



Análise dos resultados do Índice Brasileiro de Conectividade (IBC) 2024

PRUV/SPR
Julho/2025



Sumário

Sumário

1.	Introdução	4
2.	Lista de variáveis, pesos e fontes de dados	6
3.	Resultados do IBC 2024	8
3.1.	Análise de correlação do IBC com demais indicadores socioeconômicos	8
3.2.	Análise dos resultados do IBC.	10
3.3.	Mapas de calor das regiões do país pelo IBC 2024.	14
4.	Análises de agrupamentos.....	17
4.1.	Formação de quadrantes.	17
4.2.	Análise de agrupamento – preparação para formação de cluster.	21
4.2.1.	Definição de análise de cluster.	21
4.2.2.	Teste para avaliação do desempenho do agrupamento – teste de Shillouette.....	23
4.2.3.	Método para escolha da quantidade de clusters (k) – método do “cotovelo”	24
4.2.4.	Determinação das variáveis.	25
4.2.5.	Determinação das variáveis – redução de dimensionalidade.....	26
4.2.6.	Análise de agrupamento – criação dos aglomerados	30
5.	Modelo econométrico – SAR (Spatial Autoregressive Model).....	36
5.1.	Descrição do modelo SAR (Spatial Autoregressive Model).....	37
5.2.	Adoção da variável dummy para os demais clusters	38
5.3.	Teste I de Moran	38
5.4.	Grau de contiguidade dos municípios e aplicação do teste de Moran.	39
5.4.1.	Critério de contiguidade estabelecido.	39
5.4.2.	Aplicação do teste de Moran.....	41
5.5.	Aplicação do modelo SAR com clusters, efeitos diretos e indiretos	42
5.5.1.	Resultados obtidos na aplicação do modelo SAR.	43
5.5.2.	Estimativa dos efeitos diretos e indiretos e grau de ajustamento	44
5.6.	Testes de robustez do modelo SAR	45

5.6.1.	<i>Significância dos parâmetros (teste z)</i>	45
5.6.2.	<i>Multicolinearidade (VIF – Variance Inflation Factor)</i>	46
5.6.3.	<i>Autocorrelação espacial dos resíduos (Moran's I)</i>	46
5.6.4.	<i>Heterocedasticidade (Breusch-Pagan)</i>	46
5.6.5.	<i>Avaliação geral dos testes e robustez</i>	47
6.	Conclusões finais	48
7.	Referências bibliográficas	51
8.	Anexos	53

1. Introdução

Após extensa elaboração metodológica e busca de parâmetros nacionais e internacionais a versão do Índice Brasileiro de Conectividade (IBC) 2024 representa uma evolução metodológica na mensuração da infraestrutura de conectividade no Brasil. Algumas modificações foram feitas em relação à versão anterior, com a incorporação de novas variáveis, como a conectividade rural, e o aprimoramento dos critérios de normalização e ponderação, o índice proporciona uma visão mais atualizada e sensível às desigualdades regionais e tecnológicas do país. O objetivo do IBC é oferecer uma ferramenta objetiva e comparável ao longo do tempo para auxiliar na formulação e avaliação de políticas públicas no setor de telecomunicações.

É neste sentido que a Gerência de Universalização e Ampliação de Acessos da Superintendência de Planejamento e Regulamentação apresenta uma análise dos principais resultados obtidos pelo IBC em 2024. Destaca-se a manutenção de forte correlação entre o IBC e o Índice de Desenvolvimento Humano Municipal (IDHM), bem como a relação inversa com o Índice de Vulnerabilidade Social (IVS). A distribuição regional revelou disparidades persistentes, com Sul e Sudeste apresentando maior conectividade, enquanto o Norte e o interior do Nordeste ainda enfrentam desafios significativos. Além disso, a análise de um elemento específico do Indicador de Conectividade Rural (ICR) feito pela parceria ConectarAgro/UFV confirmou a importância das áreas rurais no diagnóstico da conectividade no nível dos municípios.

Complementando a análise descritiva, o estudo elaborou técnicas de agrupamento para segmentar os municípios em clusters de conectividade. Utilizando métodos como análise de quadrantes e K-means foram identificados perfis distintos de municípios com base em características como presença de fibra ótica e escore do IBC. Esses agrupamentos revelaram padrões importantes que podem subsidiar políticas mais focalizadas e eficazes, especialmente no que diz respeito à priorização de investimentos públicos em localidades com baixa conectividade.

Por fim, o relatório apresenta um modelo econométrico espacial do tipo SAR (*Spatial Autoregressive Model*), que incorpora o efeito de vizinhança entre os municípios. Com um $R^2 = 0,8261$, o modelo mostrou um grau de ajuste bem significativo, ou seja, o modelo explica 82,6%

da variabilidade do IBC. Com um coeficiente espacial ($\rho = 0,4318$) evidenciando uma autocorrelação espacial positiva significativa, ou seja, o IBC de um município é positivamente influenciado pelo IBC médio dos seus vizinhos imediatos. Esse valor sugere que aproximadamente 43% da variação local pode ser explicada pela vizinhança, evidenciando efeitos de difusão regional. A análise econométrica reforça o papel da conectividade como um bem público com externalidades positivas, e oferece subsídios técnicos para orientar ações de planejamento regional e otimização de recursos na expansão da infraestrutura digital e execução bem como avaliação de políticas públicas.

2. Lista de variáveis, pesos e fontes de dados

Para iniciar o presente trabalho, apresentamos um quadro resumo das variáveis componentes para nova versão do IBC em comparação a versão anterior. Neste caso, são detalhadas a adição da nova variável e a nova ponderação feita, de forma a atualizar o indicador. A partir da versão 2024, a metodologia de cálculo de índice foi atualizada para refletir com mais precisão a conectividade dos municípios brasileiros e acompanhar a constante evolução tecnológica do setor de telecomunicações. As principais mudanças nas variáveis estão descritas abaixo:

- Densidade de acessos móveis: considera apenas tecnologia 4G ou superior e acessos do tipo “padrão”, desconsiderando acessos M2M e ponto de serviço;
- Densidade de acessos de banda larga fixa: inclui apenas acessos classificados como “tipo Internet” e com velocidade contratada de 100 Mbps ou superior;
- Presença de *backhaul* fibra ótica: nova categorização da variável dividindo-se em 3 categorias;
- Inserção de uma variável para medir a conectividade rural: avalia a cobertura móvel em áreas passíveis de uso agrícola. Essa variável faz parte do Indicador da Conectividade Rural (ICR), desenvolvido em associação pela ConectarAGRO com a Universidade Federal de Viçosa (UFV), a partir de dados do Ministério da Agricultura e Anatel;
- Novo critério de normalização que utiliza limites inferiores e superiores teóricos pré-estabelecidos para viabilizar a comparação intertemporal do IBC a partir de 2024;
- Nova ponderação dos componentes do índice: definida a partir de um questionário eletrônico aplicado aos servidores da Anatel e especialistas de diversas entidades do setor.

Tabela 1 - Alteração da ponderação e variáveis das versões do IBC

Listagem e respectivos pesos das variáveis componentes das versões do IBC

Nome da variável	IBC ate 2023		IBC 2024	
	Fonte de dados	Peso	Fonte de dados	Peso
Densidade de Acessos Móveis	ANATEL	19%	ANATEL	15%
Densidade de Acessos de Banda Larga Fixa	ANATEL	19%	ANATEL	14%
Adensamento de ERBs	ANATEL	13%	ANATEL	13%
Presença de Backhaul de Fibra	ANATEL	20%	ANATEL	16%
Índice de Competitividade (HHI) do SMP	ANATEL	6%	ANATEL	7%
Índice de Competitividade (HHI) do SCM	ANATEL	6%	ANATEL	7%
População com coberta com sinal de telefonia móvel	ANATEL	17%	ANATEL	15%
Percentual de cobertura móvel em áreas passíveis de uso agrícola	-----	-----	UFV/ConectarAgro	13%

Fonte: SPR/ANATEL

Densidade de Acessos Móveis e Densidade de Acessos de Banda Larga Fixa estão no Painel de Acessos da ANATEL, com dados de população baseados no IBGE. Os índices de Competitividade são calculados baseados pela participação de mercado calculada pelo número de acessos da empresa em relação aos acessos totais. Link: <https://informacoes.anatel.gov.br/paineis/acessos>

População com cobertura com sinal de telefonia móvel, bem como a quantidade de empresas de Backhaul de fibra, estão no painel de infraestrutura; a cobertura está na sessão cobertura móvel e a presença de fibra está na sessão rede de transporte. Link: <https://informacoes.anatel.gov.br/paineis/infraestrutura>.

Adensamento de ERBs está em painel de Outorga e licenciamento. Cabe ressaltar que neste painel de estações não é registrado o histórico, portanto a extração de dados de ERBs é feita quando se calcula o IBC, e pode alterar ao longo do tempo. Link: <https://informacoes.anatel.gov.br/paineis/outorga-e-licenciamento/estacoes-do-smp>.

O percentual de cobertura móvel em áreas de uso agrícola se encontra no painel da ConectarAgro. Painel do Indicador de Conectividade Rural (ICR). Link: <https://www.conectaragro.com.br/indicadordeconectividaderural/>.

Para maiores detalhes consultar o Relatório Metodológico da nova versão do IBC, disponível na página do “Meu Município” do painel de dados da ANATEL¹.

1

https://sei.anatel.gov.br/sei/modulos/pesquisa/md_pesq_documento_consulta_externa.php?HWH32bONvibUcMC3mewfUpIX7e-9fyZZC4iEjI2QHwXAoLCOrVZwNzRf5vR3YcCMWZN4eCgQDLmVzIOFPcg7Rh3LXmDINOEyKQjxWvmQHMF7ng9whS0n_BgQ9jRf8MdF



3. Resultados do IBC 2024

3.1. Análise de correlação do IBC com demais indicadores socioeconômicos

Um dos primeiros exercícios estatísticos conduzidos foi a comparação do IBC com distintos índices e indicadores de estágios de infraestrutura, de renda e de desenvolvimento dos municípios. Nesse caso, usou-se uma análise de correlação linear entre o IBC e demais índices. Isso permite avaliar se existe relação entre duas variáveis. Essa é uma medida muito usada quando se estuda variáveis quantitativas. Em FÁVERO e BELFIORE (2017), é explicado o conceito de correlação de Pearson, critério adotado no presente trabalho. Segundo eles:

“O coeficiente de correlação de Pearson é uma medida que varia entre -1 e 1. Por meio do sinal, é possível verificar o tipo de relação linear entre as duas variáveis analisadas (direção em que a variável Y aumenta ou diminui em função da variação de X); quanto mais próximo dos valores extremos, mais forte é a correlação entre elas”. (FÁVERO e BELFIORE, 2017)

Desta forma:

- Se o coeficiente for positivo, existe uma relação diretamente proporcional entre as variáveis; se $\rho=1$, tem-se uma correlação linear positiva perfeita.
- Se ρ for negativo, existe uma relação inversamente proporcional entre as variáveis; se $\rho=-1$, tem-se uma correlação linear negativa perfeita.
- Se ρ for nulo, não existe correlação entre as variáveis.

Cabe ressaltar, ainda, que não existe uma única padronização universal para avaliar o grau de correlação. Nesses casos, um critério bastante comum para interpretar o valor do coeficiente é descrito na tabela seguinte:

Tabela 2 – Grau de correlação de Pearson

Níveis de correlação de Pearson	
0,00 a 0,19	Correlação muito fraca
0,20 a 0,39	Correlação fraca
0,40 a 0,59	Correlação moderada
0,60 a 0,79	Correlação forte
0,80 a 1,00	Correlação muito forte

Essa regra é válida tanto para correlação positiva quanto negativa: o sinal somente indica a direção da relação e o valor absoluto indica a força. Desta feita, realizamos uma análise de correlação entre o IBC 2024 e alguns outros indicadores. São eles:

- O Índice de Desenvolvimento Humano Municipal (IDHM)² é uma medida composta de indicadores de três dimensões do desenvolvimento humano: longevidade, educação e renda. O índice varia de 0 a 1. Quanto mais próximo de 1, maior o desenvolvimento humano.
- O Índice de Vulnerabilidade Social (IVS)³ é um indicador desenvolvido pelo Instituto de Pesquisa Econômica Aplicada (IPEA) para medir o grau de vulnerabilidade social da população em diferentes territórios do Brasil; infraestrutura urbana, capital humano, renda e Trabalho.
- Indicador da Conectividade Rural (ICR)⁴, desenvolvido em associação da ConectarAgro em parceria com a Universidade Federal de Viçosa (UFV) a partir de dados do Ministério da Agricultura e da Anatel.
- O PIB per capita municipal⁵ é um indicador econômico que representa o Produto Interno Bruto (PIB) de um município dividido pela sua população estimada em um determinado ano. Ele mede a riqueza média por habitante, servindo como referência para avaliar o nível de desenvolvimento econômico local. No Brasil, ele é medido pelo IBGE.

Após estas descrições, apresenta-se a tabela com os resultados da correlação do IBC contra os demais indicadores.

Tabela 3 – Matriz de correlação IBC e demais indicadores

Matriz de correlação - IBC (2024) e demais indicadores sociais e econômicos										
Ano	Matriz de correlação (IBC e demais indicadores)	IBC 2024	IDHM	idhm_renda	IVS	ivs infra_urbana	ivs cap_humano	ivs renda_e_trabalho	ICR	PIB per capita 2021
2024	IBC 2024	1								
2010	IDHM	0,72	1							
2010	idhm_renda	0,68	0,95	1						
2010	IVS	-0,63	-0,88	-0,88	1					
2010	ivs_infraestrutura_urbana	-0,37	-0,55	-0,57	0,83	1				
2010	ivs_capital_humano	-0,64	-0,92	-0,87	0,91	0,59	1			
2010	ivs_renda_e_trabalho	-0,68	-0,90	-0,90	0,89	0,52	0,84	1		
2024	ICR	0,68	0,42	0,37	-0,43	-0,29	-0,45	-0,40	1	
2021	PIB per capita (preços correntes) 2021	0,25	0,36	0,39	-0,33	-0,22	-0,32	-0,35	0,06	1

Fontes: IDHM (PNUD/ONU), IVS(IPEA), ICR(ConectarAgro/UFV), PIB per capita (IBGE), PRUV/ANATEL

² <https://www.undp.org/pt/brazil/desenvolvimento-humano/painel-idhm>

³ <https://ivs.ipea.gov.br/#/>

⁴ <https://www.conectaragro.com.br/indicadordeconectividaderural/>

⁵ <https://www.ibge.gov.br/busca.html?searchword=lista+de+munic%C3%ADpios+do+Brasil+por+renda+per+capita>



Foram destacados os níveis de correlação mais significativos. Neste caso, evidencia-se muita correlação do IBC 2024 com o IDHM (0,72), sendo correlação positiva. Isso indica, como já antes estava mensurado na versão anterior do IBC, uma manutenção da relação positiva entre o IBC e o IDHM. O IDHM de renda também teve uma correlação significativa.

A relação do IVS com o IBC mostrou-se negativa, o que confirma a tese de que quanto menor a vulnerabilidade social, maior a conectividade (correlação de -0,63). Dois subíndices do IVS também apresentaram uma correlação negativa forte: -0,64 para o IVS em capital humano e -0,68 em IVS para renda e trabalho.

A correlação entre o ICR (Indicador de Conectividade Rural) também se mostrou alta (0,68). Isso é confirmado pelo fato de que muito dos indicadores do ICR estão presentes no IBC.

Cabe ressaltar também a correlação de dois indicadores com o IBC: a primeira é a correlação com o IVS de infraestrutura urbana e a segunda a relação do IBC e o PIB per capita. Apesar de intuitivamente essas correlações seriam em tese mais fortes, uma interpretação dos resultados pode evidenciar uma boa atuação regulatória. Isso porque as correlações mais fracas mostram que o IBC está sendo equilibrado na avaliação entre municípios grandes e pequenos e entre municípios ricos e pobres. No caso, o IVS de infraestrutura evidencia-se que a conectividade nos municípios contribui para a redução da vulnerabilidade social em infraestrutura, porém, outros equipamentos e serviços de infraestrutura urbana são considerados como, por exemplo; iluminação, asfaltamento, limpeza urbana, rede de energia elétrica, tratamento de água potável e esgoto etc. Isso pode indicar um ponto forte do IBC bem como um indicador de missão institucional executada da Anatel.

