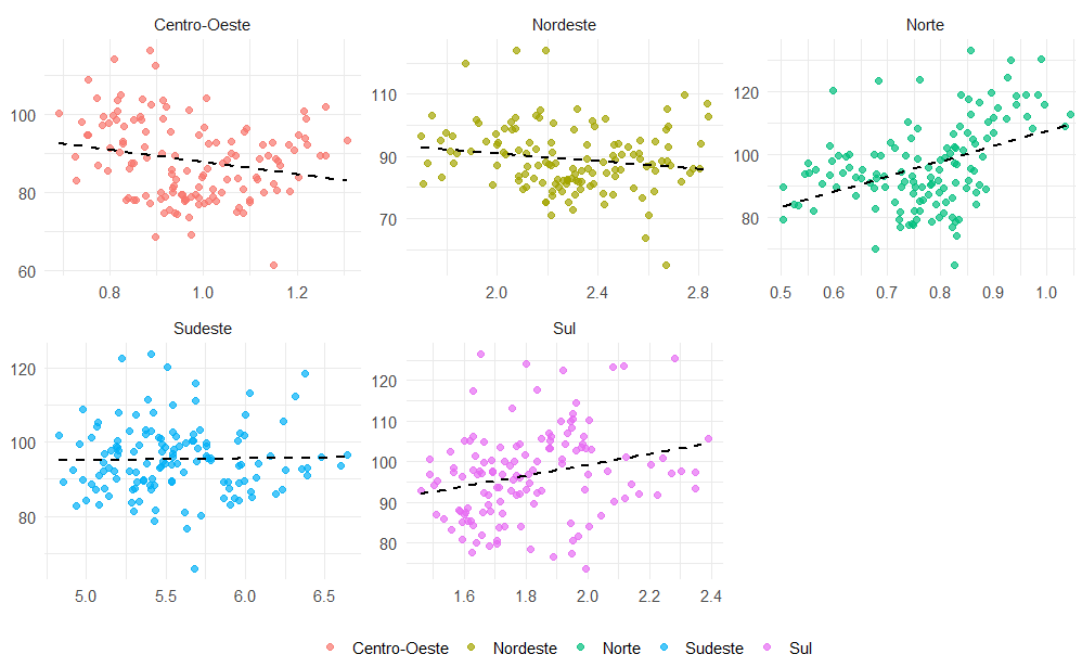


Figura 6 - Dispersão do IVV versus Credito PF calculado em cada região do Brasil



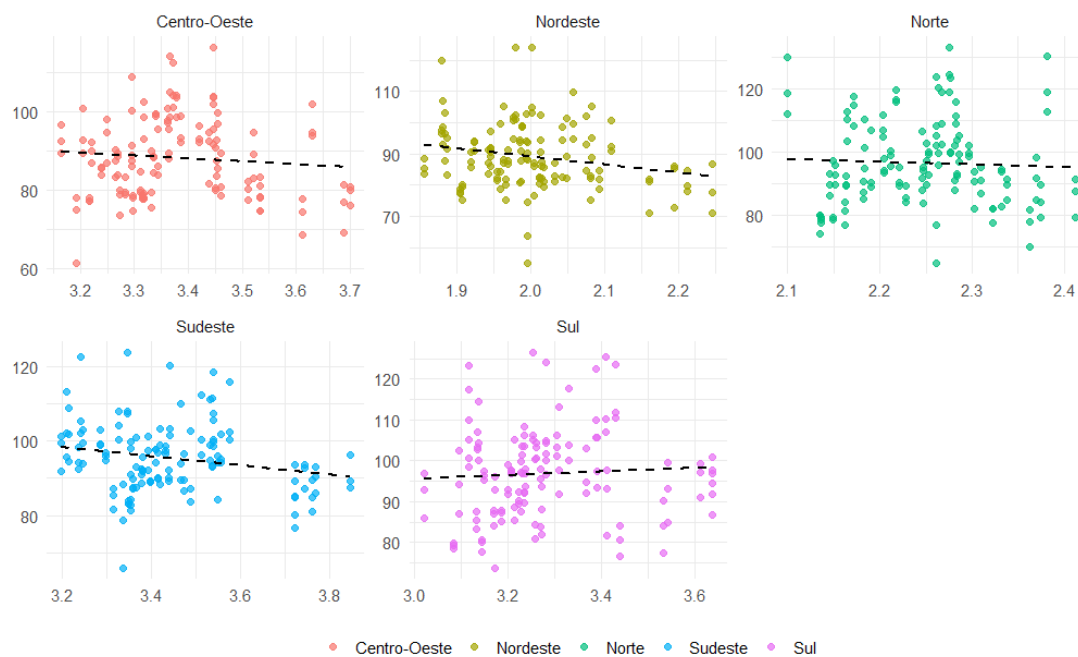
A Figura 7 apresenta a relação entre consumo de energia e IVV. As regiões Norte e Sul mostram tendência positiva mais definida, indicando que nestas regiões o consumo residencial de energia elétrica pode refletir maior dinamismo econômico regional ou aumento de renda doméstica. Em contraposição, Centro-Oeste e Nordeste não apresentam uma relação clara, indicando baixa relevância desta variável para explicar o comportamento do IVV. Por fim o Sudeste, apesar da relação positiva, apresenta fraca associação, indicando a importância limitada dessa variável como determinante do desempenho do varejo.

Figura 7 - Dispersão do IVV versus EnergRes calculado em cada região do Brasil



Finalizando a análise de dispersão temos a Figura 8 na qual é possível verificar uma relação fraca e pouco consistente entre IVV e a renda média. A maioria das regiões mostra retas com inclinação próxima de zero ou levemente negativa, acompanhadas de alta dispersão dos dados. Este comportamento sugere que o Rendimento Médio mensal real não é um forte preditor direto do volume de vendas no varejo regional, pelo menos de forma isolada.

Figura 8 - Dispersão do IVV versus RM calculado em cada região do Brasil



A análise das matrizes de correlação, Figura 9, revelou relações importantes entre as variáveis independentes e a dependente (IVV) com cada um desses indicadores macroeconômicos que são calculados separadamente para cada região do país. Além disso, os resultados permitem também avaliar o impacto das correlações existentes entre as próprias variáveis independentes, pois isso pode gerar problemas de multicolinearidade em modelos estatísticos preditivos.

