



UNIVERSIDADE DO ESTADO DE MATO GROSSO
CAMPUS DE SINOP
FACULDADE DE CIÊNCIAS EXATAS E TECNOLÓGICAS
MESTRADO PROFISSIONAL EM MATEMÁTICA EM REDE
NACIONAL - PROFMAT



MÁRCIO TORRES DE ALMEIDA

**UMA PROPOSTA DE ÍNDICE DE DESENVOLVIMENTO REGIONAL:
UM ESTUDO DE CASO**

SINOP

2018

MÁRCIO TORRES DE ALMEIDA

**UMA PROPOSTA DE ÍNDICE DE DESENVOLVIMENTO REGIONAL:
UM ESTUDO DE CASO**

Dissertação apresentada ao Programa de Mestrado Profissional em Matemática em Rede Nacional – PROFMAT, do departamento de Matemática da Universidade Estadual do Mato Grosso – UNEMAT, como requisito parcial para obtenção do grau de Mestre em Matemática.

Prof. Dr. Giovane Maia do Vale
Orientador

Prof. Me João Gabriel Ribeiro
Coorientador

SINOP
2018

A447u ALMEIDA, Márcio Torres de .
Uma Proposta de Índice de Desenvolvimento Regional: um
Estudo de Caso / Márcio Torres de Almeida - Sinop, 2018.
66 f.; 30 cm.(ilustrações) Il. color. (sim)

Trabalho de Conclusão de Curso
(Dissertação/Mestrado) - Curso de Pós-graduação Stricto Sensu
(Mestrado Profissional) Profmat, Faculdade de Ciências Exatas e
Tecnológicas, Câmpus de Sinop, Universidade do Estado de
Mato Grosso, 2018.
Orientador: Prof. Dr. Giovane Maia do Vale
Coorientador: Prof. Me João Gabriel Ribeiro

1. Índice de Desenvolvimento Regional. 2. Padronização. 3.
Distribuição Normal. 4. Método dos Mínimos Quadrado
Aparados. I. Márcio Torres de Almeida. II. Uma Proposta de
Índice de Desenvolvimento Regional: um Estudo de Caso: .
CDU 311



ESTADO DE MATO GROSSO
SECRETARIA DE ESTADO DE CIÊNCIA E TECNOLOGIA
UNIVERSIDADE DO ESTADO DE MATO GROSSO
CAMPUS UNIVERSITÁRIO DE SINOP
FACET – FACULDADE DE CIÊNCIAS EXATAS E TECNOLÓGICAS.
MESTRADO PROFISSIONAL EM MATEMÁTICA EM REDE NACIONAL
PROFMAT UNEMAT - SINOP

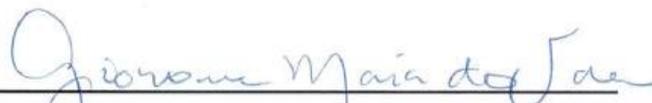


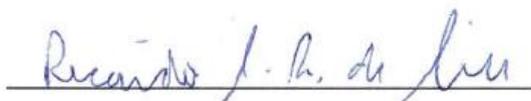
MARCIO TORRES DE ALMEIDA

“Uma proposta de índice de desenvolvimento regional: um estudo de caso”

Dissertação apresentada ao Programa de Mestrado Profissional em Matemática em Rede Nacional da Universidade do Estado de Mato Grosso - UNEMAT no *Campus* Universitário de Sinop, para obtenção do título de Mestre em Matemática.

Aprovado em: 22/06/2018


Prof. Dr. Giovanna Maia do Vale - UNEMAT


Prof. Dr. Ricardo S.R. da Silva – UNESP


Prof. Dr. Rogério Dias Dalla Riva - UNEMAT

SINOP – JUNHO - 2018



Programa de Mestrado Profissionalizante em Matemática em Rede Nacional
UNEMAT- Sinop Avenida dos Ingás, nº 3001 - Centro – CEP: 78.555-000 – Sinop–
MT.Tel./Fax: (66)9601-8925 – Cx. Postal: 680 – profmat-unemat@unemat-net.br



Dedico esse trabalho à minha esposa Katiúscia, pois seu amor e incentivo foram fundamentais para que eu chegasse até aqui, sempre acreditando que podíamos, ante aos momentos de incertezas, vencê-los. Sem dúvida, ela e suas orações, fizeram toda a diferença em nossas vidas.

Da mesma forma, esse trabalho é dedicado aos meus filhos, Maria Eduarda e Arthur Henrique, minhas sinceras desculpas nos momentos de ausência e irritável estresse. Espero ter sido exemplo e inspiração para a busca do sonho que todo homem deseja realizar. Suas vidas em minha vida aspiram por dias cada vez melhores.

AGRADECIMENTOS

À Deus, que nos dá a vida e a ânsia por sempre aprender mais.