3.2. Análise dos resultados do IBC.

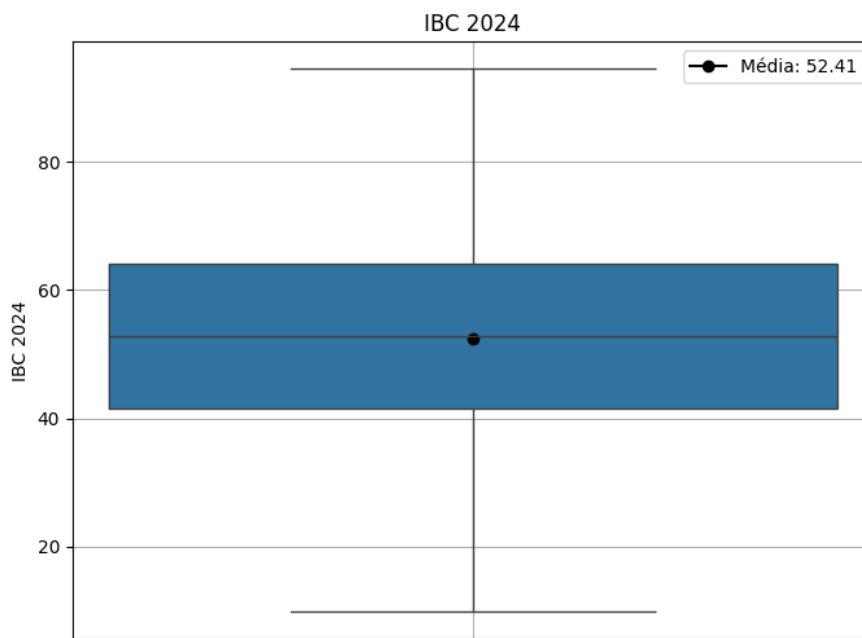
Após feito o cálculo do IBC para 2024, obtiveram-se os dados em estatísticas descritivas. Observa-se a média e a mediana muito próximas, sendo a média em 52,4 e a mediana em 52,9. Ressalta-se ainda que foram analisados os dados em 5.565 municípios, visto que o mapeamento utilizado foi do Censo de 2010 e ainda os 5 municípios novos (após 2010) não haviam sido criados.

Tabela 4 – Estatísticas Descritivas do IBC 2024

Estatísticas Descritivas IBC 2024	
Média	52,405
Erro padrão	0,200
Mediana	52,896
Desvio padrão	14,909
Variância da amostra	222,287
Curtose	-0,635
Assimetria	-0,146
Intervalo	84,755
Mínimo	9,815
Máximo	94,569
Soma	291635
Contagem	5565

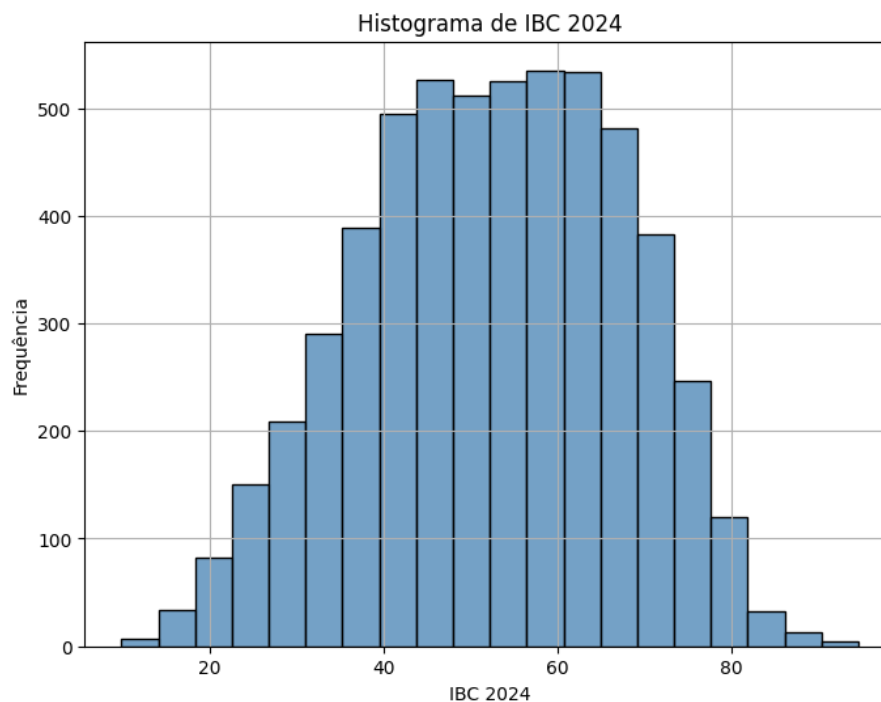
Somado à tabela de estatísticas descritivas, apresenta-se o gráfico *box-plot* do IBC de 2024. Esse gráfico mostra a parte de concentração dos dados para os 5.565 municípios.

Figura 1 – Gráfico box-plot do IBC 2024



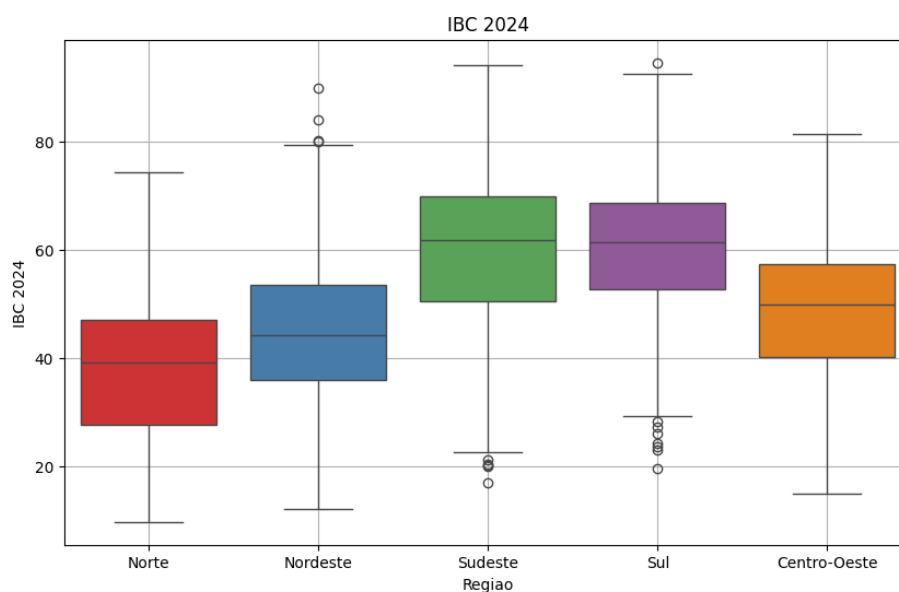
Em complemento ao *box-plot*, encontra-se o histograma do valor do IBC de 2024. Nesse caso, observa-se uma distribuição pouco assimétrica, visto que a média e a mediana estão muito aproximadas.

Figura 2 – Histograma do IBC 2024



Em seguida, foi segmentado o gráfico *box-plot* pelas cinco regiões do país, já apresentando um início de aprofundamento da análise. Observa-se em destaque os *box-plot* das regiões Sul e Sudeste acima da média, e os gráficos das regiões Norte e Nordeste abaixo, em especial a região Norte.

Figura 3 – Box-- plot do IBC por regiões do Brasil – 2024



Além disso, apresentam-se as estatísticas descritivas segmentadas também por região. Os valores gerais descritivos corroboram os dados ilustrados no gráfico de *box-plot*.

Tabela 5 – Estatísticas Descritivas por Região – IBC 2024

	Região Centro-Oeste	Região Nordeste	Região Norte	Região Sudeste	Região Sul
Contagem (n)	466.00	1794.00	449.00	1668.00	1188.00
Média	48.59	44.83	38.56	59.71	60.32
Desvio padrão	12.15	12.43	12.95	13.25	11.76
Mínimo	14.91	12.16	9.81	16.97	19.67
1º Quartil (Q1)	40.15	35.97	27.77	50.50	52.72
Mediana (Q2)	49.88	44.25	39.29	61.92	61.38
3º Quartil (Q3)	57.31	53.46	47.08	69.81	68.70
Máximo	81.45	89.85	74.26	94.01	94.57
Moda	14.91	12.16	9.81	16.97	19.67
Variância	147.58	154.50	167.82	175.69	138.29
Coef. de Variação	0.25	0.28	0.34	0.22	0.19
Assimetria (Skewness)	-0.23	0.18	0.15	-0.49	-0.37
Curtose	-0.46	-0.31	-0.48	-0.33	0.01
Intervalo (Máx - Mín)	66.54	77.69	64.44	77.04	74.90

3.3. Mapas de calor das regiões do país pelo IBC 2024.

Seguem os mapas de calor das cinco regiões do país, evidenciando as diferenças ao nível municipal. No caso do mapa de calor, quanto mais escuro foi a coloração, maior o valor do IBC.