Na Região Norte, o IVV apresentou correlação positiva moderada com o Crédito à Pessoa Física (0,50) e com o consumo de Energia Residencial (0,42), sugerindo que o consumo no varejo está associado à expansão do crédito e à atividade domiciliar. No entanto, a alta correlação entre Crédito à Pessoa Física e Energia Residencial (0,83) indica forte multicolinearidade, o que dificulta a identificação do efeito isolado de cada variável em modelos de regressão. Adicionalmente, a Inadimplência revelou correlação negativa com o IVV (-0,44) e ainda mais forte com Energia Residencial (-0,81), apontando para sua influência difusa sobre variáveis explicativas relevantes, o que reforça a necessidade de cautela na sua inclusão em modelos simultâneos.

Na Região Nordeste, o IVV apresentou correlação positiva moderada com o Índice de Atividade Econômica Regional (IBCR) (0,64) e negativa com o Desemprego (-0,49). A Energia Residencial, por sua vez, manteve forte associação com o Crédito à Pessoa Física (0,88) e elevada correlação com a Inadimplência (-0,73), configurando um cenário de interdependência estrutural entre variáveis que, embora relevantes para o IVV, podem dificultar a distinção de seus efeitos diretos em modelos econométricos. Essa redundância potencial pode comprometer a robustez das estimativas se não for devidamente tratada.

Na Região Centro-Oeste, a Inadimplência mostrou-se fortemente correlacionada negativamente tanto com a Energia Residencial (-0,79) quanto com o Crédito à Pessoa Física (-0,78), o que pode afetar a estabilidade de modelos que incluam ambas simultaneamente. Embora o Crédito à Pessoa Física tenha exibido uma das maiores correlações com Energia Residencial (0,86), sua associação direta com o IVV foi negativa e fraca (-0,17), sugerindo que seu papel no consumo é indireto ou mediado por outras variáveis. O Desemprego apresentou forte correlação negativa com IVV (-0,73).

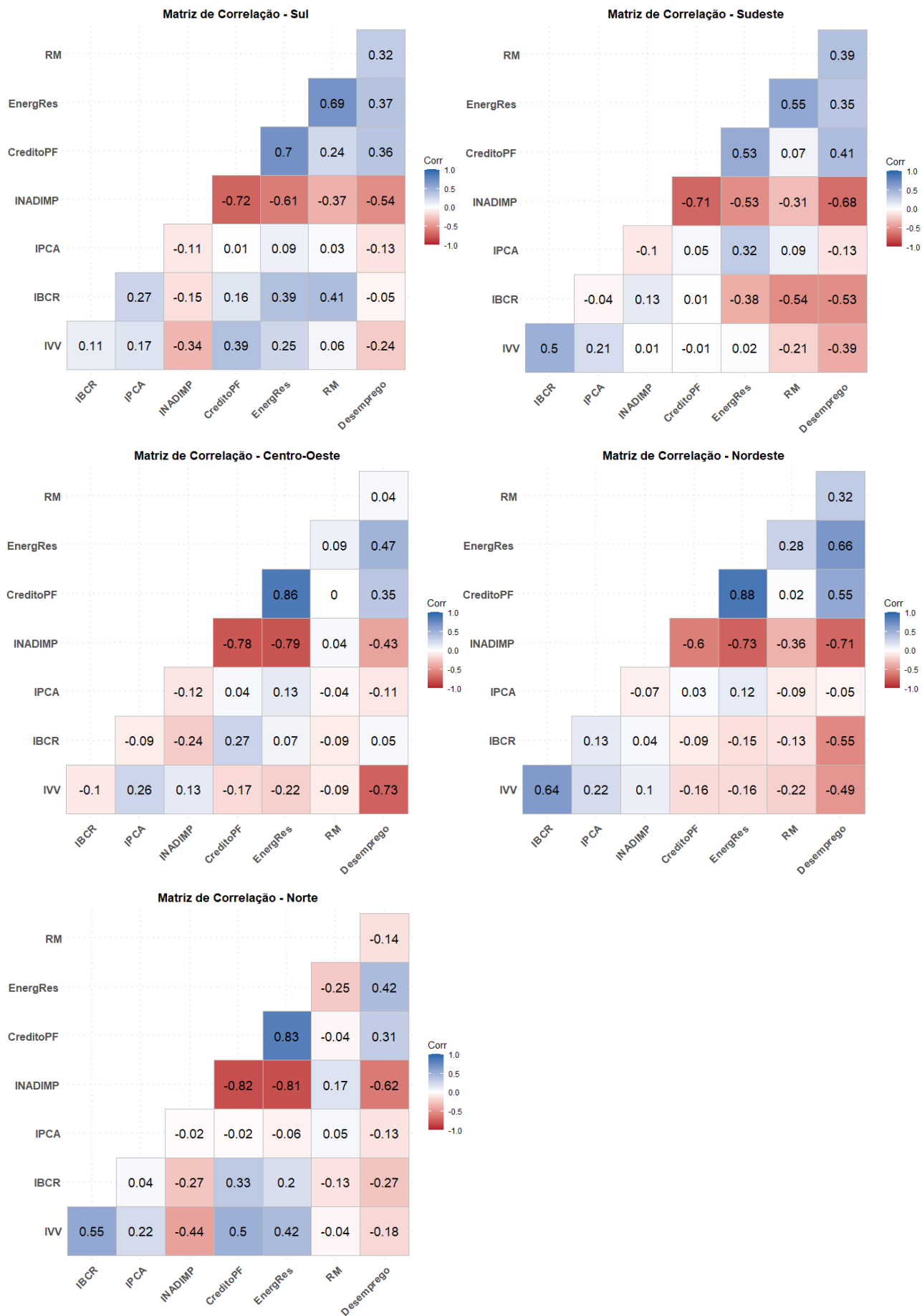
Na Região Sudeste, o IBCR demonstrou correlação positiva com IVV (0,50), destacando a atividade econômica regional como principal motor do consumo varejista. A Inadimplência apresentou forte correlação negativa com o Crédito à Pessoa Física (-0,71), mas, ao contrário das demais regiões, não exibiu correlações elevadas com as outras variáveis independentes, o que sugere menor risco de multicolinearidade. A Energia Residencial e a Renda Média apresentaram correlação positiva entre si (0,55), porém seus efeitos sobre o IVV foram mais limitados.

Na Região Sul, o IVV se correlacionou fracamente com a Energia Residencial (0,25) e com o Crédito à Pessoa Física (0,39), enquanto a correlação entre essas duas variáveis foi igualmente elevada (0,70), indicando possível redundância. A Inadimplência apresentou correlação negativa acentuada com o Crédito à Pessoa Física (-0,72) e com a Energia Residencial (-0,61), formando um bloco de variáveis que se movimentam em conjunto, dificultando a distinção de seus efeitos individuais em um modelo explicativo.