À minha esposa Katiúscia, pelos momentos que estive ausente e aos meus filhos Maria Eduarda e Arthur Henrique, que espero fielmente, ter sido exemplo na busca perseverante desta conquista.

Aos meus Pais, Waldeth e Leolino os quais verdadeiramente e com amor torcem pelo meu sucesso e pelas minhas conquistas.

Às minhas manas queridas que incondicionalmente torcem pelo meu sucesso.

Ao Amigo e Parceiro, professor Marcelo Hoffmann perseverante nos milhares de quilômetros indispensáveis a concretização deste sonho. E, aos Amigos Antônio, Bruno e Rayan pelo companheirismo e pela humildade em compartilhar o conhecimento, atitudes essenciais ao sucesso desta formação.

À UNEMAT, por oportunizar esse Mestrado nesta região, de modo que eu pudesse ser beneficiado.

À SBM, CAPES e IMPA pela coordenação do PROFMAT e suporte financeiro.

A todos os professores envolvidos com o PROFMAT no Campus da UNEMAT em Sinop pela dedicação na arte de ensinar.

Aos professores Giovane Maia do Vale e João Gabriel Ribeiro, pelo apoio e orientação.

Aos amigos e companheiros de curso. Juntos, lutamos e vencemos.

“Talvez não tenhamos conseguido fazer o melhor, mas lutamos para que o melhor fosse feito. Não somos o que deveríamos ser, não somos o que iremos ser... mas Graças a Deus, não somos o que éramos”.

Martin Luther King

RESUMO

Quando se busca na literatura por resultados de estudos regionais, verifica-se que, frequentemente, o desenvolvimento dos municípios é citado. Porém, a maioria destes estudos aborda apenas elementos do desenvolvimento como, por exemplo, aspectos econômicos e/ou políticos, não trazendo uma modelagem matemática holística do desenvolvimento, que abarque simultaneamente várias variáveis e/ou indicadores. Nestes termos, o trabalho aqui apresentado traz a formulação de um índice de desenvolvimento regional baseado em 5 indicadores: área de município, população de município, densidade demográfica, produto interno bruto (PIB) e índice de desenvolvimento humano (IDH). O índice proposto, de formulação simples e concisa, consistiu da média não ponderada dos valores padronizados dos indicadores. Para fins de experimentação e validação do índice, foram tomados os valores dos indicadores de 33 municípios do entorno de Sinop – MT. Assim, tal conjunto foi particionado em subconjuntos de acordo com o bioma a que pertenciam: Amazônia, Cerrado-Amazônia e Cerrado. Logo, dois experimentos relativos aos biomas Amazônia e Cerrado-Amazônia foram efetuados. Estes experimentos consistiram da geração de um segundo índice de desenvolvimento, via Método dos Mínimos Quadrados Aparados, o qual foi comparado estatisticamente com o primeiro índice criado. A análise comparativa teve início com a ordenação hierárquica dos subconjuntos de dados, baseada nos valores calculados por ambos os índices. A comparação se deu ainda por meio de histogramas, gráficos boxplot, gráficos de confiança, mapas coroplédicos, gráficos de dispersão e pelo cálculo do coeficiente de correlação de Pearson. Além das ordenações dos subconjuntos de dados coincidirem, verificou-se por meio da análise comparativa que o índice criado é válido e se equiparou ao segundo índice, criado por meio de teorias consolidadas.

Palavras-chave: Índice de Desenvolvimento Regional, Padronização, Distribuição Normal, Método dos Mínimos Quadrado Aparados.

ABSTRACT

When one searches in the literature for results of regional studies, it is verified that, frequently, the development of the municipalities is quoted. However, most of these studies only address elements of development such as economic and / or political aspects, not bringing a holistic mathematical modeling of development, which simultaneously encompasses several variables and / or indicators. In these terms, the work presented here brings the formulation of a regional development index based on 5 indicators: area of municipality, population, demographic density, gross domestic product (GDP) and human development index (HDI). The proposed index, with a simple and concise formulation, consisted of the unweighted mean of the standardized values of the indicators. For purposes of experimentation and validation of the index, the values of the indicators of 33 municipalities in Sinop - MT were taken. Thus, this set was partitioned into subsets according to the biome to which they belonged: Amazon, Cerrado-Amazônia and Cerrado. Therefore, two experiments related to the Amazon and Cerrado-Amazônia biomes were carried out. These experiments consisted of the generation of a second development index, using the Trimmed Least Squares Method, which was compared statistically with the first index created. The comparative analysis began with the hierarchical ordering of the data subsets, based on the values calculated by both indices. The comparison was also made by means of histograms, boxplot graphs, confidence charts, choropleth charts, scatter plots and by the Pearson correlation coefficient. In addition to the ordering of the data subsets, it was verified through the comparative analysis that the created index is valid and equated to the second index, created through consolidated theories.