Figura 4 – Mapa de calor do IBC de 2024 – Região Sudeste

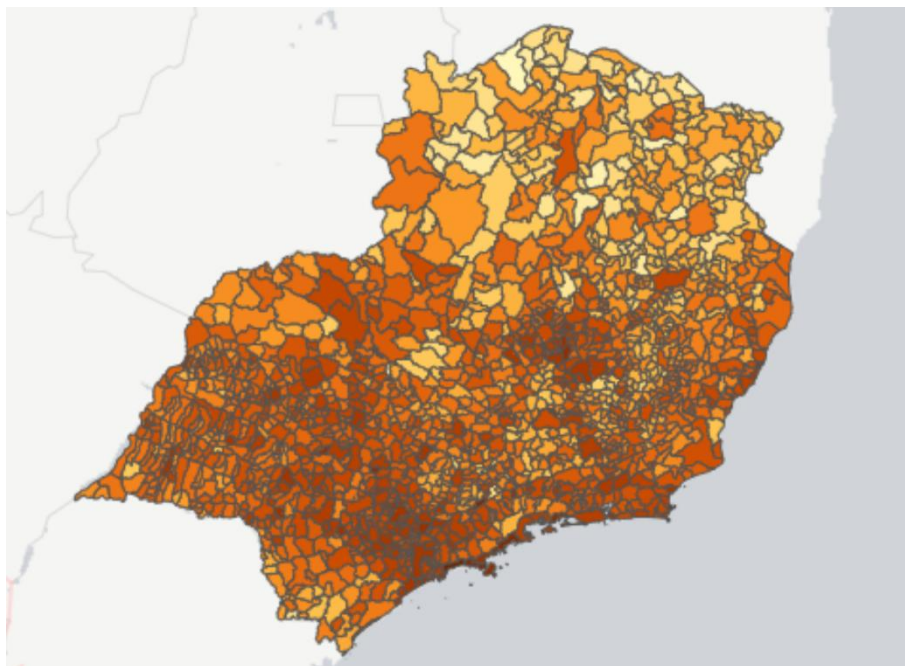


Figura 5 – Mapa de calor do IBC de 2024 – Região Sul

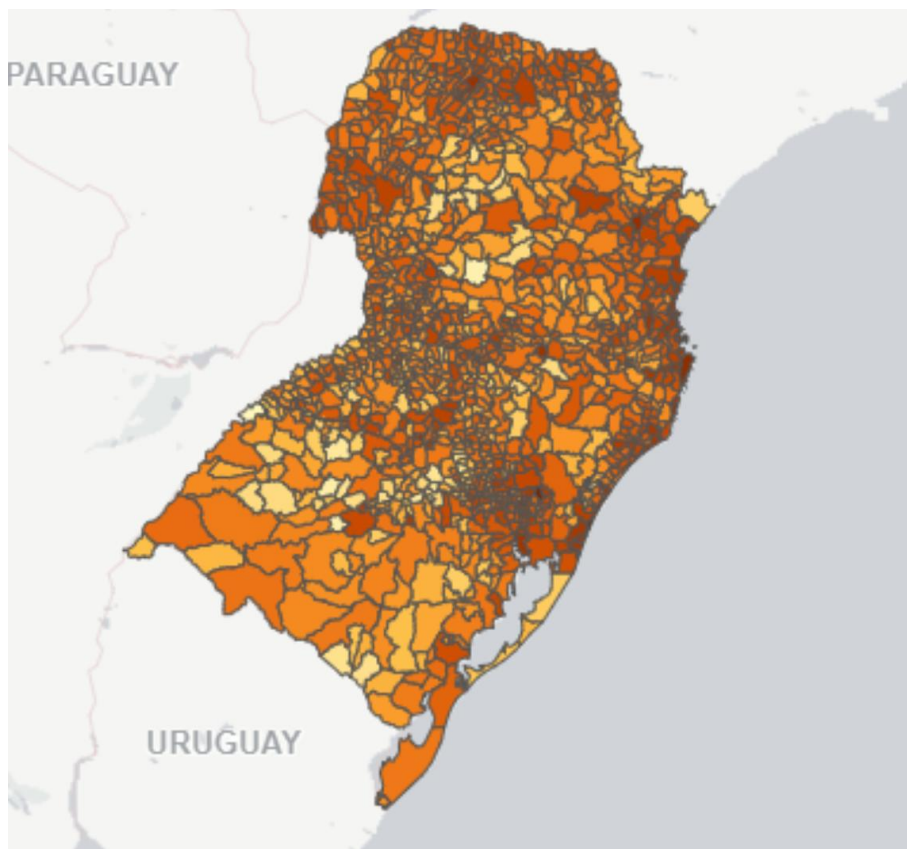


Figura 6 – Mapa de calor do IBC de 2024 – Região Centro-Oeste

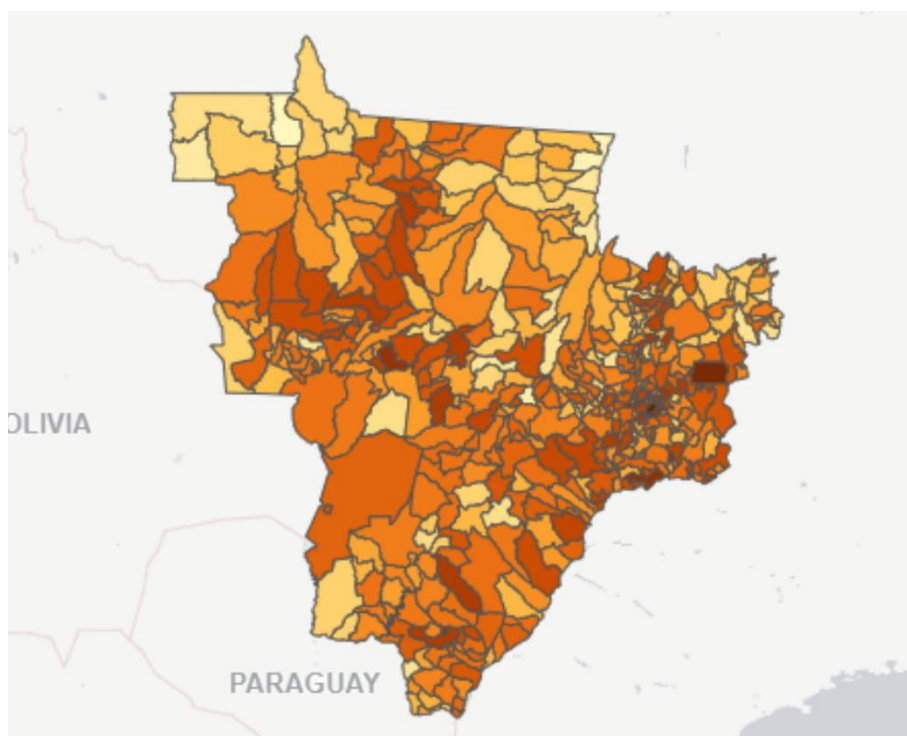


Figura 7 – Mapa de calor do IBC de 2024 – Região Nordeste

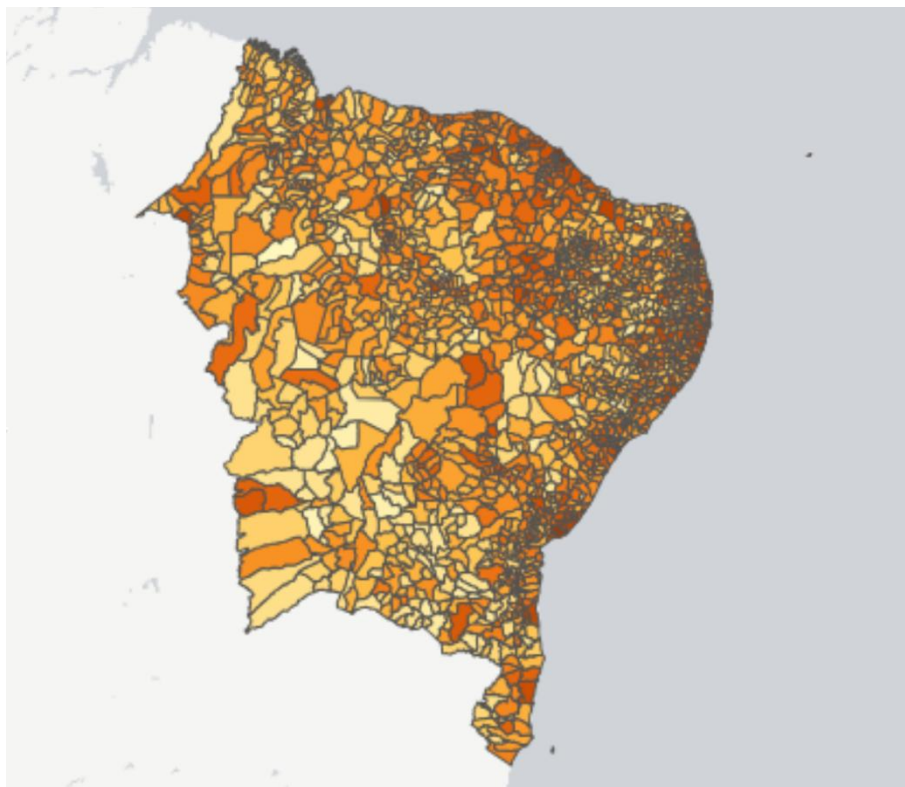
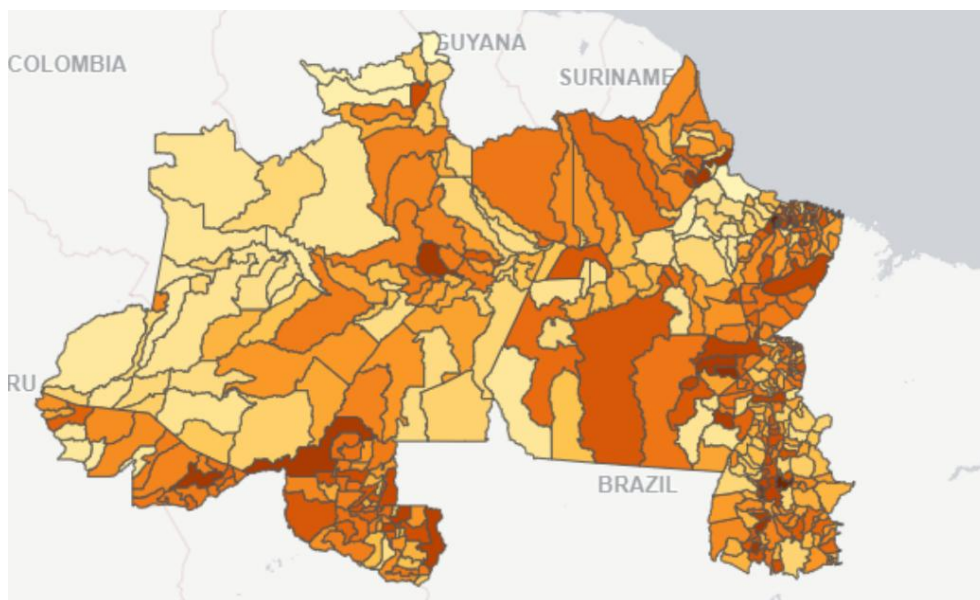


Figura 8 – Mapa de calor do IBC de 2024 – Região Norte



4. Análises de agrupamentos

4.1. Formação de quadrantes.

Após uma análise geral dos resultados, partimos para um procedimento mais detalhado, onde foi feita uma análise de agrupamento. Uma definição de análise de agrupamento pode ser apresentada em HAIR Jr⁶:

“A análise de agrupamentos é uma técnica analítica para desenvolver subgrupos significativos de indivíduos ou objetos. Especificamente, o objetivo é classificar uma amostra de entidades (indivíduos ou objetos) em um número menor de grupos mutuamente excludentes, com base nas similaridades entre as entidades”.

“A análise de agrupamentos geralmente envolve pelo menos três passos. O primeiro é a medida de alguma forma de similaridade ou associação entre as entidades para determinar quantos grupos realmente existem na amostra. O segundo passo é o real processo de agrupamento, onde entidades são particionadas em grupos (agrupamentos). O último passo é estabelecer o perfil das pessoas ou variáveis para determinar sua composição. Muitas vezes, isso é possível pela aplicação da análise discriminante aos grupos identificados pela técnica de agrupamento.”

A primeira abordagem que se tomou como base consistiu na análise por quadrantes utilizando duas variáveis: Índice Brasileiro de Conectividade (IBC) e o Índice de Desenvolvimento Humano Municipal (IDHM)⁷. A metodologia baseia-se na representação gráfica dos dados em um plano cartesiano bidimensional, categorizando por quadrantes. Ressalta-se que tais variáveis são as que obtiveram a maior correlação positiva entre si (0,72).

Pontos de corte de cada variável são utilizados para dividir o plano em quatro quadrantes, cada um representando uma combinação específica de níveis de conectividade e desenvolvimento humano. Nesse caso, como valor de referência para definir os quadrantes, o valor de medianas do IBC e do IDHM, respectivamente 52,89 e 0,665. Assim, o município será considerado com alto

⁶ HAIR JR., Joseph F. et al. *Análise multivariada de dados*. 6. ed. Porto Alegre: Bookman, 2009

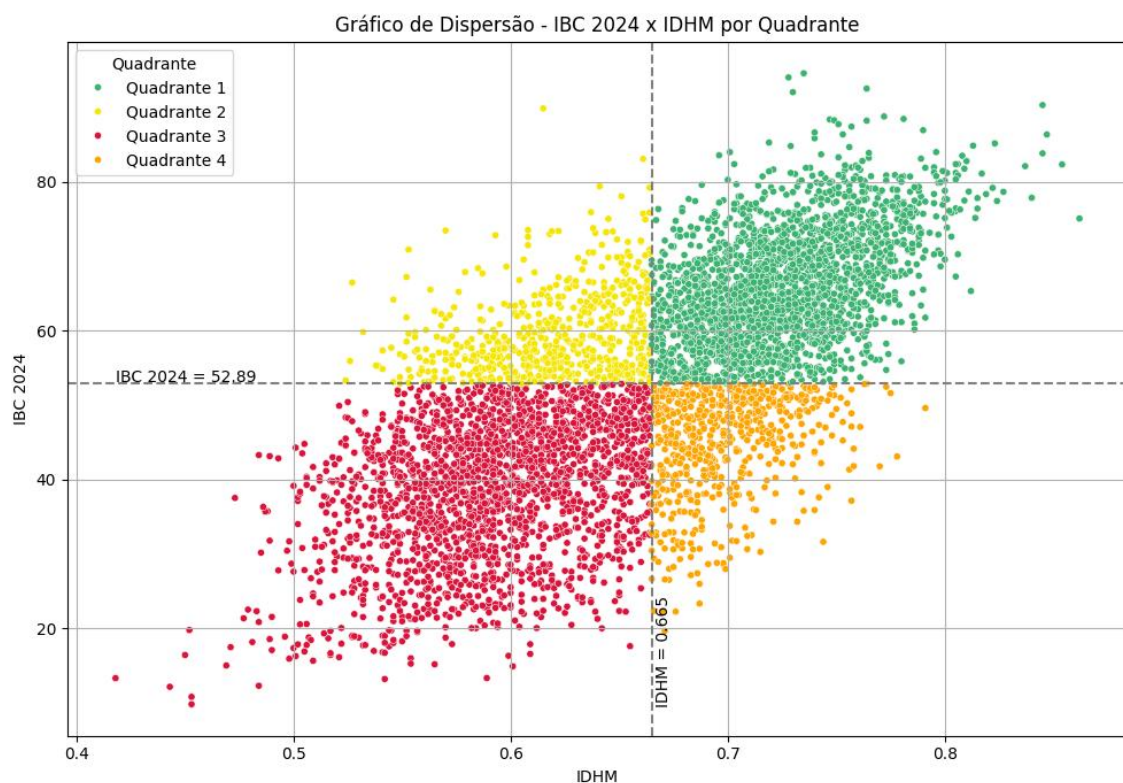
⁷ ALVES, ANA, *Estatística Aplicada: Análise de Dados (Volume I — Análises)*. Aprender Estatística Fácil, 2024. 196 p. 2nd ed.

IBC o que tiver um indicador de mais de 52,89 e baixo IBC aquele que estiver abaixo desta marca.

A classificação dos quadrantes se dá da seguinte maneira:

- Quadrante 1 (Alto IBC, Alto IDHM): Municípios com alto nível de conectividade e alto desenvolvimento humano.
- Quadrante 2 (Alto IBC, Baixo IDHM): Municípios com alta conectividade, mas baixo desenvolvimento humano.
- Quadrante 3 (Baixo IBC, Baixo IDHM): Municípios com baixos índices tanto de conectividade quanto de desenvolvimento.
- Quadrante 4 (Baixo IBC, Alto IDHM): Municípios com baixo nível de conectividade, mas alto desenvolvimento humano.

Figura 9 – Gráfico de dispersão dos quadrantes



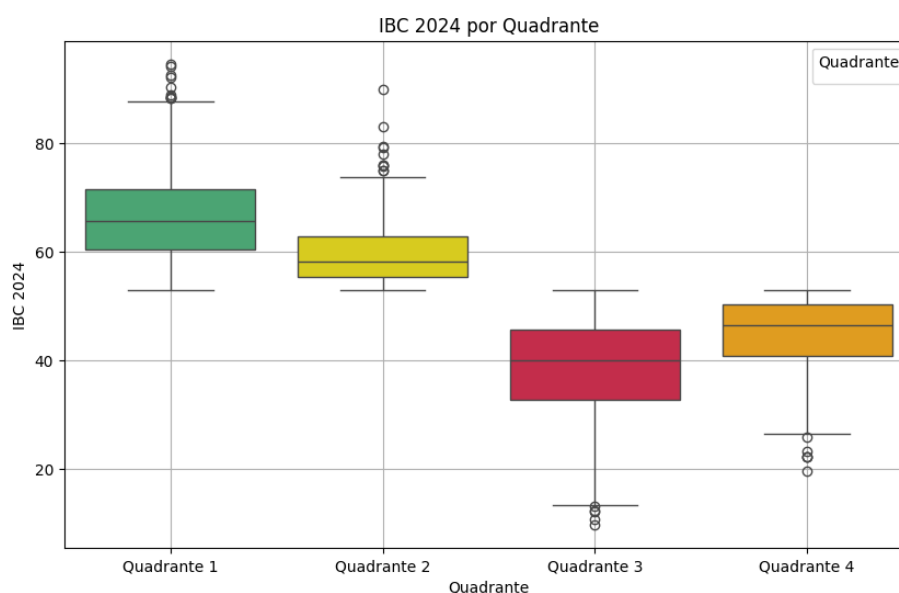
Deve-se observar que, devido à alta correlação entre as variáveis IBC e IDHM, os quadrantes 1 e 3 (alto e alto; baixo e baixo) foram os com mais municípios, visto que a nuvem de dispersão dos pontos é ascendente. Após a apresentação do gráfico de dispersão, adiciona-se as estatísticas descritivas com cortes dos quatro quadrantes.

Tabela 6 – Estatísticas Descritivas por Quadrante – IBC 2024

	Quadrante 1	Quadrante 2	Quadrante 3	Quadrante 4
Contagem (n)	2185.00	598.00	2166.00	616.00
Média	66.21	59.63	38.65	44.80
Desvio padrão	7.53	5.61	8.92	6.81
Mínimo	52.94	52.90	9.81	19.67
1º Quartil (Q1)	60.38	55.31	32.75	40.80
Mediana (Q2)	65.75	58.12	40.00	46.58
3º Quartil (Q3)	71.49	62.80	45.76	50.38
Máximo	94.57	89.85	52.85	52.87
Moda	52.94	52.90	9.81	19.67
Variância	56.76	31.51	79.49	46.35
Coef. de Variação	0.11	0.09	0.23	0.15
Assimetria (Skewness)	0.38	1.26	-0.54	-1.01
Curtose	-0.33	1.95	-0.43	0.43
Intervalo (Máx - Min)	41.63	36.95	43.04	33.20

A seguir, apresenta-se um gráfico de *box-plot* com a segmentação dos quatro quadrantes, já com coloração específica. Nota-se, por óbvio, destaque para os quadrantes 1 e 3.

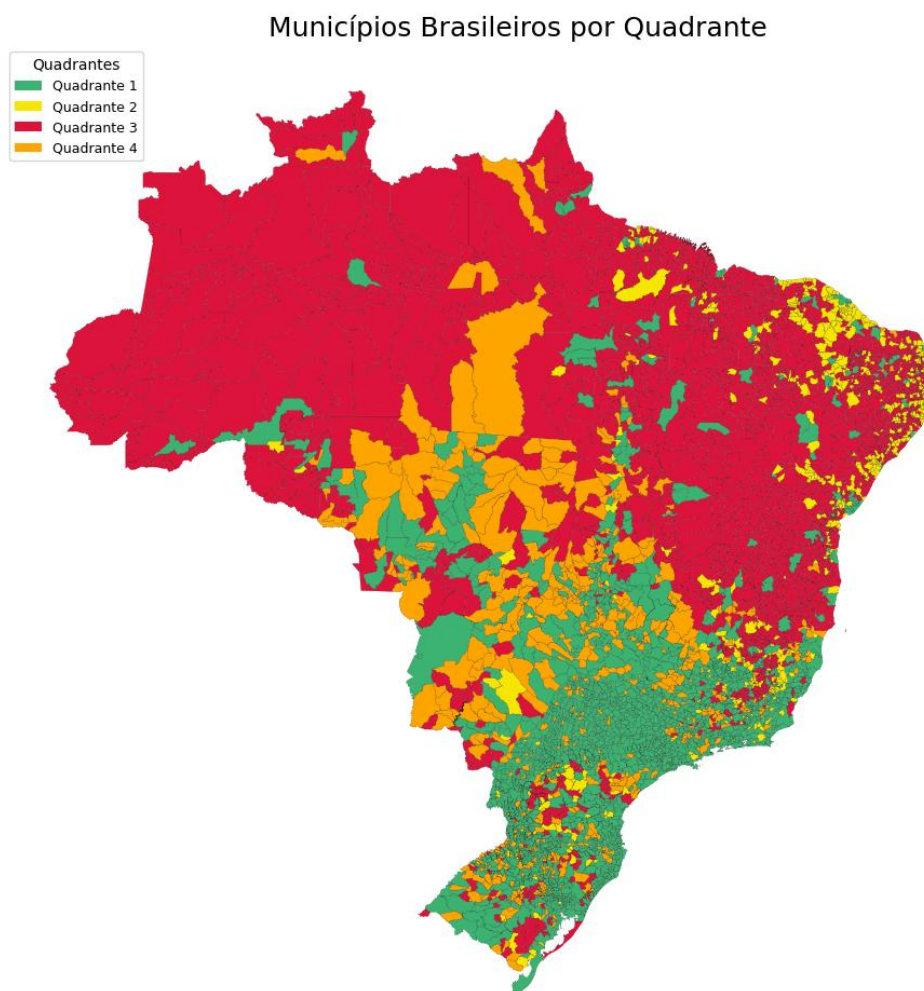
Figura 10 – Gráfico de box-plot dos quadrantes – IBC 2024



Finalizando a análise por quadrante, construiu-se um mapa do país, com recorte ao nível municipal, hachurando as áreas dos municípios com cores específicas segundo o gráfico de

dispersão. Isso já fornece uma noção da distribuição e do estágio de conectividade e desenvolvimento humano nos municípios brasileiros. Essa ilustração já evidencia as disparidades de estágio de conectividade com alguns componentes regionais, porém não se limitando a eles.