De modo geral, observa-se que a variável consumo de energia elétrica apresentou correlações elevadas com diversas variáveis explicativas em todas as regiões analisadas.

Especificamente, foram verificadas correlações expressivas com o Crédito à Pessoa Física nas regiões Norte (0,83), Nordeste (0,88), Centro-Oeste (0,86) e Sul (0,70). Além disso, a energia residencial também apresentou elevada correlação negativa com a inadimplência nas regiões Centro-Oeste (-0,79) e Norte (-0,81), bem como com o desemprego no Nordeste (-0,73). Tais resultados indicam a existência de um bloco de variáveis altamente interdependentes, sugerindo a existência de multicolinearidade.

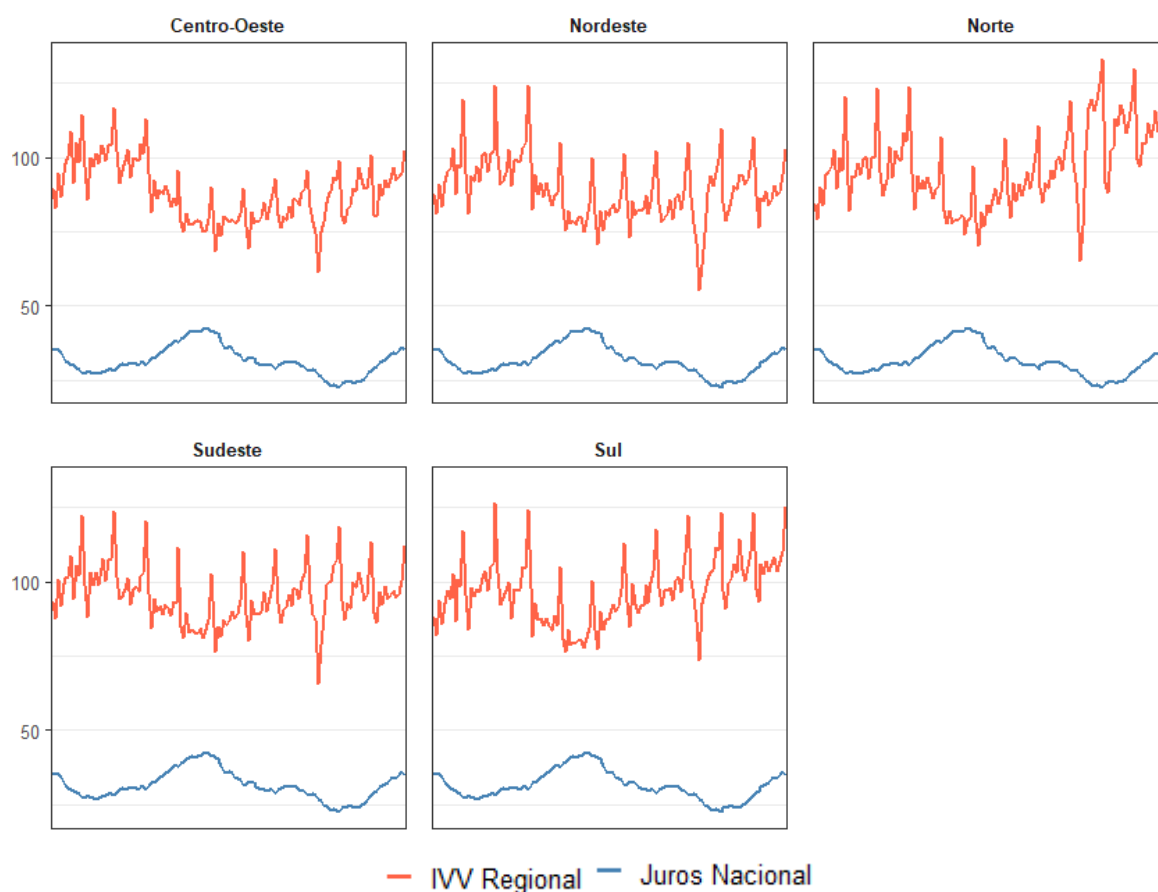
Figura 9 – Correlação do IVV com os indicadores calculados em cada região do Brasil



Por fim, as Figuras 10 e 11 trazem uma análise da relação entre a variável resposta IVV e os indicadores macroeconômicos que são calculados ao nível nacional, sendo a análise estratificada para cada região afim de identificar possíveis padrões regiões.