Figura 11 – Mapa dos municípios brasileiros por quadrante – IBC 2024



Observa-se claramente o Sul e Sudeste do país com muitos municípios em alto grau de conectividade, o litoral do nordeste se destacando regionalmente, bem como municípios do Centro-oeste. Os municípios de baixo estágio de conectividade no Norte e interior do Nordeste. Isso já fornece uma noção da distribuição e do estágio de conectividade e desenvolvimento humano nos municípios brasileiros.

4.2. Análise de agrupamento – preparação para formação de cluster.

4.2.1. Definição de análise de cluster.

Após a análise de agrupamento mediante quadrantes, foram utilizadas técnicas de análise de “aglomerado” ou análise de cluster. Essa técnica é bem descrita em FÁVERO e BELFIORE (2017), que destacam algumas definições importantes. Segue transcrição:

“Muitos são os procedimentos para ser elaborada uma análise de agrupamentos, visto que existem diferentes medidas de distância ou de semelhança para, respectivamente, variáveis métricas ou binárias. Além disso, definida a medida de distância ou de semelhança, o pesquisador ainda precisa determinar, entre diversas possibilidades, o método de aglomeração das observações, a partir de determinados critérios hierárquicos ou não hierárquicos⁸”. (FÁVERO e BELFIORE, 2017) (grifo nosso)

Ainda segundo os autores:

“Os esquemas de aglomeração podem ser classificados, basicamente, em dois tipos, conhecidos por hierárquicos e não hierárquicos. Enquanto os primeiros caracterizam-se por privilegiar uma estrutura hierárquica (passo a passo) para a formação dos agrupamentos, os esquemas não hierárquicos utilizam algoritmos para maximizar a homogeneidade dentro de cada agrupamento, sem que haja um processo hierárquico para tal”. (FÁVERO e BELFIORE, 2017)

Já MALHORTA (2006) diz que a análise de clusters é uma técnica usada para classificar objetos ou casos em grupos relativamente homogêneos chamados de clusters (ou conglomerados)⁹.

Em SILVA et al. (2025), é destacado que os objetos em cada cluster tendem a ser semelhantes entre si, mas diferentes de objetos em outros clusters. Neste tipo de análise não há nenhuma informação a priori sobre a composição do grupo ou cluster para qualquer um de seus objetos¹⁰.

⁸ FÁVERO, Luiz Paulo; BELFIORE, Patrícia. *Manual de análise de dados: estatística e modelagem multivariada com Excel®, SPSS® e Stata®*. 2. ed. Rio de Janeiro: Elsevier, 2017

⁹ MALHOTRA, Naresh K. *Pesquisa de Marketing: Uma Orientação Aplicada*. Ed. Bookman; 4ª edição; Porto Alegre; 2006, p.572.

¹⁰ SILVA, Elvis Magno da et al. Análise multivariada de Cluster: uma ferramenta estatística para pesquisa em ciências sociais aplicadas. *International Journal of Scientific Management and Tourism*, Curitiba, v. 10, n. 5, p. 1–25, 2024. DOI: 10.55905/ijsmtv10n5-002. Disponível em: <https://www.ijsmt.com/>. Acesso em: 7 maio 2025

Nesse sentido, o que inicialmente parecer trivial, ao se desejar agrupar observações em clusters (aglomerados ou conglomerados) internamente homogêneos, pode se tornar um tanto complexo, enquanto há uma multiplicidade de combinações entre diferentes medidas de distância ou de semelhança e métodos de aglomeração.

Por isso, destacamos a transcrição importante de FÁVERO e BELFIORE (2017): *“É de fundamental importância, portanto, que o pesquisador defina, com base na teoria subjacente e em seus objetivos de pesquisa, bem como em sua experiência e intuição, os critérios a partir dos quais as observações serão alocadas em cada um dos grupos”*.

Dessa forma, como descrito anteriormente, para fazer uma análise de cluster, então deve-se estabelecer os critérios para a metodologia de medida de semelhança. Inclusive, um esquema de aglomeração não hierárquico requer a estipulação, a priori, da quantidade de clusters a partir da qual serão definidos os centros de aglomeração e alocadas as observações. Nesses casos, nos esquemas de aglomeração não hierárquicos, entre os quais o mais popular é o procedimento *k-means*, ou k-médias, em que se referem a processos em que são definidos centros de aglomeração a partir dos quais são alocadas as observações pela proximidade a eles.

A partir da definição da análise de agrupamento, parte para decidir qual metodologia aplicar para fazer diferenciação e qual critério de similaridade adotar (medida de diferença). Neste caso, optou-se pelo método do *K-means* pelas razões descritas seguintes.

O *K-means* é um método de *clustering* não supervisionado que agrupa dados em k clusters com base na similaridade. O método é um algoritmo de clusterização muito comum e popular, usado para agrupar dados em clusters distintos. Um ponto importante é que o *K-means* solicita que os dados tenham padronização para serem realizados; o que é o caso do IBC, uma vez que os dados dos componentes do índice são padronizados na escala de 0 a 100.

Destaca-se também que o método funciona iterativamente, atribuindo cada ponto de dados ao cluster mais próximo de seu centroide e atualizando os centroides até que a atribuição dos pontos se estabilize. O objetivo então é minimizar a soma das distâncias internas (*intra-cluster*) entre os pontos e seus respectivos centroides.

Resumidamente, o processo pode ser descrito da seguinte maneira:

1. Define-se o número de clusters (k) a priori;
2. O algoritmo escolhe aleatoriamente k centroides iniciais;
3. Cada ponto do conjunto de dados é atribuído ao centroide mais próximo (formando cluster). O algoritmo calcula a distância de cada ponto no conjunto de dados até cada um dos k centroides. A métrica mais comum é a distância Euclidiana. Para cada ponto é atribuído ao cluster cujo centroide está mais próximo, logo esse processo cria k grupos temporários;
4. Os centroides são então recalculados como a média dos pontos em cada cluster;
5. Os passos 3 e 4 se repetem até que os centroides não mudem mais (ou até atingir um número máximo de iterações).

Uma vez então escolhido o método (*K-Means*), parte-se então para escolher quais as variáveis serão aplicadas para realizar a análise e a quantidade “ k ” de aglomerados a serem criados. Ambas as definições influenciam no desempenho da análise de cluster. Ambos os casos possuem métodos para ajudar nas definições, com a execução de teste estatístico apropriado.

4.2.2. Teste para avaliação do desempenho do agrupamento – teste de *Shillouette*

O teste de Silhouette (ou índice de *Silhouette*) é uma métrica usada para avaliar a qualidade de agrupamentos (clusters) em uma análise de cluster. Ele mede o quão bem cada objeto está inserido dentro do seu próprio cluster, comparando a proximidade do objeto com os pontos do mesmo cluster e com os pontos de outros clusters¹¹.

Para cada observação i , o índice de *Silhouette* $s(i)$ é calculado da seguinte forma:

- $a(i)$: distância média entre i e todos os outros pontos do mesmo cluster;
- $b(i)$: distância média entre i e todos os pontos do cluster mais próximo (ou seja, o segundo melhor cluster para i).

A fórmula do índice de *Silhouette* é:

¹¹ FÁVERO, Luiz Paulo; BELFIORE, Patrícia. *Manual de análise de dados: estatística e modelagem multivariada com Excel®, SPSS® e Stata®*. 2. ed. Rio de Janeiro: Elsevier, 2017

$$s(i) = \frac{b(i) - a(i)}{\max [a(i), b(i)]}$$

O índice de *Silhouette* varia de -1 a $+1$:

- Próximo de $+1$: o ponto está bem ajustado ao seu próprio cluster e mal ajustado aos clusters vizinhos — boa alocação;
- Próximo de 0 : o ponto está na fronteira entre dois clusters;
- Próximo de -1 : o ponto pode estar alocado ao cluster errado — má alocação.

O índice médio de *Silhouette* de todas as observações é frequentemente usado como medida global da qualidade do *clustering*:

- Valores acima de $0,5$: boa estrutura de cluster;
- Valores entre $0,25$ e $0,5$: estrutura razoável, mas pode haver sobreposição entre clusters;
- Valores abaixo de $0,25$: indicam que a estrutura de cluster pode não ser significativa.

Como descrito, este teste será executado como um dos parâmetros para julgar a qualidade ou desempenho do agrupamento realizado.

4.2.3. Método para escolha da quantidade de clusters (k) – método do “cotovelo”

Em sequência, também se apresenta um método para avaliar uma quantidade adequada de clusters. Nesse caso, um método popular é o chamado “método do cotovelo”. Tal método é uma técnica gráfica usada para determinar o número ideal de clusters (k) em uma análise de agrupamento, especialmente quando se utiliza algoritmos como o *K-Means*.

A ideia central é avaliar como a soma dos erros quadráticos *intra-cluster* (também conhecida como soma das distâncias *intra-grupo* ou Within-Cluster Sum of Squares – WCSS) diminui à medida que o número de clusters aumenta.

Para diferentes valores de k, calcula-se o WCSS. Esses valores são então plotados em um gráfico onde o eixo x representa o número de clusters e o eixo y representa o WCSS.

O gráfico tende a mostrar uma queda acentuada no WCSS à medida que k aumenta, mas a taxa de redução diminui após certo ponto. Esse é o comportamento da chamada curva do gráfico do “cotovelo” que corresponde ao valor de k a partir do qual os acréscimos no número de clusters não produzem reduções significativas no WCSS. Assim, esse ponto é interpretado empiricamente como o número ótimo de clusters, por representar um equilíbrio entre complexidade do modelo e qualidade da segmentação. Cabe ressaltar que tal interpretação tem um certo grau de subjetividade do pesquisador.

4.2.4. Determinação das variáveis.

Uma vez estabelecido o tipo de análise a ser feita, bem como o método de avaliação, partimos para determinar um primeiro conjunto de variáveis e escolher dentre elas quais as melhores para agrupar os municípios.

Cabe salientar que a definição das variáveis pode afetar a qualidade da análise de agrupamento, de tal forma que se definiu então um ponto de partida, e deste ponto, iniciamos rodadas de clusterização com respectivos testes de *Silhouette* para avaliação.

Esse procedimento permite encontrar a melhor forma de agrupamento, bem como realizar um procedimento chamado de redução de dimensionalidade; onde se busca ter o mesmo desempenho de clusterização com menos variáveis utilizadas.

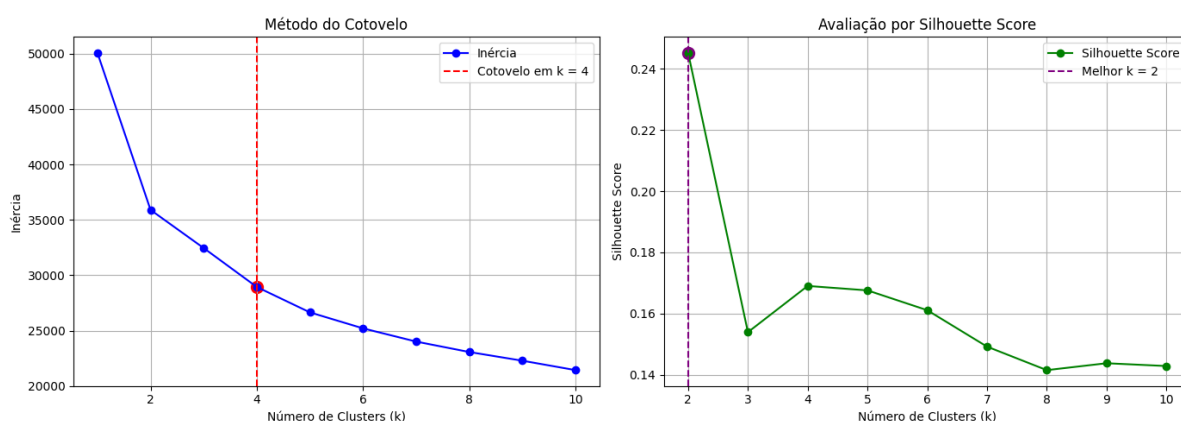
Portanto, para fazer os aglomerados, inicialmente buscou-se utilizar todas as variáveis componentes do IBC bem como o próprio IBC como indicador, afinal, uma vez construído, ele também pode ser utilizado como variável/critério para os aglomerados. As variáveis listadas inicialmente elencadas:

1. IBC 2024,
2. Cobertura Pop. 4G5G,
3. Densidade SMP,
4. HHI SMP,
5. Densidade SCM,
6. HHI SCM,
7. Adensamento de Estações,
8. Existência de Fibra Ótica,

9. Cobertura de área agricultável com *input*¹².

Foram realizados o gráfico do “cotovelo” e o teste de *Silhouette* em codificação *Python* para obter o resultado da clusterização e o número otimizado de conglomerados. Os resultados estão descritos nos gráficos abaixo:

Figura 12 –Gráfico de Cotovelo e Silhouette score para o aglomerado com 9 componentes



Verifica-se, pelo gráfico do método do cotovelo, que o método julgou que o número ótimo de k em 4 clusters. Porém, o teste de *Silhouette* máximo foi de 0,2452 para um número ótimo de 2 clusters. Para k=4, o resultado do teste de *Silhouette* foi de 0,169. Porém, em ambos os casos os resultados foram abaixo de 0,25, significando que estrutura de cluster pode não ser significativa ou bem estruturada.

4.2.5. Determinação das variáveis – redução de dimensionalidade.

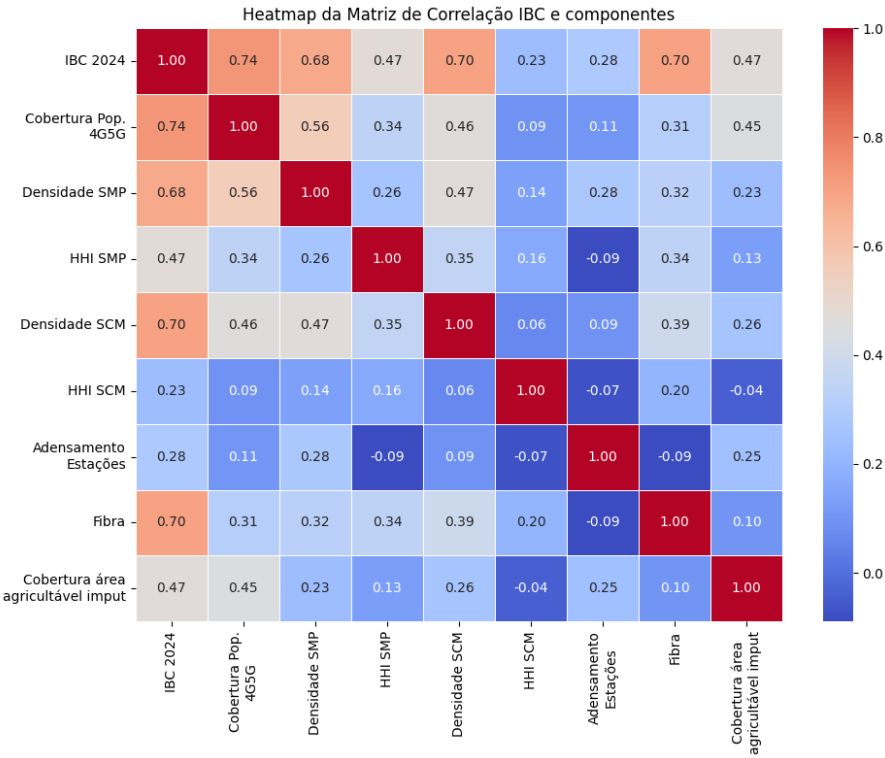
Como o desempenho da análise do conglomerado não foi satisfatório, partiu-se para um procedimento de redução de dimensionalidade, ou seja, partir para uma análise de conglomerado com menos variáveis.

No caso, a redução de dimensionalidade é uma etapa importante quando se trabalha com dados que possuem muitas variáveis (dimensões). Ela consiste em transformar os dados originais em um conjunto com menos variáveis, preservando ao máximo a estrutura e a variabilidade dos dados. Os objetivos principais desta técnica são descritos da seguinte

¹² A variável necessitou de imputação de dados para 13 municípios, uma vez que estes não possuem área agricultável legal. Nesse caso usamos os mesmos dados de cobertura da área urbana. Detalhes constam no Relatório Metodológico do IBC 2024.

forma: i) eliminar redundâncias, como variáveis altamente correlacionadas; ii) reduzir ruídos e complexidade, facilitando a identificação de padrões; iii) melhorar o desempenho dos algoritmos de clusterização, como k-means; iv) permitir visualização, já que dados em 2D ou 3D podem ser mais facilmente interpretados. Assim, nesse caso o primeiro passo foi realizar a matriz de correlação das variáveis para buscar possíveis variáveis redundantes. A matriz de correlação também foi gerada em código Python, justamente para verificar se há alto grau de correlação para eliminar variáveis redundantes. Nesse caso, uma possível redundância seria as variáveis com correlação maiores que 0,8 ou 0,85. Segue a matriz de correlação.

Figura 13 – Matriz de correlação do IBC e seus componentes.



Percebe-se que nenhuma variável possui uma correlação superior a 0,8. Logo, não foi possível a eliminação de variáveis por redundância a priori.

Não conseguindo reduzir a dimensionalidade a partir da correlação partiu-se para uma segunda técnica, que foi uma sequência de clusterizações com combinações de variáveis, partindo de 2 a 9 variáveis, alterando também a quantidade de clusters. Novamente, foi criado um código em Python que carrega um conjunto de variáveis que representam indicadores diversos e realiza a padronização dos dados.

Considerando a nova lista de variáveis analisadas, o código testou todas as combinações possíveis entre duas ou mais variáveis, e, para cada combinação, aplicou o algoritmo *K-Means* com diferentes valores de *k* (número de clusters), de 2 até 9. Para cada caso, calculou-se o *Silhouette score*, que mede o quão bem os dados estão agrupados — valores mais altos indicam clusters mais definidos e separados.

$$\sum_{r=2}^9 \binom{9}{r} = \binom{9}{2} + \binom{9}{3} + \binom{9}{4} + \binom{9}{5} + \binom{9}{6} + \binom{9}{7} + \binom{9}{8} + \binom{9}{9}$$

Tal exercício totalizou 502 combinações de variáveis possíveis; multiplicando o total de combinações pelas diferentes determinações de *k* em 8 possibilidades (2 a 9 clusters), obteve-se um total de 4.016 (502 × 8) possíveis clusterizações diferentes. Para otimização no processamento, nesta fase foram reduzidas as amostras para 1.000 municípios para obter um melhor desempenho computacional. Abaixo estão os 20 melhores resultados encontrados.

Tabela 7 – Melhores resultados de aglomerados por *Silhouette score*

Fibra, Cobertura área agricultável imput	9	0.622384
Fibra, Cobertura área agricultável imput	8	0.621916
Fibra, Cobertura área agricultável imput	7	0.603371
Densidade SCM, Fibra	8	0.60269
Densidade SCM, Fibra	7	0.597739
Densidade SCM, Fibra	9	0.597655
Cobertura Pop. 4G5G, Fibra	8	0.584354
Cobertura Pop. 4G5G, Fibra	7	0.58225
IBC 2024, Fibra	2	0.57924
HHI SMP, Fibra	9	0.576289
Fibra, Cobertura área agricultável imput	6	0.573486
Adensamento Estações, Fibra	7	0.570285
HHI SCM, Fibra	9	0.570168
HHI SMP, Fibra	8	0.566149
Adensamento Estações, Fibra	8	0.563525
IBC 2024, Fibra	6	0.562976
Cobertura Pop. 4G5G, Fibra	9	0.561053
Densidade SCM, Fibra	6	0.559699
IBC 2024, Fibra	7	0.558748
HHI SMP, Fibra	6	0.557384

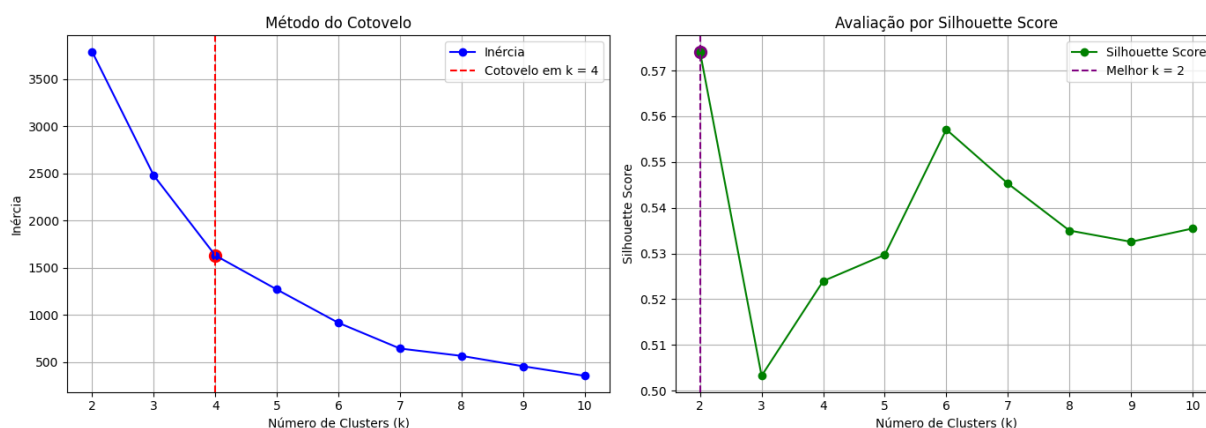
Como se verifica, o melhor desempenho em clusterização foi a combinação de duas variáveis (“Presença de fibra” e “cobertura de área agricultável”), com um número de *Silhouette score*

de 0,622. Porém, tal como praticamente todos os vinte melhores resultados, os números clusters “k” foram muito altos. No caso do melhor resultado de clusterização, o k foi de 9. Além disso, todos os testes ficaram com *Silhouette* score acima de 0,55.

Assim, entende-se que é necessária uma redução da dimensionalidade (quantidade de variáveis usadas para fazer os clusters), porém, tal ação resultou num número elevado de clusters. Há, portanto, que se buscar uma solução para achar um número não tão alto de clusters, mas com um desempenho medida pelo teste de *Silhouette* adequada.

Nesse caso, a própria tabela de resultados nos evidencia um caminho. A exceção observada dos melhores resultados encontrados foi a combinação das duas variáveis: ‘IBC 2024’ e ‘Presença de Fibra’, tendo um score de 0,579. Essa é uma informação de orientação para seguir com a análise de aglomeração. Aplicamos então novamente o gráfico de cotovelo e o teste de *Silhouette* para a combinação destas variáveis.

Figura 14 –Gráfico de Cotovelo e *Silhouette* score para o aglomerado: ‘IBC 2024’ e ‘Presença de fibra’



Como se observa, verificamos pelo método do cotovelo que o total de k em clusters seria recomendado em 4. Além disso, o comportamento pelo gráfico de *Silhouette* score mostra (como já dito) que o total de clusters recomendados seria de 2 (score de 0,574). Quando aumenta o k para 3, o desempenho cai para próximo de 0,50. Posteriormente, com a quantidade de k aumenta, também o score de *Silhouette* aumenta, mas ainda fica abaixo de 0,574. Em k=4 temos o score de 0,524. Seja como for, todos os valores atingiram acima de

0,5, o que indica uma boa estrutura de cluster, substancialmente melhor do que a clusterização com todas as variáveis componentes.

Observam-se resultados muito aproximados entre os mais bem obtidos. Optou-se então pela combinação “IBC 2024” e “Fibra” por mais simplicidade e pelo próprio índice IBC ser uma composição dos demais indicadores, assim a perda de qualidade na estrutura de aglomeração é significativamente baixa. Além disso, mantém um bom grau de representatividade na composição dos clusters. Por esses motivos, a combinação “fibra” e “IBC 2024”, com $k=4$, foi a escolhida.

4.2.6. Análise de agrupamento – criação dos aglomerados

Feito então o agrupamento formando por 4 aglomerados (clusters) conforme mencionado, escolhendo como variáveis para os conglomerados o “IBC 2024” e a variável “Presença de Fibra Ótica” nos municípios.

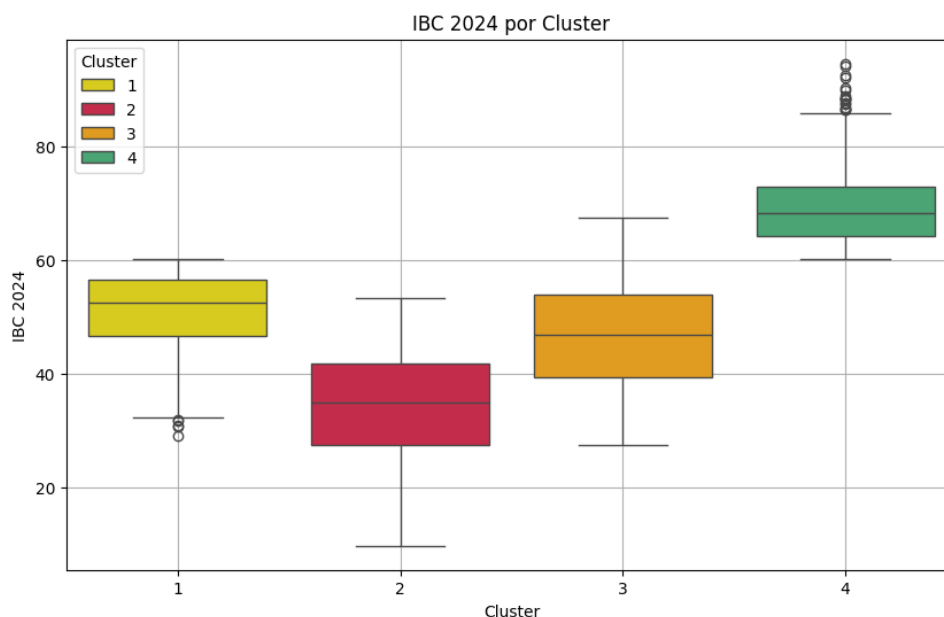
Após a execução da clusterização, seguem os resultados de estatísticas descritivas segmentadas por clusters. Verificou-se um número equilibrado de municípios nos quatro clusters, e percebe-se claramente o cluster nº 4 com a média mais elevada e o cluster nº 2 sendo aquele com a média mais baixa.

Tabela 8 – Estatísticas Descritivas por clusters

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Contagem (n)	1380.00	1187.00	1217.00	1781.00
Média	51.14	34.61	46.76	69.11
Desvio padrão	6.51	9.05	9.22	5.99
Mínimo	29.04	9.81	27.52	60.16
1º Quartil (Q1)	46.75	27.52	39.47	64.33
Mediana (Q2)	52.45	35.01	46.98	68.19
3º Quartil (Q3)	56.51	41.85	54.01	72.98
Máximo	60.15	53.33	67.37	94.57
Moda	29.04	9.81	27.52	60.16
Variância	42.33	81.90	84.95	35.90
Coef. de Variação	0.13	0.26	0.20	0.09
Assimetria (Skewness)	-0.67	-0.14	-0.03	0.72
Curtose	-0.29	-0.80	-0.90	0.28
Intervalo (Máx - Mín)	31.10	43.52	39.85	34.41

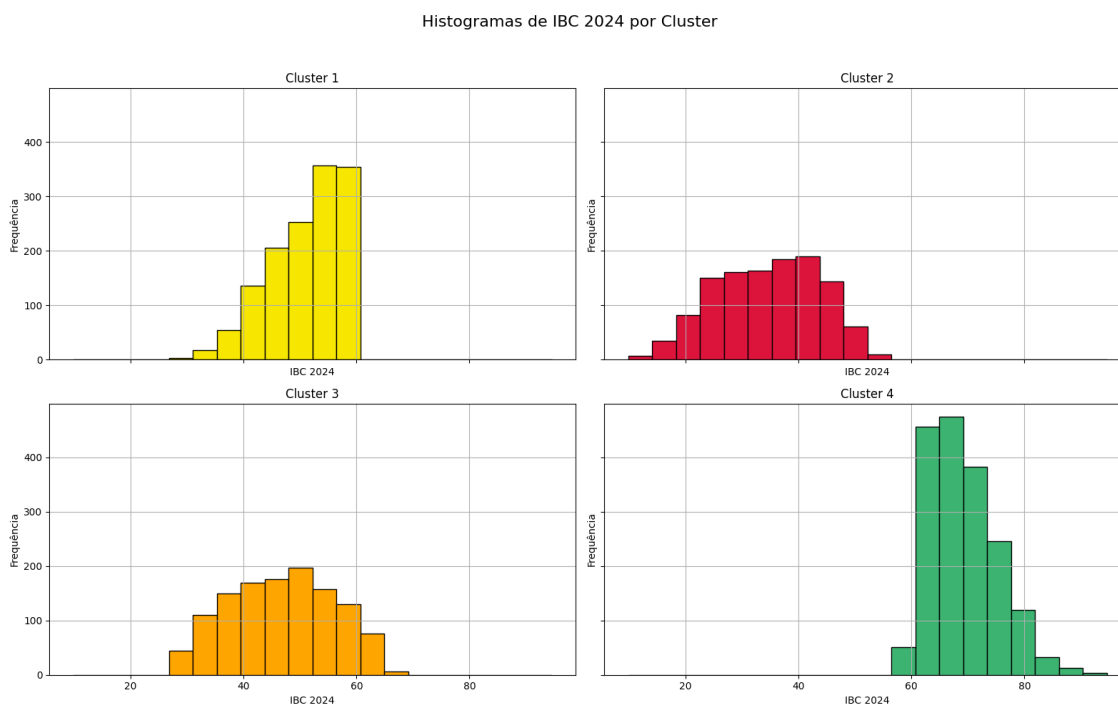
O gráfico de *box-plot* por clusters suplementa esta análise. Ilustram-se as diferenças.

Figura 15 – Gráfico *box-plot* por cluster



Em complementar ao gráfico *box-plot* seguem quatro histogramas detalhados pelos clusters. Nota-se a distribuição com níveis de IBC mais altos no cluster 4 (verde), bem como uma distribuição mais baixa (com resultados e notas mais baixas do IBC) nos clusters 2 (vermelho).

Figura 16 – Histogramas por clusters do IBC 2024



É interessante notar a segmentação executada pelos clusters. Usamos como ilustração a média do IBC e dos componentes com segmentação pelos clusters. Como se percebe, o cluster 4 é o de maior média (69,11); já com o cluster 2 é o de menor média (34,61). Além disso, entre os dois clusters de médias de IBC próximas (cluster 1 a 3), nesse caso se percebe que a variável “Presença de Fibra” é a que segmenta ambos os clusters. Aqui, percebe-se o cluster 1 como uma média de 100 para a variável “Fibra” e o cluster 3 com média de 48,15. Isso mostra que esse cluster tem municípios com nenhuma ou somente uma empresa de fibra no município. Lembrando que a nota “0” é destinada aos municípios sem presença de fibra e a nota “0,5” é atribuída aos municípios com uma empresa com fibra. Isso reforça a evidência da importância da presença de fibra na conectividade do município.

Além disso, a variável que mais diferencia os clusters 1 e 3 é “Fibra”. Isso mostra a predominância do cluster 3 em relação ao 1 na priorização de possíveis investimentos de recursos no contexto de implantação, arquitetura e avaliação de políticas públicas.