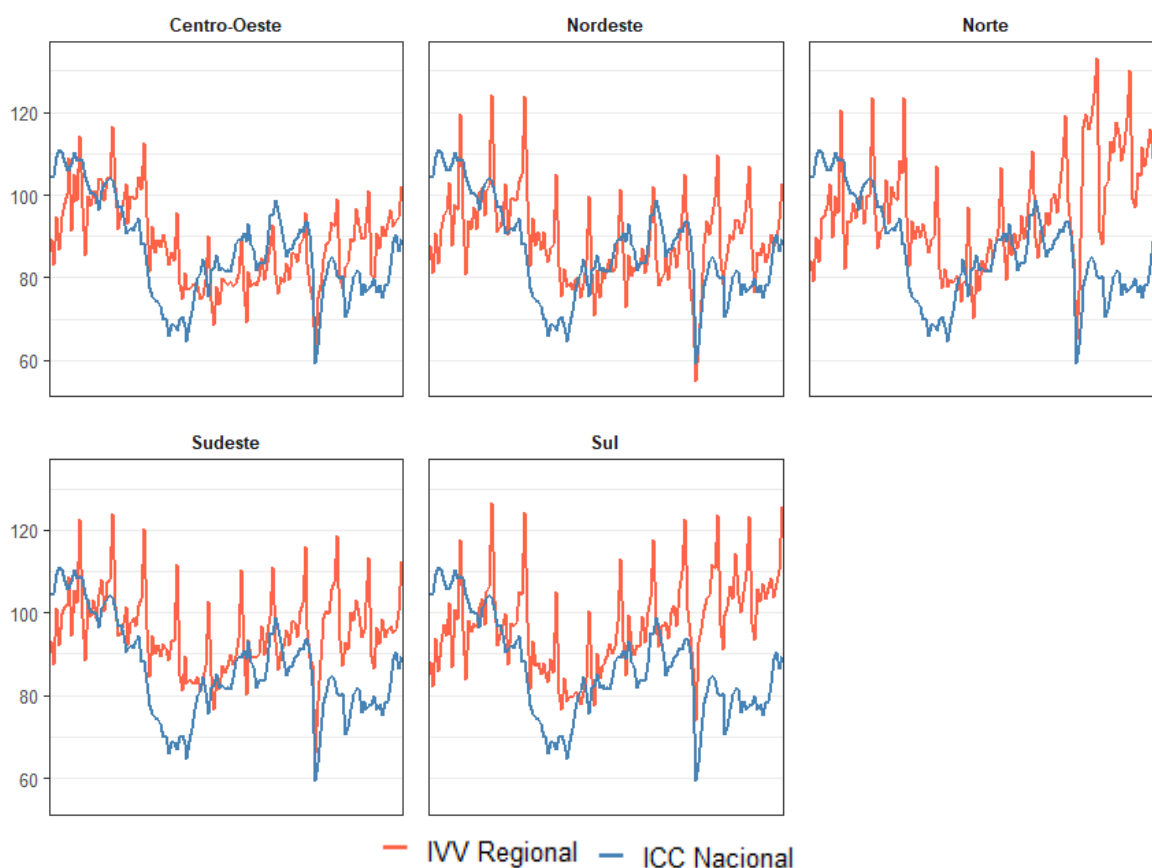
A Figura 10 apresenta a comparação temporal entre o IVV, variável regional, e a taxa de Juros nacional, nas diferentes regiões brasileiras. Observa-se claramente que o comportamento dos juros nacionais segue um padrão homogêneo e cíclico bem definido em todas as regiões analisadas, caracterizado por períodos alternados de aumento e redução das taxas. Em contrapartida, o IVV exibe uma dinâmica regional mais heterogênea, evidenciando diferentes graus de sensibilidade às variações das taxas de juros. Notadamente, as regiões Centro-Oeste, Nordeste, Sudeste e Sul mostram uma correlação visual evidente, com períodos de juros elevados coincidindo com redução ou estabilização no IVV, sugerindo uma contração das vendas varejistas sob condições monetárias mais restritivas. Por outro lado, reduções nas taxas de juros são consistentemente acompanhadas por períodos de crescimento gradual no IVV. A região Norte, embora apresente comportamento semelhante, demonstra uma resposta menos imediata e de menor intensidade aos ciclos das taxas de juros nacionais, sugerindo que fatores econômicos ou estruturais locais possam atenuar ou moderar o impacto dessas variações sobre o desempenho regional do varejo.

Figura 10 – Comparação Temporal: IVV Regional versus Juros Nacional



A Figura 11 exibe a comparação entre o IVV regional e o ICC nacional. Observa-se que o comportamento do ICC nacional é homogêneo entre as regiões, com períodos claros de queda acentuada. O IVV regional exibe forte correspondência com as variações no ICC, revelando que quedas na confiança do consumidor são acompanhadas por retrações ou desacelerações visíveis nas vendas varejistas em todas as regiões analisadas. As regiões Centro-Oeste, Nordeste, Sudeste e Sul demonstram padrões semelhantes, com respostas imediatas e perceptíveis às mudanças no ICC, especialmente em momentos críticos. Destaca-se, no entanto, a região Norte, que embora apresente uma tendência alinhada às demais, exibe maior volatilidade e uma reação mais intensa às oscilações do ICC, sugerindo uma maior vulnerabilidade ou exposição regional às variações na confiança do consumidor.

Figura 11 - Comparação Temporal: IVV Regional versus ICC Nacional



A comparação das correlações, Tabela 2, com as análises gráficas reforça as observações feitas visualmente. A correlação negativa entre IVV e taxas de juros, mais intensa nas regiões Norte (-0,554) e Sul (-0,559), está claramente refletida nos gráficos, onde se observa uma visível redução das vendas nos períodos de juros elevados. Por outro lado, apesar da correlação menos forte observada no Nordeste (-0,308) e Centro-Oeste (-0,364), as análises gráficas mostram,

ainda assim, sensibilidade às mudanças cíclicas das taxas de juros, porém com menos intensidade.

Quanto ao ICC, as correlações positivas mais elevadas (Centro-Oeste: 0,525; Sudeste: 0,450; Nordeste: 0,419) também estão evidentes nos gráficos. Nessas regiões, períodos marcantes de queda ou recuperação do ICC coincidem claramente com movimentos semelhantes no IVV regional. Em contrapartida, as correlações mais baixas encontradas nas regiões Norte (0,106) e Sul (0,253) são coerentes com as observações gráficas, nas quais a associação visual entre ICC e IVV é menos nítida ou mais volátil, especialmente na região Norte, destacando a influência de fatores adicionais sobre o comportamento das vendas.

Tabela 2 - Correlação do IVV com os indicadores calculados em cada região do Brasil

Região	Corr. IVV x Juros	Corr. IVV x ICC
Centro-Oeste	-0,364	0,525
Nordeste	-0,308	0,419
Norte	-0,554	0,106
Sudeste	-0,448	0,450
Sul	-0,559	0,253

3.2 Ajuste da regressão linear múltipla

Para investigar a relação entre os indicadores econômicos e o IVV, foram conduzidos ajustes progressivos, conforme ilustrado no Tabela 3. Todos os modelos apresentados nesta seção foram estimados com base no método dos mínimos quadrados ordinários (MQO), exceto quando explicitamente indicado o uso de ponderações ou transformações. Na construção das variáveis de controle (*dummies*), a Região Sudeste e o primeiro trimestre foram utilizados como categorias de referência, permitindo a interpretação dos coeficientes das demais categorias em relação a esses grupos-base.

No Ajuste 1, foi aplicada a estimação por mínimos quadrados ordinários, incluindo todas as variáveis explicativas disponíveis – IBCR, IPCA, INADIMP, CreditoPF, EnergRes, Desemprego, Juros, ICC e Região, sem transformações ou inclusão de termos defasados.

Este ajuste apresentou elevada multicolinearidade, em especial decorrente da variável EnergRes, conforme indicado pelo elevado valor do VIF, Tabela 4. Além disso, os pressupostos fundamentais do modelo de regressão foram substancialmente violados. O teste de normalidade de Shapiro-Wilk evidenciou que os resíduos não seguem uma distribuição, comprometendo a validade dos testes de hipóteses, Tabela 5.

O teste de autocorrelação de Breusch-Godfrey revelou a presença de autocorrelação significativa nos resíduos, Tabela 5. Essa dependência serial implica que os erros padrão dos coeficientes pode estar subestimados, o que afeta a robustez das inferências estatísticas derivadas do modelo.

A partir da análise descritiva dos dados apresentada na Seção 4.1, constatou-se que o IVV exibe características temporais marcantes, com evidentes padrões sazonais – notadamente picos acentuados em dezembro – e uma queda abrupta em abril de 2020, atribuída aos impactos iniciais da pandemia de COVID-19. Tais evidências empíricas justificaram a incorporação de variáveis que pudessem capturar tanto a sazonalidade quanto os efeitos exógenos associados à pandemia.

Nesse contexto, o Ajuste2, mantendo-se o método de estimação por MQO, foi elaborado com o intuito de aprimorar a especificação do modelo para atender aos padrões observados na autocorrelação dos dados, evidenciada nos gráficos de ACF e PACF, foi incluído o lag do IVV (lag_IVV), o qual permitiu capturar a persistência temporal inerente à série.

Complementarmente, foram incorporadas variáveis controle para os trimestres, a fim de representar as variações sazonais, bem como indicadores específicos para os efeitos da pandemia (pandemia³ e abril_2020⁴), que refletem os eventos atípicos identificados na análise descritiva. Essas modificações resultaram em uma melhora substancial do desempenho do modelo, elevando o R² ajustado de 0,5689 para 0,757, além de aprimorar os diagnósticos relativos à heterocedasticidade e à autocorrelação dos resíduos, conforme comprovado pelos testes estatísticos aplicados, Tabela 5.

No Ajuste3, a especificação adotada no Ajuste2 foi mantida, mas a estimação dos parâmetros foi feita via método de mínimos quadrados ponderados considerando a ponderação baseada no inverso do quadrado do IVV, isto é, $weights = 1/(IVV^2)$. Essa abordagem foi escolhida após a realização de testes comparativos com outras alternativas de ponderação (tais como $weights = IVV$ e $weights = 1/IVV$), os quais não demonstraram desempenho superior quanto à correção da heterocedasticidade. A aplicação dessa estratégia resultou em uma melhora significativa no modelo, evidenciada pelo aumento do R² ajustado para 0,7734. Este ajuste também apresentou melhoria relevante no teste de variância dos resíduos, Tabela 5.

A transição do Ajuste 3 para o Ajuste 4 envolveu duas etapas metodológicas complementares e sequenciais. Inicialmente, manteve-se a estimação via mínimos quadrados ponderados, e foi aplicada uma transformação logarítmica à variável dependente (IVV), com o objetivo de melhorar a aderência dos resíduos à suposição de normalidade. Importante destacar

³ 1 para o período 01-2020 a 06-2020, 0 para o restante

⁴ 1 para 04-2020, 0 para o restante

que, após essa transformação, verificou-se que os níveis de significância das variáveis explicativas permaneceram praticamente inalterados, não comprometendo a estrutura inicial do modelo.

Em seguida, aplicou-se o procedimento de seleção backward, método detalhado na seção 2.2.1, buscando otimizar a especificação do modelo por meio da exclusão sistemática de variáveis não significativas, de modo a alcançar uma estrutura mais enxuta e estatisticamente consistente. Das variáveis inicialmente incluídas no modelo, foram mantidas apenas aquelas que apresentaram significância estatística, totalizando 13 termos: IBCR, Desemprego, Juros, RegS, RegNO, RegNE, RegCo, Lag_IVV, Trimestre2, Trimestre3, Trimestre4, Pandemia e Abril_2020. Consequentemente, foram excluídas as variáveis IPCA, INADIMP, CréditoPF e ICC.

O modelo resultante, denominado Ajuste 4, apresentou desempenho superior em termos dos diagnósticos dos resíduos e aderência aos pressupostos da regressão linear. O teste de Shapiro-Wilk evidenciou que os resíduos se aproximam de uma distribuição normal (p -value = 0.08), o que reforça a adequação desta especificação para inferências estatísticas.

Adicionalmente, apesar de o R^2 ajustado ter aumentado para 0,7947 e o AIC ter atingido o valor de -1700.258, é importante salientar que o uso do AIC como medida comparativa entre modelos pode ser problemático quando há transformação logarítmica da variável dependente. Essa transformação altera a escala dos resíduos e, consequentemente, inviabiliza comparações diretas do AIC com modelos que operam com a variável IVV em sua forma original.

Portanto, diante da limitação do uso do AIC como critério comparativo, em razão da transformação logarítmica aplicada à variável dependente, adotou-se como principal critério de seleção o fato de o Ajuste 4 ser o único a atender satisfatoriamente aos pressupostos da regressão linear. Especificamente, seus resíduos apresentaram distribuição aproximadamente normal, homocedasticidade e ausência de autocorrelação, conforme evidenciado na Tabela 5.

Em síntese, a evolução dos modelos evidenciou a importância de uma abordagem iterativa, criteriosa e metodologicamente fundamentada. A exclusão de variáveis com elevada autocorrelação, como EnergRes, a inclusão de termos capazes de captar a dinâmica temporal e os efeitos sazonais e pandêmicos, a aplicação de técnicas de ponderação e transformações apropriadas, bem como o uso do procedimento de seleção backward para a definição de um modelo mais parcimonioso, culminaram na formulação do Ajuste 4. Esse modelo final demonstrou-se estatisticamente robusto, com melhor ajuste aos dados observados e boa aderência aos pressupostos da regressão linear.

Tabela 3 – Resumo e métricas de avaliação dos modelos ajustados

Coefficiente	Ajuste1	Ajuste2	Ajuste3	Ajuste4 (Log(IVV))
Intercepto	149,2051***	65,5350***	64,060***	4,1666***
IBCR	0,0079	0,1861***	0,1839***	0,0021***
IPCA	3,9006***	0,3859	-0,1515	
INADIMP	-8,1611***	-0,7771	-0,6038	
CreditoPF	0,0090 ^o	-0,0055	-0,0026	
EnergRes	-2,0857			
Desemprego	-2,4302***	-0,6464***	-0,7063***	-0,0077***
Juros	-0,5368***	-0,6682***	-0,6310***	-0,0071***
ICC	0,2506***	0,0039	0,0056	
RegS	-21,6692***	-5,9185**	-5,1400*	-0,0389***
RegNO	-0,9692**	-7,7623**	-6,6224*	-0,0616***
RegNE	2,3661	-5,0914*	-3,9855 ^o	-0,0376***
RegCO	-22,8047**	-15,1630***	-14,0342***	-0,1388***
Lag_IVV		0,3509***	0,3378***	0,0037***
Trimestre2		3,1810***	2,9310***	0,0325***
Trimestre3		4,9154***	4,5076***	0,0502***
Trimestre4		13,1644***	12,0503***	0,1272***
Pandemia		-1,7570	-2,5688*	-0,0362**
Abril_2020		-19,0814***	-18,6884***	-0,2602***
Medidas				
R2 Ajustado	0,5689	0,757	0,7734	0,7947
QMRes	57,3057	32,52402	0,0034	3,866936e-07
AIC	4146,623	3780,036	3715,086	-1700,258