Tabela 9 – Estatísticas Descritivas por clusters

Cluster	1	2	3	4
IBC 2024	51.14	34.61	46.76	69.11
Cobertura Pop. 4G5G	72.83	69.55	75.00	92.63
Densidade SMP	46.96	42.46	46.93	63.31
HHI SMP	38.22	28.37	32.50	49.13
Densidade SCM	20.27	13.56	21.44	51.59
HHI SCM	53.28	47.93	49.18	59.91
Adensamento Estações Fibra	21.95	29.98	34.15	35.07
Cobertura área agricultável imput	100.00	1.10	48.15	98.60
	43.74	52.01	59.36	76.00

Em complemento, tal como foi feito na análise por quadrantes, construiu-se um mapa do país, com recorte ao nível municipal, hachurando as áreas dos municípios com cores específicas segundo os clusters. Essa visão é um avanço em relação à segmentação regional e suplanta a análise de quadrantes. Somado ainda ao mapa nacional segmentado pelos cluster, foi feito um gráfico de dispersão 3D para ilustrar a segmentação por clusters. Tal mapa utilizou o IDHM, o IBC e o IVS, mas “invertendo a polaridade” do IVS, uma vez que este tem relação inversa com o IDHM. Para tal foi feita a equação $(1-IVS)$, criando o indicador IDS. Seguem as ilustrações:

Figura 17 – Mapa dos municípios brasileiros por clusters do IBC 2024

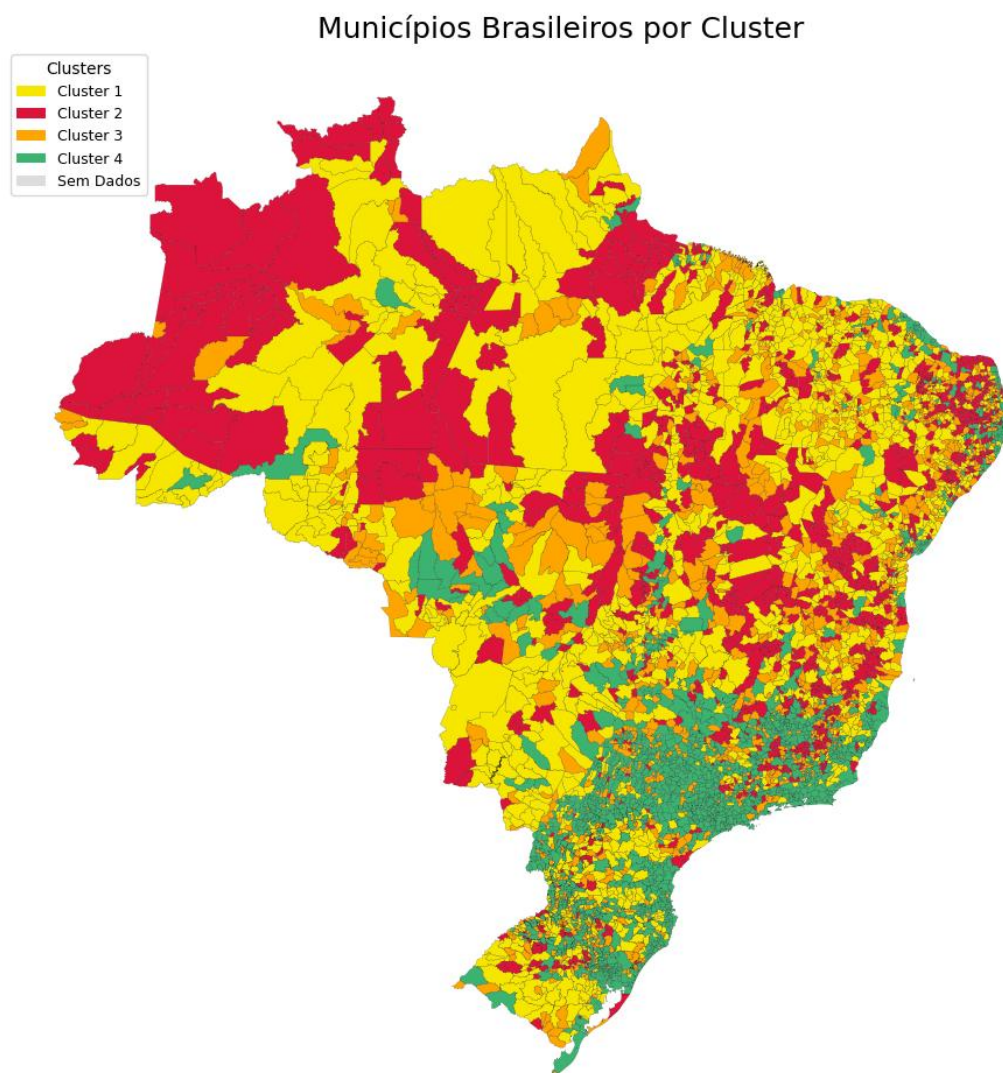
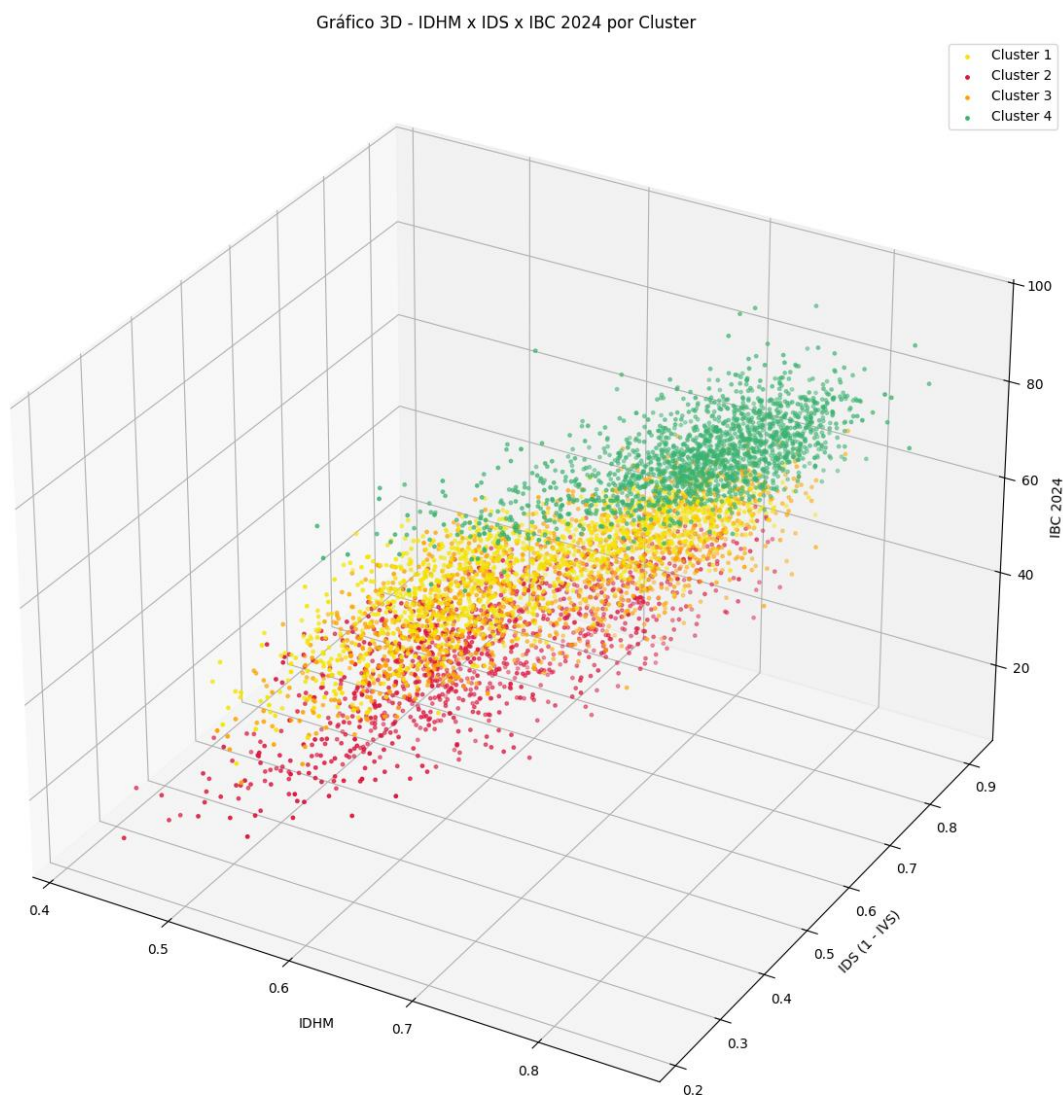


Figura 18 – Gráfico 3D com IBC (2024), IDHM e Índice de Desenvolvimento Social em 4 clusters



Após a apresentação da análise de clusters com a apresentação geográfica (mapa) e um gráfico ilustrativo tridimensional, apresentamos uma matriz com cruzamento de dados que segmenta ambas as análises. Nesse caso, as variáveis que segmentaram a matriz foram o 'Quadrante 3' e o 'Cluster 2'. Essas são aglomerações que apresentaram municípios com os menores índices de conectividade.

A razão para tal é, por óbvio, passar a indicar ou evidenciar quais os municípios diagnosticados como de "Baixa conectividade" por dois métodos distintos aplicados. Assim, tem-se um diagnóstico mais apurado de quais municipalidades há uma defasagem relativa na conectividade na realidade do Brasil para o ano em 2024.

Tal ação visa contribuir com camadas informacionais para possíveis tomadores de decisão ou formuladores de políticas públicas na conduta de busca de ampliação de conectividade e universalização de serviços, permitindo uma melhor orientação de onde aplicar recursos públicos limitados, melhorando potencialmente a eficácia e a efetividade de políticas de conectividade.

Sugere-se também a complementação de tal perspectiva com outros pontos para a elaboração das políticas, tais como: i) análise de eficácia per capita do uso dos recursos públicos; ii) adoção de critérios de redundância e resiliência da rede em complementariedade à ampliação da conectividade; iii) analisar aspectos de ociosidade de redes implantadas. A matriz de segmentação segue apresentada.

Tabela 10 – Matriz de cruzamento de dados de clusters e quadrantes

Segmentação dos municípios com cruzamento "Quadrante x Cluster"				
Cluster 2	Quadrante 3			
		SIM	NÃO	TOTAL
	SIM	962	225	1187
	NÃO	1204	3174	4378
	TOTAL	2166	3399	5565

Percebe-se que dos 5.565¹³ municípios analisados, 2.166 pertencem ao Quadrante 3 enquanto 1.187 pertencem ao cluster 2. Destes aglomerados, um conjunto de 962 municípios são classificados em ambos os aglomerados. Ou seja, em dois métodos de segmentação/aglomeração foram classificados 962 municípios com baixa conectividade (baixo IBC) e baixo desenvolvimento humano e social (medidos por IDHM e IVS). Assim, tais municípios possuem uma carência relativa de conectividade em relação ao restante do país. Sugere-se, portanto, na elaboração de políticas públicas, uma orientação para priorização a esse conjunto de municípios brasileiros.

¹³ A listagem dos municípios, com as respectivas classificações em clusters e quadrantes, segue em arquivo de planilha ‘.xls’ avulso. Já a orientação para os 5 municípios faltantes nesta análise é que o elaborador de política pública analise a realidade dessas localidades no caso concreto da elaboração da política.

5. Modelo econométrico – SAR (*Spatial Autoregressive Model*)

Nesta parte do trabalho, para aprofundar a análise e trazer mais camadas informacionais, buscou-se fazer um exercício econométrico. A aplicação de um modelo econométrico permite captar com maior precisão os efeitos (que podem ser temporais ou espacial) da conectividade digital entre municípios. Ao incluir a influência de elementos regionais, por exemplo, na explicação dos resultados locais, o modelo pode captar uma abordagem mais sofisticada para entender como a infraestrutura de conectividade se propaga territorialmente.

Um adequado modelo econométrico pode oferecer uma métrica clara de retorno indireto de investimentos. Trata-se de uma ferramenta interessante para subsidiar decisões públicas, especialmente em escala regional. O exercício econométrico fornece uma base para formular políticas públicas orientadas por evidências, priorizando intervenções que buscam otimizar não só os benefícios locais, mas também os ganhos territoriais — algo que é especialmente relevante em contextos de redes de serviços compartilhados.

Nesse sentido, foi arquitetada a possibilidade de estudar o(s) efeito(s) local(is) do estágio de conectividade de municípios em demais localidades. Para tal exercício, foi feita uma análise prévia para escolher o tipo de modelo econométrico. Nesse caso, avaliando a estrutura de dados do próprio IBC, observam-se algumas características nos dados que favorecem um modelo econométrico do tipo SAR.

As características dos dados são descritas da seguinte forma: i) o IBC mede a infraestrutura de conectividade e depende de malhas físicas partilhadas (backhaul, torres, ERBs); ii) os dados do IBC são influenciados por investimentos regionais ou consorciados; iii) o IBC apresenta, conforme já verificado em mapas e nas análises de aglomerados, efeitos de difusão regional (ex.: expansão da fibra ou cobertura móvel). Essas características sugerem que o nível de IBC de um município pode ser direta e indiretamente afetado pelo nível de IBC nos municípios vizinhos — e isso é uma condição típica para aplicar o SAR (*Spatial Autoregressive Model*).

Um modelo econométrico SAR é um tipo de modelo de regressão espacial que considera a influência da proximidade geográfica entre as observações na variável de resposta. Em outras palavras, o valor de uma observação pode ser afetado pelos valores de outras observações próximas a ela, criando uma autocorrelação espacial. Nesse caso, tal modelagem permite buscar uma forma de mensuração do chamado “efeito vizinhança” ou “efeito *spillover*”, referindo-se ao impacto que uma ação, política ou fenômeno tem sobre aqueles que estão por perto e que não foram diretamente afetados.

Assim, para modelar a influência espacial direta entre os valores do IBC do município, foi adotado o modelo SAR, que inclui uma defasagem espacial da variável dependente, sendo apropriado quando há alta autocorrelação espacial (ANSELIN, 1988; SAGE; PACE, 2009).

A escolha deve ser apoiada, entretanto, pela aplicação de um teste estatístico cujo resultado seja significativo do índice de Moran sobre os resíduos do modelo de mínimos quadrados ordinários (MQO ou OLS), o que indicaria a dependência espacial sobre o valor do IBC entre municípios vizinhos.

5.1. Descrição do modelo SAR (Spatial Autoregressive Model)

Após a definição do modelo, apresenta-se abaixo a equação geral do modelo a ser adotado. Ao modelo SAR tradicional, optou-se por adicionar uma variável do tipo “dummy”. Uma variável dummy (também chamada de variável indicadora ou variável binária) é uma variável que assume somente dois valores: 0 ou 1. Ela é utilizada para representar dados categóricos, ou seja, variáveis que não podem ser medidas numericamente. Fazendo uma analogia, a variável dummy é um artifício criado para representar as variáveis qualitativas de forma numérica. Em FÁVERO e BELFIORE (2017) é afirmado que “As variáveis dummy devem, portanto, ser utilizadas quando desejarmos estudar a relação entre o comportamento de determinada variável explicativa qualitativa e o fenômeno em questão, representado pela variável dependente”. Nesse sentido, é exatamente o que se apresenta neste tipo de modelagem, uma vez que se pretende estudar efeitos distintos nos clusters por diferenças no grau de conectividade. No caso em específico, justifica-se a inserção desta variável em função da análise de cluster feita previamente.

$$y_i = \rho \left(\sum_j w_{ij} y_j \right) + \beta_0 + \sum_{k=1}^{K-1} \beta_k D_{ik} + \varepsilon_i$$

Onde:

y_i é a variável dependente (IBC estimado do município i);

w_{ij} é o elemento da matriz de contiguidade espacial (W);

ρ é o coeficiente espacial (força do efeito dos vizinhos);

β_0 é o intercepto para o grupo de referência (ex: cluster 2);

D_{ik} é a variável dummy (1 se o município i pertence ao cluster k, 0 caso contrário);

ε_i é o resíduo quando estimado a variável dependente;

K é o total de clusters.

5.2. Adoção da variável dummy para os demais clusters

Uma das escolhas da modelagem foi inserir variáveis dummies dos clusters já previamente feitos. Primeiramente, os clusters (1 a 4) são agrupamentos classificados por homogeneidade, ou seja, os municípios foram agrupados em quatro grupos distintos com similaridades entre eles. No caso, a similaridade foi classificada pelo próprio IBC bem como a existência ou não de empresas de fibra ótica na localidade. Não foram feitos clusters, portanto, de regionalidade ou de proximidade.

Diante dessa abordagem, uma opção seria escolher um modelo com quatro equações, uma para cada cluster. Porém, tomou-se a decisão de uma única equação para facilitar a capacidade interpretativa. Isso pode capturar os efeitos de proximidade geográfica (spillover, difusão regional, etc.) e o modelo permite separar efeitos estruturais (tipo de município).