*** p < 0,001; ** p < 0,01; * p < 0,05; ^o p < 0,1

Valores arredondados para quatro casas decimais

Tabela 4 – Fator de Inflação de Variância (VIF) nos modelos ajustados

Coefficiente	Ajuste1	Ajuste2	Ajuste3	Ajuste4
IBCR	1,8352	1,9722	1,9761	1,8967
IPCA	1,0840	1,2842	1,3104	
INADIMP	2,4512	2,8500	2,8228	
CreditoPF	4,2345	4,2421	4,4531	
EnergRes	10,0929			
Desemprego	2,1898	2,5321	2,5967	1,5428
Juros	1,3802	1,6802	1,6963	1,2865
ICC	1,4528	1,8033	1,7842	
Região	2,5359	2,1199	2,1596	1,2681
Lag_IVV		1,4886	1,4964	1,4054
Trimestre		1,1071	1,1086	1,0430
Pandemia		1,3237	1,4560	1,2901
Abril_2020		1,1545	1,2533	1,1716

Valores arredondados para quatro casas decimais

Tabela 5 – Testes normalidade, homocedasticidade e autocorrelação para os resíduos dos modelos ajustados

Ajuste	Shapiro-Wilk	Breusch-Pagan Modificado	Breusch-Godfrey
Ajuste1	W = 0,9729, p-value = 4,304e-09	Chisquare = 22,4146, df = 1, p-value = 2,197e-06	LM test = 246,45, df = 1, p-value = 2,2e-16
Ajuste2	W = 0,9762, p-value = 3,122e-08	Chisquare = 38,1859, df = 1, p-value = 6,4315e-10	LM test = 0,4406, df = 1, p-value = 0,5068
Ajuste3	W = 0,9868, p-value = 3,37e-05	Chisquare = 8,1344, df = 1, p-value = 0,0043	LM test = 0,4406, df = 1, p-value = 0,5068
Ajuste4	W = 0,9955, p-value = 0,08	Chisquare = 2,6037 , df = 1 , p-value = 0,1066	LM test = 0,19553, df = 1, p-value = 0,6584;

3.3 – Análise preditiva

Conforme Tabela 6, os resultados globais obtidos indicam que o MAE no conjunto de teste é de aproximadamente 5,48 unidades, enquanto o RMSE é de cerca de 6,97 unidades. Esses valores sugerem que, em média, as previsões do modelo diferem dos valores observados em 5,48 unidades do IVV, com erros maiores que podem chegar, tipicamente, a 7 unidades. Se considerarmos que os valores médios do IVV nas diferentes regiões giram em torno de 90 a 100, esses erros representam uma variação percentual relativamente pequena, sugerindo uma boa capacidade preditiva do ajuste final. Além disso, a diferença entre RMSE e MAE indica que, embora a maioria dos erros seja moderada, há alguns casos com desvios maiores, o que é comum em séries com características sazonais ou eventos atípicos.

Adicionalmente, testou-se a eficiência do modelo por região e os resultados são também apresentados na Tabela 6. Observa-se que o modelo apresenta os menores valores de erro na região Nordeste, com MAE e RMSE de 3,29 e 3,85, respectivamente. Esses indicadores sugerem que as previsões para o Nordeste estão, em média, mais próximas dos valores observados, evidenciando uma melhor capacidade preditiva do modelo nessa região. Em contrapartida, as regiões Norte e Sul apresentam erros significativamente maiores, indicando que o modelo é menos eficiente nesses contextos. A região Centro-Oeste situa-se em um patamar intermediário, enquanto o Sudeste, embora com métricas um pouco superiores às do Nordeste, ainda demonstra uma performance satisfatória. Portanto, pode-se concluir que, entre as regiões analisadas, o ajuste final foi mais eficiente na previsão dos valores do IVV para a região Nordeste e menos precisos para a região Norte.

Tabela 6 – Métricas de avaliação do melhor modelo ajustado considerando os dados de teste

Região	MAE	RMSE
Nordeste	3,29	3,85
Sudeste	3,82	4,74
Sul	6,67	8,09
Centro-Oeste	4,47	5,38
Norte	9,17	10,6
Brasil	5,48	6,97

Tabela 7 – Métricas de avaliação do melhor modelo ajustado considerando os dados de treino

Região	MAE	RMSE
Nordeste	3,97	5,21
Sudeste	3,94	5,19
Sul	4,29	5,65
Centro-Oeste	4,19	5,43
Norte	5,20	6,63
Brasil	4,32	5,65

Quando comparamos os resultados com o conjunto de treino, Tabela 7, é possível verificar que para as regiões Nordeste e Sudeste, o modelo apresentou desempenho consistente tanto no conjunto de treinamento quanto no conjunto de teste, demonstrando boa capacidade de generalização. Na região Nordeste, por exemplo, o erro médio absoluto (MAE) apresentou uma redução de 3,97 unidades no treinamento para 3,29 no teste, enquanto o erro quadrático médio (RMSE) caiu de 5,21 para 3,85 unidades, sugerindo uma adequada robustez do modelo para capturar as dinâmicas econômicas locais recentes. A região Sudeste apresentou comportamento semelhante, com um leve decréscimo no MAE (de 3,94 para 3,82) e no RMSE (de 5,19 para 4,74), reforçando sua confiabilidade preditiva.