5.3. Teste I de Moran

Como mencionado anteriormente, é necessário aplicar um teste estatístico do Índice de Moran (I), que é uma estatística utilizada para medir o grau de autocorrelação espacial entre unidades geográficas. O índice trata da questão de saber se os valores de uma variável (por exemplo, o IBC dos municípios) estão distribuídos de forma aleatória ou espacialmente agrupada (*clusterizada*). Um Índice de Moran aplicado aos resíduos do modelo OLS (*Ordinary Least Squared* ou Mínimos Quadrados Ordinários) significativo é um indício de haver autocorrelação espacial que não é explicada pelo modelo tradicional.

Nesse caso:

- O modelo de mínimos quadrados ordinários (OLS) torna-se inadequado por violar o pressuposto de independência dos resíduos;
- No caso das estimativas, os coeficientes podem apresentar vieses ou ineficiências e os testes estatísticos associados (como os p-valor) se tornam inválidos.

Assim, o modelo SAR (*Spatial Autoregressive*) passa a ser mais adequado enquanto esse combina dependência espacial com equações de defasagem espacial para a série. Simplificando, o modelo SAR ajusta o valor de uma variável (por exemplo, o IBC) em função dos valores encontrados nos municípios vizinhos. Essa abordagem ajuda a melhorar a precisão e a validade estatística. A descrição matemática do teste segue abaixo, bem como a descrição da hipótese nula e da hipótese alternativa.

$$I = \frac{N}{S_o} \times \frac{\sum_{i=1}^N \sum_{j=1}^N w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^N (y_i - \bar{y})^2}$$

Onde:

N: número de observações (municípios);

y_i Valor da variável no município;

$\bar{y}$: média da variável.

w_{ij} elemento da matriz de pesos espaciais W, que define a vizinhança entre i e j;

S_o : soma total dos pesos.

A hipótese nula do teste (H_0) é não haver autocorrelação espacial (os valores são distribuídos aleatoriamente no espaço). Enquanto a hipótese alternativa (H_1) é que há autocorrelação espacial (positiva ou negativa) entre os valores da variável. Portanto, se o p-valor do teste for menor que 0,05 então rejeita-se a hipótese nula e presume-se haver autocorrelação espacial estatisticamente significativa.

5.4. Grau de contiguidade dos municípios e aplicação do teste de Moran.

5.4.1. Critério de contiguidade estabelecido.

A fim de verificar a presença de autocorrelação espacial nos resíduos do modelo tradicional de regressão linear (OLS), foi realizado o teste de Moran (I) com base no índice de

conectividade (IBC) dos municípios brasileiros em 2024. Utilizando uma matriz de pesos espaciais do tipo contiguidade de primeira ordem (Queen), os resíduos do modelo OLS — que incluía variáveis *dummy* para clusters de municípios — foram submetidos ao teste estatístico.

Cabe mencionar o conceito de “contiguidade”, que no contexto de modelos espaciais é necessário estabelecer algum critério para se definir o que é uma localidade (no caso em concreto um município) vizinha de outra localidade. A definição de vizinhança mais usual é a contiguidade de primeira ordem, onde dois municípios são considerados vizinhos se compartilham uma divisa em comum.

Há outras formas, como a contiguidade de segunda ordem, onde não só os vizinhos coligados são definidos, mas também os vizinhos dos municípios vizinhos iniciais. Pode-se definir vizinhança também por distância geográfica fixa (por exemplo até 50 km de distância) ou pelos n-vizinhos mais próximos, ou seja, estabelece um número limitado de municípios vizinhos.

No caso, a contiguidade de primeira ordem é amplamente utilizada em estudos espaciais e é popular por ser intuitiva e eficiente empiricamente. É muito comum observar que municípios adjacentes muitas vezes unem-se em infraestrutura, na rede de serviços e na economia social; portanto, eles têm uma inter-relação indiscutível no desenvolvimento econômico ou social das regiões vizinhas. Outro ponto é que esse tipo de matriz resulta numa estrutura que reduz a carga computacional. Por isso, como critério de vizinhança neste trabalho, foi adotada a contiguidade de primeira ordem.

Assim, a matriz adotada assume um valor para os municípios que são contíguos (adjacentes) e multiplica esse valor pelo IBC de cada respectivo município (o peso espacial vai depender da quantidade de municípios adjacentes). Abaixo, segue um mapa ilustrativo de uma região hipotética com cinco municípios e seus vizinhos, bem como a matriz de vizinhança¹⁴, também chamada matriz de proximidade espacial.

¹⁴ As ilustrações de matriz de vizinhança abaixo estão no artigo: SEFFRIN, Rodolfo; ARAÚJO, Everton Coimbra de; BAZZI, Claudio Leones. Análise espacial de área aplicada à produtividade de soja na região oeste do Paraná utilizando o software R. *Revista Brasileira de Geomática*, Curitiba, v. 6, n. 1, p. 23–43 **40**

Figura 19 – Exemplo ilustrativo de matriz de vizinhança

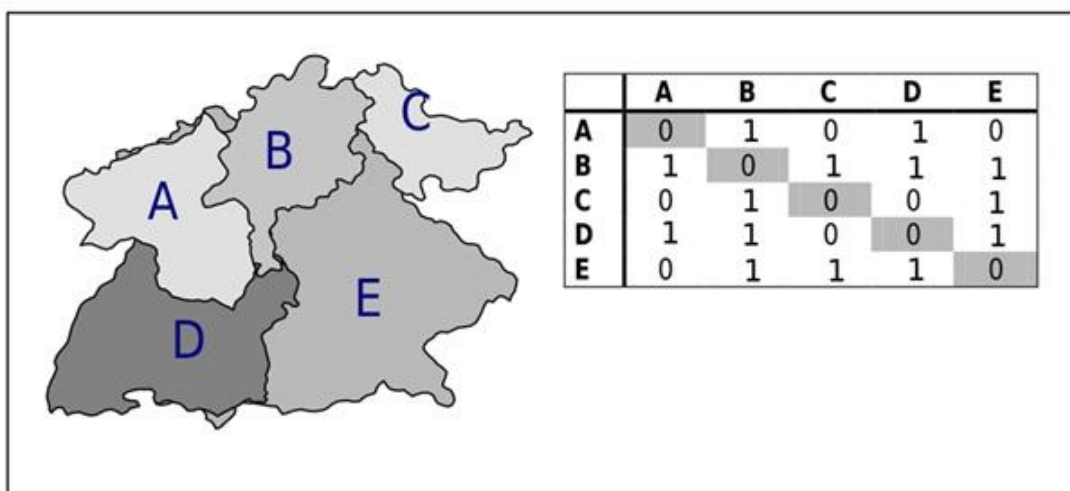
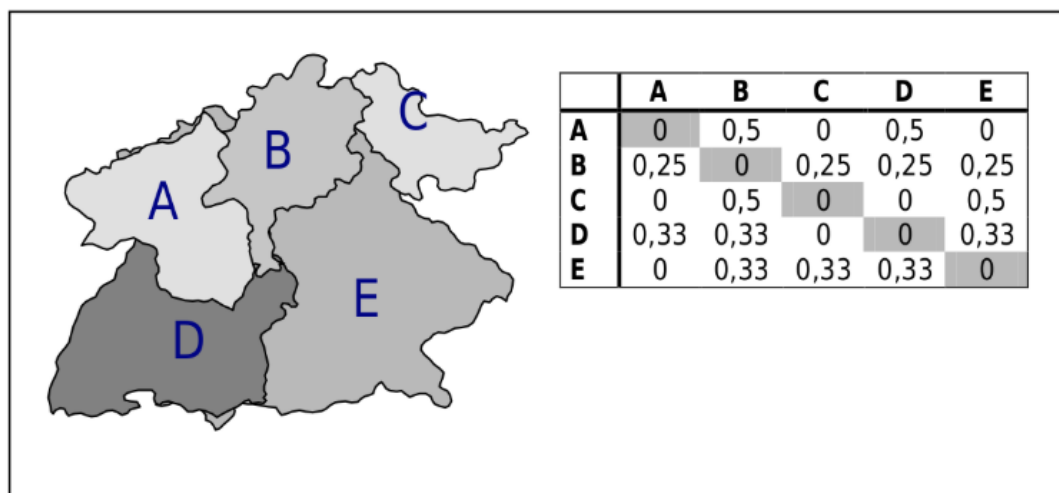


Figura 20 – Exemplo ilustrativo de matriz de vizinhança ponderada



5.4.2. Aplicação do teste de Moran

Os dados foram carregados em código Python, adotando alguns passos:

- Carregou-se os dados de “shape file” simplificado dos municípios brasileiros via GeoJSON;
- Juntaram-se esses dados com um *data frame* (df) contendo o IBC de 2024 e a classificação em clusters (1 a 4);
- Criou-se variáveis dummy para os clusters e ajustou-se um modelo de regressão linear (OLS) explicando o IBC como função dos clusters;
- Obtiveram-se os resíduos do modelo OLS e calculou-se a matriz de pesos espaciais de contiguidade de 1ª ordem (Queen);

- v) Aplicou-se o teste do índice de Moran sobre os resíduos para testar autocorrelação espacial.

Após processar o código obteve-se o então o seguinte resultado:

I de Moran: 03434.

Pseudo p-valor: 0,001.

A conclusão do teste foi que há autocorrelação espacial significativa nos resíduos do OLS, portanto um modelo espacial (ex SAR) passa a ser recomendado.

5.5. Aplicação do modelo SAR com clusters, efeitos diretos e indiretos

Ao se constatar a viabilidade de modelo espacial, partiu-se para aplicar um modelo SAR para capturar a dependência espacial presente no Índice Brasileiro de Conectividade (IBC) ao nível municipal. A variável dependente foi o valor do IBC em 2024, enquanto as variáveis independentes foram o próprio IBC, pois o modelo é autorregressivo, e *dummies* para clusters de municípios, previamente definidos de acordo com características de conectividade. O cluster 2 foi tomado como categoria de referência, permitindo comparar clusters adicionais com este grupo base.

A matriz de pesos espaciais utilizada foi do tipo contiguidade de primeira ordem (Queen), que considera dois municípios como vizinhos quando compartilham uma divisa em comum. Para garantir a comparação de coeficientes e a estabilidade numérica, as variáveis explicativas foram todas padronizadas pelo escore “z”.

A estimação do modelo foi realizada via máxima verossimilhança, utilizando a biblioteca spreg (PySAL). O modelo SAR incorpora um termo de defasagem espacial da variável dependente ($\rho W y$), onde ρ representa o coeficiente de autocorrelação espacial e W é a matriz de contiguidade.

Também buscou-se calcular os efeitos diretos, indiretos e totais das variáveis explicativas usando álgebra matricial. Nesse caso, o efeito direto mede o impacto médio de uma variável sobre a própria unidade; o efeito indireto mede impacto transmitido aos vizinhos (*spillover*) e o efeito total é a soma de ambos os efeitos.

O modelo SAR passa a ser descrito pela seguinte equação:

$$IBC_i = \rho * (W * IBC_j) + \beta_0 + \beta_1 * Cluster_1 + \beta_2 * Cluster_3 + \beta_3 * Cluster_4$$

Onde:

o cluster 2 (de menor grau conectividade) é a base de referência;

IBC_i representa o valor do IBC em determinado município i ;

$(W * IBC_j)$ representa a média ponderada do IBC nos municípios vizinhos de i , segundo matriz de pesos espaciais cuja contiguidade é de primeira ordem¹⁵;

ρ representa o coeficiente espacial;

β_0 representa o intercepto;

$\beta_1, \beta_2, \beta_3$ são os coeficientes caso o município pertença a outros clusters.

5.5.1. Resultados obtidos na aplicação do modelo SAR.

Após a explicação do arcabouço teórico do modelo e dos procedimentos estatísticos adotados para realizar a estimativa, os resultados obtidos seguem abaixo. Com base na estimação do modelo SAR, a seguinte equação foi obtida para explicar o Índice Brasileiro de Conectividade (IBC) dos municípios:

$$IBC_i = 0,4318 * (W * IBC_j) + 29,5945 + 6,3985 * Cluster_1 + 3,8367 * Cluster_3 + 12,2467 * Cluster_4$$

Intercepto ($\beta_0 = 29,5945$) representa o valor médio estimado do IBC para municípios do Cluster 2 (grupo de referência), na ausência de efeitos espaciais. O p-valor ($< 0,0001$) confirma que esse valor é estatisticamente significativo.

$\beta_1 = 6,3985$: municípios pertencentes ao Cluster 1 apresentam, em média, um IBC 6,40 pontos maior que os do Cluster 2, antes de considerar os efeitos de vizinhança.

$\beta_2 = 3,8367$: municípios do Cluster 3 têm, em média, um IBC 3,84 pontos superior ao do grupo de referência.

$\beta_3 = 12,2467$: apresenta o maior impacto direto entre os grupos, com um IBC 12,25 pontos acima da média do Cluster 2.

¹⁵ Para evitar ambiguidades, neste trabalho adotamos a notação IBC_i para representar o valor do índice de conectividade no município i , e IBC_j para os valores correspondentes nos municípios vizinhos de i , conforme definidos pela matriz de contiguidade espacial.

Tanto o intercepto quando os demais “betas” dos demais clusters são positivos, o que se coaduna com a teoria, uma vez que o cluster de referência escolhido foi o de menor estágio de conectividade. Os “betas” representam, então, o valor médio estimado de aumento do IBC de um município se ele “sair” de um cluster 2 para um cluster de maior estágio de conectividade. Por exemplo, o cluster 4 apresenta o maior impacto no IBC.

O coeficiente espacial ($\rho = 0,4318$) indica uma autocorrelação espacial positiva significativa, ou seja, o IBC de um município é positivamente influenciado pelo IBC médio dos seus vizinhos imediatos. **O valor de 0,4318 sugere que aproximadamente 43% da variação local pode ser explicada pela vizinhança, evidenciando efeitos de difusão regional.**

5.5.2. Estimativa dos efeitos diretos e indiretos e grau de ajustamento

Cabe mencionar, contudo, que o efeito do coeficiente “ ρ ” (*rho*) não é ele mesmo o efeito de *spillover* (efeito indireto), mas sim um indicativo da presença e intensidade da dependência espacial direta na variável dependente.

O efeito *spillover* é um impacto indireto que ocorre quando uma mudança em uma variável explicativa em um município influencia o resultado nos municípios vizinhos — ou seja: não é o IBC em si que se transmite, mas sim o efeito de uma variável X (como estar em um cluster) sobre outros municípios. Esse efeito só pode ser quantificado depois que o modelo SAR é estimado, com base no coeficiente espacial, na matriz de pesos e nos “betas”. Além dos efeitos marginais, foram calculados os efeitos espaciais compostos de cada cluster. Seguem os resultados a seguir.

Tabela 11 – Efeitos espaciais do modelo SAR para IBC 2024 e clusters de conectividade

Cluster	Efeito Direto	Efeito Indireto	Efeito Total
Cluster 1	6,655	4,604	11,259
Cluster 3	3,991	2,761	6,751
Cluster 4	12,738	8,812	21,549

Como exemplo, o efeito do cluster 4 não só beneficia diretamente o município ao qual pertence, mas também exerce um efeito positivo expressivo sobre seus vizinhos — sugerindo que estratégias concentradas neste perfil de município podem produzir ganhos regionais em conectividade.

Soma-se ainda que o pseudo- R^2 do modelo foi calculado em $R^2 = 0,8261$, o que indica que o modelo explica 82,6% da variabilidade do IBC. O que pode ser considerado um ajuste bastante significativo.

Log-likelihood = -18.165,01, AIC = 36.340,01, BIC = 36.373,13 são valores que ajudam a reforçar a análise do modelo e podem ser usados para comparações com outras especificações.

5.6. Testes de robustez do modelo SAR

Para completar a construção do modelo, foram conduzidos alguns testes de robustez. A robustez e a consistência do modelo SAR ajustado foram avaliadas por meio de um conjunto de testes estatísticos complementares. Assim, os referidos testes têm o objetivo de auferir a força, estabilidade e confiabilidade do modelo. Optou-se por criar uma tabela de resumo dos testes elaborados, sem adentrar na descrição completa de cada teste respectivo, pois o objetivo geral, como dito, é a avaliação do modelo. Segue a tabela de testes de robustez efetuados.

Tabela 12 – Testes de robustez propostos para o modelo SAR

Teste	Objetivo	Critério de Interpretação	Resultado Esperado
Moran's I dos resíduos OLS z e p-valor de ρ (rho) p-valores dos coeficientes (β) VIF (Variance Inflation Factor) Moran's I dos resíduos do SAR AIC / BIC / Log-likelihood Breusch-Pagan	Detectar autocorrelação espacial	$p < 0,05$ indica autocorrelação	Recomendado uso de modelo SAR
	Verificar dependência espacial	$p < 0,05$ valida o uso do modelo SAR	Justifica uso de estrutura espacial
	Verificar relevância das variáveis	$p < 0,05 \rightarrow$ coeficiente significativo	Identifica variáveis explicativas
	Avaliar multicolinearidade	$VIF < 5 \rightarrow$ sem multicolinearidade	Modelo estável
	Verificar autocorrelação nos resíduos	$p > 0,05 \rightarrow$ sem autocorrelação residual	Modelo bem especificado
	Avaliar e comparar o ajuste do modelo	AIC/BIC menores $\rightarrow$ melhor modelo	Comparar com modelo OLS ou outros
	Verificar heterocedasticidade nos resíduos	$p > 0,05 \rightarrow$ resíduos com variância constante	Heterocedasticidade não detectada

5.6.1. Significância dos parâmetros (teste z)

Todos os coeficientes estimados, incluindo o intercepto e as variáveis dummy de cluster, apresentaram valores de z extremamente elevados e p-valores inferiores a 0,0001, o que indica serem estatisticamente diferentes de zero com alto grau de confiança.

Tabela 13 – Resultados dos testes de parâmetros

Parâmetro	β	z-valor	p-valor
Intercepto	29,595	62,900	0,000
Cluster 1	6,399	58,170	0,000
Cluster 3	3,837	35,460	0,000
Cluster 4	12,247	90,020	0,000
Coef. Espacial (ρ)	0,432	49,370	0,000

Estes resultados confirmam que o intercepto bem como as diferenças entre clusters são estatisticamente significativas e que a dependência espacial entre municípios é relevante.

5.6.2. Multicolinearidade (VIF – Variance Inflation Factor)

O VIF foi calculado para as variáveis explicativas (dummies dos clusters), resultando em valores iguais a 1,0 para todas. Isso indica ausência multicolinearidade, ou seja, as variáveis não estão linearmente correlacionadas entre si, garantindo a estabilidade das estimativas e precisão dos testes estatísticos.

5.6.3. Autocorrelação espacial dos resíduos (Moran's I)

Apesar de o modelo SAR já incluir um componente espacial, foi aplicado o Índice de Moran aos resíduos do próprio SAR, como teste de diagnóstico adicional. O valor obtido foi:

Moran's I = 0,0640
p-valor = 0,0010

Embora o índice seja pequeno, o p-valor indica que ainda há resíduos espacialmente correlacionados significativamente, mas de forma bem baixa. Esse número no p-valor levanta a evidência de concluir que apesar significativo, o nível de autocorrelação espacial dos resíduos do modelo SAR é baixo.

5.6.4. Heterocedasticidade (Breusch-Pagan)

O teste de *Breusch-Pagan* foi aplicado para verificar se há variância não constante nos resíduos (heterocedasticidade), o que pode comprometer a eficiência dos estimadores. Os resultados obtidos estão descritos abaixo:

LM Stat = 490,15 ($p < 0,0001$)
F Stat = 179,04 ($p < 0,0001$)

Ambos os testes indicam que a heterocedasticidade está presente nos resíduos. Isso significa que a variância do erro não é constante entre os municípios, o que pode ocorrer, por exemplo, em regiões com padrões extremos de conectividade. Apesar disso, o modelo SAR ainda é válido, e o problema pode ser mitigado com métodos robustos à heterocedasticidade, caso necessário.

Todavia, a heterocedasticidade constatada pode ser classificada como leve ou moderada, pois nesse caso até se buscou fazer um procedimento de correção do modelo, reestimando o modelo SAR com erro padrão robusto à heterocedasticidade, utilizando o estimador robusto tipo White. Também foi feita uma tentativa de modificação do modelo, considerando um modelo tipo SARAR, mas a questão da heterocedasticidade permaneceu. Assim, após um novo exercício de estimativa do modelo, os resultados foram praticamente os mesmos, com resultados similares no teste de *Breusch-Pagan*. Desta maneira, opta-se por permanecer com a estimativa do modelo SAR.

5.6.5. Avaliação geral dos testes e robustez.

Os testes realizados indicam que o modelo SAR ajustado é estatisticamente robusto, com coeficientes significativos, ausência de multicolinearidade e redução substancial da autocorrelação espacial dos resíduos.

Contudo, constatou-se a presença de heterocedasticidade nos resíduos, sugerindo a possibilidade de ajustes adicionais (como erros robustos). Porém, tal heterocedasticidade pode ser considerada baixa ou moderada. Na tentativa de ajustar o modelo para corrigir tal problema, os novos resultados principais praticamente foram muito similares aos resultados iniciais. O modelo, portanto, se mostra adequado para a análise do IBC 2024 ao nível municipal com componentes espaciais (clusters de homogeneidade). Para trabalhos futuros, há a possibilidade de ajustes com modelagens alternativas que podem ser consideradas. De todo modo, o modelo SAR se apresentou estatisticamente consistente e superior a modelos de mínimos quadrados ordinários (OLS).

6. Conclusões finais

Após a adaptação do IBC para uma nova versão, com a instauração de nova variável, nova ponderação e uma revisão com alterações metodológicas, foram feitos alguns testes e análises para avaliar a qualidade e robustez desta nova versão.

Foi constatada correlação elevada do IBC 2024 com alguns índices e indicadores que avaliam a infraestrutura e o grau de desenvolvimento socioeconômico dos municípios. Notadamente, verificou-se uma correlação com o Índice de Desenvolvimento Humano Municipal (IDHM), com o Índice de Vulnerabilidade Social (IVS) e especialmente com o Indicador da Conectividade Rural (ICR). Essas evidências mostram um grau de confiabilidade na construção do índice.

Cabe ressaltar a correlação de dois indicadores com o IBC: a primeira é a correlação com o IVS de infraestrutura urbana e a segunda a relação do IBC e o PIB per capita. Apesar de intuitivamente essas correlações seriam em tese mais fortes, uma interpretação dos resultados pode evidenciar uma boa atuação regulatória. Isso porque as correlações mais fracas mostram que o IBC está sendo equilibrado na avaliação entre municípios grandes e pequenos e entre municípios ricos e pobres. No caso, o IVS de infraestrutura evidencia que a conectividade nos municípios contribui para a redução da vulnerabilidade social em infraestrutura, porém, outros equipamentos e serviços de infraestrutura urbana são considerados, como por exemplo iluminação, asfaltamento, limpeza urbana, rede de energia elétrica, tratamento de água potável e esgoto, etc. Isso pode indicar um ponto forte do IBC bem como um indicador de missão institucional executada da ANATEL.

Foi feita a análise de dados gerais do IBC, com apresentação de estatísticas descritivas gerais, gráfico box-plot por região, histograma e mapa de calor. Tais gráficos e tabelas mostraram a dinâmica do grau de conectividade do país para 2024. Acrescentado a esses exercícios duas análises de agrupamento foram feitas: i) uma análise de quadrante, fazendo uma segmentação dos municípios num corte da distribuição pela mediana de duas variáveis – IDHM e IBC e ii) análise de clusters para agrupar os municípios por similaridade ou homogeneidade.

No caso da análise de cluster, após testes estatísticos (Silhouette score) constatou-se uma boa estrutura de agrupamento usando como variáveis decisórias o próprio IBC de 2024 e a existência

de fibra ótica nos municípios, com uma abordagem com 4 clusters (aglomerados). Um cluster foi considerado com o maior grau de conectividade, dois clusters com conectividade moderada e o último (chamado de cluster 2) tendo baixa conectividade. Entre os dois aglomerados intermediários, a variável que segmentou os clusters foi a pontuação da variável categórica de “existência de fibra”.

Com os dois exercícios de aglomerados, foi feito um cruzamento de dados onde obtive o conjunto de municípios que, tanto pela análise de quadrantes como pela análise de clusters, são de baixo grau de conectividade. Foram 962 municípios identificados nesta situação e, segundo a avaliação deste trabalho, essas seriam as localidades a receberem alguma priorização em eventuais políticas públicas.

Após a confecção dos dois exercícios de análise de agrupamentos, utilizou-se os mesmos dados produzidos da análise de aglomerados para construir um modelo econométrico espacial, na tentativa de estimar o(s) efeito(s) local(is) do estágio de conectividade de municípios em demais localidades. Para tal exercício, foi feita uma análise prévia para escolher a adoção de qual tipo de modelo econométrico a ser adotado. Nesse caso, avaliando a estrutura de dados do próprio IBC observam-se algumas características nos dados que favorecem um modelo econométrico do tipo SAR (Spatial Autoregressive Model), dadas as características específicas dos dados do IBC (interdependência de malhas físicas de infraestrutura conjuntas, possível efeito de difusão regional, influência por dinâmicas de investimentos regionalizadas).

Um modelo econométrico SAR é um tipo de modelo de regressão espacial que considera a influência da proximidade geográfica entre as observações na variável de resposta. Tal modelagem permite buscar uma forma de mensuração do chamado efeito vizinhança" ou "efeito *spillover*", referindo-se ao impacto que uma ação, política ou fenômeno tem sobre aqueles que estão por perto e que não foram diretamente afetados.

O modelo construído então usou como variável dependente o IBC dos municípios específicos e como variáveis dependentes o IBC dos municípios adjacentes do município específico (ponderados pelo peso indicados por uma vizinhança ponderada), além de variáveis dummy em função a qual cluster de conectividade o município pertence (baseado no exercício de aglomerado anteriormente feito).

Com um $R^2 = 0,8261$, o modelo mostrou um grau de ajuste significativo, ou seja, o modelo explica 82,6% da variabilidade do IBC. Com um coeficiente espacial ($p = 0,4318$) o modelo indica uma autocorrelação espacial positiva significativa, ou seja, o IBC de um município é positivamente influenciado pelo IBC médio dos seus vizinhos imediatos. Esse valor de 0,4318 sugere que aproximadamente 43% da variação local pode ser explicada pela vizinhança, evidenciando efeitos de difusão regional. O modelo se mostrou consistente na maioria dos testes de robustez.

Além disso, foram calculados os efeitos diretos e indiretos baseando-se nos quatro tipos de clusters. Os valores indicam forte influência dos efeitos indiretos no efeito total da conectividade de um município, mostrando, assim, significativo efeito de transbordamento. Tais achados podem corroborar em levar o foco de aplicação de recursos e políticas públicas nas localidades de menor grau de conectividade, como os indicados no presente trabalho, assim considerando os efeitos vizinhança na implantação de infraestruturas de redes e de conectividade nos municípios brasileiros.

7. Referências bibliográficas

AGÊNCIA NACIONAL DE TELECOMUNICAÇÕES (Brasil). Painel de Acessos – Informações de acessos móveis e banda larga fixa. Disponível em: <https://informacoes.anatel.gov.br/paineis/acessos>. Acesso em: 7 maio 2025.

AGÊNCIA NACIONAL DE TELECOMUNICAÇÕES (Brasil). Painel de Infraestrutura – Cobertura móvel e rede de transporte (backhaul de fibra). Disponível em: <https://informacoes.anatel.gov.br/paineis/infraestrutura>. Acesso em: 7 maio 2025.

AGÊNCIA NACIONAL DE TELECOMUNICAÇÕES (Brasil). Painel de Outorga e Licenciamento – Estações do SMP. Disponível em: <https://informacoes.anatel.gov.br/paineis/outorga-e-licenciamento/estacoes-do-smp>. Acesso em: 7 maio 2025.

AGÊNCIA NACIONAL DE TELECOMUNICAÇÕES (Brasil). Relatório Metodológico do Índice Brasileiro de Conectividade – IBC 2024. Brasília: Anatel, 2025. Acesso em: 7 maio 2025

ALVES, Ana. Estatística Aplicada: Análise de Dados (Volume I — Análises). 2. ed. Aprender Estatística Fácil, 2024. 196 p.

ANSELIN, Luc. Spatial econometrics: methods and models. Dordrecht: Kluwer Academic Publishers, 1988

CONECTARAGRO. Painel do Indicador de Conectividade Rural (ICR). Disponível em: <https://www.conectaragro.com.br/indicadordeconectivaderural/>. Acesso em: 7 maio 2025.

FÁVERO, Luiz Paulo; BELFIORE, Patrícia. Manual de análise de dados: estatística e modelagem multivariada com Excel®, SPSS® e Stata®. 2. ed. Rio de Janeiro: Elsevier, 2017.

HAIR JR., Joseph F. et al. Análise multivariada de dados. 6. ed. Porto Alegre: Bookman, 2009.

IBGE – Instituto Brasileiro de Geografia e Estatística. Estimativas da população dos municípios brasileiros: 2023. Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/populacao/9103-estimativas-de-populacao.html>. Acesso em: 7 maio 2025.

LESAGE, James P.; PACE, R. Kelley. Introduction to spatial econometrics. Boca Raton: CRC Press, 2009

MALHOTRA, Naresh K. Pesquisa de marketing: uma orientação aplicada. 4. ed. Porto Alegre: Bookman, 2006. 572 p.

MILLO, Giovanni; PIRAS, Gianfranco, Splm: Spatial panel data models in R. *Journal of Statistical Software*, [s. l.], v. 47, n. 1, p. 1–38, 2012. DOI: <https://doi.org/10.18637/jss.v047.i01>

SEFFRIN, Rodolfo; ARAÚJO, Everton Coimbra de; BAZZI, Claudio Leones. Análise espacial de área aplicada à produtividade de soja na região oeste do Paraná utilizando o software R. *Revista Brasileira de Geomática*, Curitiba, v. 6, n. 1, p. 23–43, jan./mar. 2018. Disponível em: <https://periodicos.utfpr.edu.br/rbgeo>

SILVA, Elvis Magno da et al. Análise multivariada de cluster: uma ferramenta estatística para pesquisa em ciências sociais aplicadas. *International Journal of Scientific Management and Tourism*, Curitiba, v. 10, n. 5, p. 1–25, 2024. DOI: 10.55905/ijsmtv10n5-002. Disponível em: <https://www.ijsmt.com/>. Acesso em: 7 maio 2025.

8. Anexos

Link para download do arquivo com a planilha de dados dos municípios, incluído a classificação dos municípios por quadrantes e por clusters, além do IDS (1-IVS).

https://sei.anatel.gov.br/sei/modulos/pesquisa/md_pesq_documento_consulta_externa.php?HWH32bONvibUcMC3mewfUpIX7e-9fyZZC4iEjl2QHwXAoLCOrVZwNzRf5vR3YcCMWNZ4eCgQDLmVzIOFPcg7RpcrbjPE8XLmziu4OkKby5AcP5ODt5QBcJkkSNtuZs8

Link para download dos notebooks de com código em Python gerados. São, ao todo, três notebooks: i) um arquivo que gerou as estatísticas e análises gerais do IBC, ii) outro que gerou a análise de clusters e de quadrantes e iii) um que gerou o modelo econométrico¹⁶.

https://sei.anatel.gov.br/sei/modulos/pesquisa/md_pesq_documento_consulta_externa.php?HWH32bONvibUcMC3mewfUpIX7e-9fyZZC4iEjl2QHwXAoLCOrVZwNzRf5vR3YcCMWNZ4eCgQDLmVzIOFPcg7RgyT_f2LTPG2CeTQCdGdB-mVBmfCVFR6YGqDSHY7rV9Q

¹⁶Nos notebooks de análise de IBC e de cluster deve-se atentar para o nome do arquivo 'xlsx' baixado. Ao efetuar o comando de construir data frame o nome deve ser exato. Originalmente o nome do arquivo xlsx é "análise IBC v5" e "análise IBC v6" No notebook do modelo econométrico também deve-se ajustar o nome do arquivo 'xlsx' a ser baixado. Originalmente o nome do arquivo é 'df4_resultados'. 53



Siga a Anatel