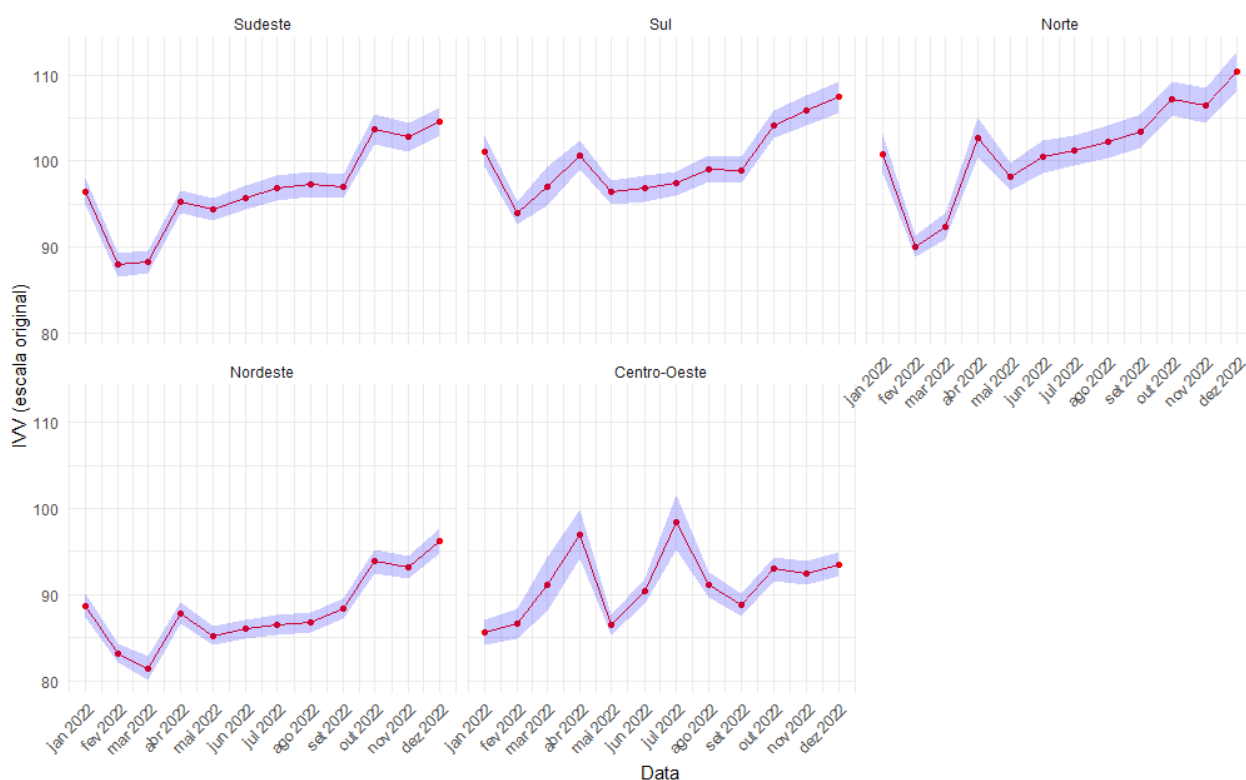
Por outro lado, regiões como Sul e Norte enfrentaram maiores desafios na generalização do modelo. Na região Sul, observou-se uma significativa deterioração da performance, com aumento do MAE de 4,29 para 6,67 unidades e do RMSE de 5,65 para 8,09 unidades na passagem do treinamento para o teste, indicando possíveis eventos econômicos atípicos ou mudanças estruturais recentes não adequadamente captadas pelo modelo. Da mesma forma, na região Norte, os erros aumentaram substancialmente (MAE subiu de 5,20 para 9,17 e RMSE de 6,63 para 10,6 unidades), revelando limitações consideráveis do modelo na previsão recente dessa região específica.

Na região Centro-Oeste, o desempenho permaneceu estável e satisfatório, com pequenas variações dos erros entre treinamento e teste (MAE de 4,19 para 4,47 e RMSE de 5,43 para 5,38), evidenciando boa adequação do modelo para acompanhar a dinâmica econômica regional.

No panorama nacional, o desempenho global do modelo mostrou-se razoavelmente satisfatório, embora tenha apresentado uma leve piora na capacidade preditiva, com o MAE passando de 4,32 para 5,48 e o RMSE aumentando de 5,65 para 6,97 entre o treinamento e o teste. Diante desse cenário, conclui-se que o modelo adotado demonstra boa capacidade preditiva para as regiões Nordeste, Sudeste e Centro-Oeste, enquanto necessita de aprimoramentos específicos para as regiões Norte e Sul, visando aumentar sua robustez e precisão.

A Figura 12 apresenta as previsões mensais do IVV, na escala original, para cada uma das cinco grandes regiões brasileiras ao longo de 2022, acompanhadas de seus intervalos de confiança. Em todas as regiões, a banda azul (intervalo de confiança) permanece relativamente estreita, os gráficos sugerem que o modelo consegue capturar, com razoável consistência, tanto a tendência ascendente de vendas em 2022 quanto as particularidades de cada região. O fato de os intervalos de confiança permanecerem relativamente estreitos na maior parte do período sugere boa precisão preditiva, embora diferenças no comportamento e nível do IVV em cada região apontem para a relevância de fatores locais e sazonais no desempenho do varejo brasileiro. A análise dos intervalos de confiança também aponta o Nordeste como a região para a qual o modelo é mais eficiente na predição do comportamento do IVV.

Figura 12 – Previsões do IVV com intervalos de Confiança para 2022 por Região



4. Conclusões e Considerações Finais

Os resultados apresentados ao longo deste estudo evidenciam a relevância de diversos indicadores econômicos na explicação do comportamento do varejo brasileiro, especialmente quando analisados de forma regionalizada. A modelagem realizada, apoiada em sucessivos ajustes, permitiu identificar os principais determinantes do Índice de Volume de Vendas (IVV) em cada região, ressaltando a importância de variáveis como o Índice de Atividade Econômica Regional (IBCR), a taxa de juros, o desemprego, além de efeitos sazonais e fatores exógenos específicos (como aqueles relacionados ao período da pandemia).

A evolução dos modelos, desde um ajuste inicial com violações de pressupostos, até a especificação final, demonstrou a importância de um processo iterativo na construção de modelos econométricos robustos. O ajuste final que demonstrou a melhor performance considerou o logaritmo do índice de volume de vendas (IVV) como variável resposta e os parâmetros foram estimados pelo método de mínimos quadrados ponderados, tendo atendido aos pressupostos fundamentais de um modelo de regressão linear. Especificamente, os resíduos do modelo apresentam distribuição próxima à normalidade, homocedasticidade e ausência de autocorrelação. Esses resultados garantem a validade das inferências estatísticas, reforçando a robustez do modelo desenvolvido. Houve significância de diversos fatores, dentre os quais as variáveis indicadoras de região, trimestre e período pandêmico de COVID-19, além de índices como taxas de desemprego e de juros e defasagem do IVV no mês anterior.

A análise preditiva confirmou a boa capacidade de previsão do modelo final, tanto no conjunto de treino quanto no conjunto de teste (ano de 2022). Os valores de MAE e RMSE, avaliados na escala original do IVV, revelaram erros médios relativamente baixos em relação ao nível de vendas típico de cada região. De maneira particular, a região Nordeste apresentou os menores valores de erro, sugerindo que a especificação adotada foi especialmente eficiente nessa região. Em contrapartida, as regiões Norte e Sul exibiram maiores desvios dos valores preditos em relação aos valores observados, apontando a existência de fatores regionais específicos ainda não totalmente capturados ou a maior variabilidade econômica peculiar a esse mercado.

No âmbito prático, os achados deste trabalho reforçam a utilidade de modelos de regressão linear múltipla para subsidiar decisões estratégicas no setor varejista. Empresas e formuladores de políticas públicas podem se beneficiar ao compreender como as variações de indicadores como juros e desemprego afetam as vendas em cada região, permitindo a adoção de medidas mais focalizadas. Adicionalmente, a incorporação de variáveis sazonais e de eventos atípicos (como a pandemia) comprova a importância de capturar efeitos conjunturais que impactam o varejo de forma heterogênea pelo território nacional.

Como limitação, ressalta-se que a complexidade dos fatores que influenciam o desempenho do varejo regional não pode ser plenamente capturada por um único modelo linear. Aspectos culturais, institucionais e políticos, bem como particularidades setoriais, podem exercer influência significativa sobre os padrões de consumo. Por essa razão, recomenda-se que pesquisas futuras explorem o desenvolvimento de modelos específicos para cada região, o que permitiria incorporar com maior precisão as características econômicas e sociais locais.

Além disso, sugere-se a inclusão de variáveis políticas e institucionais, especialmente considerando os ciclos eleitorais e mudanças administrativas ocorridas no período de 2012 a 2022, as quais podem ter afetado as expectativas econômicas, o crédito, o consumo e, por consequência, o desempenho do varejo. Um exemplo notório foi o processo de impeachment da presidente Dilma Rousseff em 2016, que gerou elevada instabilidade política e econômica, contribuindo para a retração do consumo e o agravamento da crise naquele período, um movimento visível no comportamento do IVV, Figura 1.

Em síntese, este estudo corrobora a ideia de que as disparidades regionais e as especificidades macroeconômicas exercem papel determinante no comportamento do varejo. Ao identificar os principais fatores associados ao desempenho das vendas e ao desenvolver um modelo estatístico capaz de prever o IVV com boa capacidade preditiva, este trabalho contribui para a literatura do tema e oferece subsídios práticos para a formulação de estratégias comerciais e políticas públicas voltadas ao desenvolvimento equilibrado do varejo brasileiro.

Referências

CARVALHO, A. X.; SILVA, J. R.; LIMA, L. R. Influência dos indicadores macroeconômicos sobre o consumo das famílias brasileiras. *Revista Brasileira de Economia*, v. 74, n. 3, p. 276–294, 2020.

CARVALHO, C.; FERNANDES, F.; LIMA, A. Indicadores macroeconômicos e desempenho do setor varejista brasileiro: uma análise econométrica regionalizada. *Revista de Economia Aplicada*, v. 24, n. 2, p. 58–80, 2020.

CORDEIRO, G. M.; DEMÉTRIO, C. G. B.; MORAL, R. A. *Modelos lineares generalizados e aplicações*. São Paulo: Blucher, 2024.

FIGUEIREDO, K.; ARKADER, R.; LAVALLE, C.; HIJJAR, M. F. Diferenças regionais na avaliação do serviço logístico: uma pesquisa no setor de supermercados em cinco capitais brasileiras. *Revista de Administração Mackenzie*, v. 3, n. 2, p. 117–136, 2002.

FOX, J.; WEISBERG, S. *An R Companion to Applied Regression*. 3. ed. Thousand Oaks, CA: Sage, 2019.

FOX, J.; WEISBERG, S. car: Companion to Applied Regression. R package version 3.1-2. Disponível em: <https://CRAN.R-project.org/package=car>.

KOTLER, P.; KELLER, K. L. *Administração de marketing*. 14. ed. São Paulo: Pearson Education do Brasil, 2012.

LEVY, M.; WEITZ, B. A. *Administração de varejo*. 9. ed. Porto Alegre: AMGH Editora, 2018.

MENDONÇA, M. J. C.; SACHSIDA, A. *Impacto dos indicadores econômicos no consumo das famílias brasileiras*. Brasília: Instituto de Pesquisa Econômica Aplicada – IPEA, 2019.

MINGOTI, S. A. *Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada*. Belo Horizonte: UFMG, 2005.

MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. *Introduction to linear regression analysis*. 5. ed. Hoboken, NJ: Wiley, 2015.

MONTGOMERY, D. C.; RUNGER, G. C. *Estatística aplicada e probabilidade para engenheiros*. 5. ed. Rio de Janeiro: LTC, 2012.

MORETTIN, P. A.; SINGER, J. M. *Estatística e ciência de dados*. Rio de Janeiro: LTC, 2022.

O PAPEL DO VAREJO NA ECONOMIA BRASILEIRA. São Paulo: Sociedade Brasileira de Varejo e Consumo, 2022. Disponível em: <https://sbvc.com.br/o-papel-do-varejo-naeconomia-brasileira-atualizacao-2022-sbvc/>. Acesso em: 23 abr. 2025.

R CORE TEAM. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing, 2024. Disponível em: <https://www.R-project.org/>. Acesso em: 23 abr. 2025.

ROCHA, T. V.; SILVA, M. L. Varejo regional e comportamento do consumidor: uma análise exploratória das diferenças culturais e socioeconômicas. *Revista Brasileira de Gestão de Negócios*, v. 20, n. 3, p. 363–385, 2018.

SILVA, J. A.; SARTI, F. M.; LAPLANE, M. F. Regionalização e varejo no Brasil: desafios e oportunidades. *Revista de Economia e Sociologia Rural*, v. 59, n. 2, p. e217638, 2021.

WILKS, D. S. *Statistical methods in the atmospheric sciences*. 3. ed. Oxford: Academic Press, 2011.

ANEXO I – Função de Autocorrelação (ACF) versus Função de Autocorrelação Parcial (PACF) da série temporal do IVV para cada região do Brasil