Keywords: Regional Development Index, Standardization, Normal Distribution, Least Trimmed Squares Method.

LISTA DE FIGURAS

Figura 1: Biomas do Estado de Mato Grosso.....	15
Figura 2: Localização Territorial dos Projetos de Colonização em Mato Grosso.....	26
Figura 3: Gráfico da curva de Gauss.....	35
Figura 4: Áreas sob a curva Normal em termos da Média (μ) e do Desvio Padrão (σ).	35
Figura 5: Mapa dos Municípios Segundo Seus Respectivos Biomas.	40
Figura 6: Histogramas dos valores dos índices I e \hat{I} para os municípios do bioma Amazônia.	51
Figura 7: Gráfico do tipo boxplot dos Índices I e \hat{I} para os Municípios do Bioma Amazônico	51
Figura 8: Gráfico dos Intervalos de Confiança dos Índices I e \hat{I} , com erro de 1%, para os Municípios do Bioma Amazônia	52
Figura 9: Mapas dos índices I e \hat{I} criados: Municípios do Bioma Amazônia.	53
Figura 10: Gráfico de dispersão entre os Índices I e \hat{I} – Municípios do Bioma Amazônia	54
Figura 11: Histogramas dos Índices I e \hat{I} para os Municípios do Bioma Cerrado-Amazônia.	57
Figura 12: Gráfico do tipo boxplot de I e \hat{I} - Municípios do Bioma Cerrado-Amazônia.....	57
Figura 13: Gráfico dos Intervalos de Confiança dos Índices I e \hat{I} , com erro de 1%, para os Municípios do Bioma Cerrado-Amazônia.....	58
Figura 14: Mapas dos Índices I e \hat{I} criados - Municípios do Bioma Cerrado-Amazônia.....	59
Figura 15: Gráfico de Dispersão entre os Índices - Municípios do Bioma Cerrado-Amazônia	60
Figura 16: Mapa do Município de Santa Rita do Trivelato e seus Indicadores	61

LISTA DE EQUAÇÕES

Equação 1: Média Populacional Simples	30
Equação 2: Média Amostral Simples	30
Equação 3: Elemento Mediano para uma Série Ímpar de Dados	31
Equação 4: Elemento Mediano para uma Série Par de Dados	31
Equação 5: Amplitude Total	32
Equação 6: Variância Populacional	33
Equação 7: Variância Amostral.....	33
Equação 8: Desvio Padrão Populacional	33
Equação 9: Desvio Padrão Amostral	34
Equação 10: Função Densidade de Probabilidade.....	35
Equação 11: Padronização de Varáveis	36
Equação 12: Soma dos Quadrados dos Resíduos (e^2)	37
Equação 13: Normalização dos Dados	42
Equação 14: Índice Proposto	43
Equação 15: Fórmula funcional por LTS.....	43
Equação 16: Cálculo da Quantidade de Classes	45
Equação 17: Cálculo do Intervalo de Classe	45
Equação 18: Cálculo do Intervalo de Confiança	45
Equação 19: Correlação de Pearson	46
Equação 20: Teste de Confiança.....	46
Equação 21: Fórmula/Índice Funcional para o Bioma Amazônia.....	50
Equação 22: Fórmula/Índice Funcional para o Bioma Cerrado-Amazônia	56

SUMÁRIO

LISTA DE FIGURAS	10
LISTA DE EQUAÇÕES	11
1. INTRODUÇÃO	14
1.1 Considerações Iniciais	14
1.2 Desenvolvimento Regional	18
1.3 Trabalhos Relacionados	19
1.4 Objetivos:	21
1.4.1 Objetivo Geral	21
1.4.2 Objetivos Específicos	21
1.5 Estrutura do Trabalho	22
2. FUNDAMENTAÇÃO TEÓRICA	23
2.1 Contexto Histórico	23
2.2 Tópicos de Estatística	29
2.2.1 Medidas de Posição	30
2.2.1.1 Média Populacional	30
2.2.1.2 Média Amostral	30
2.2.1.3 Moda	30
2.2.1.4 Mediana	31
2.2.2 Medidas de Dispersão	31
2.2.2.1 Amplitude Total	32
2.2.2.2 Variância Populacional	33
2.2.2.3 Variância Amostral	33
2.2.2.4 Desvio-Padrão Populacional	33
2.2.2.5 Desvio-Padrão Amostral	33
2.2.3 A Distribuição Normal e a Padronização de Variáveis	34
2.3 O Método dos Mínimos Quadrados Aparados	36

3. METODOLOGIA	39
3.1 Dados	39
4.2.2 Área de transição Cerrado-Amazônia.....	55
4.2.3 Bioma Cerrado.....	60
5. CONCLUSÕES E RECOMENDAÇÕES.....	62
5.1 Conclusões	62
5.2 Recomendações	64
REFERÊNCIAS BIBLIOGRÁFICAS	65

1. INTRODUÇÃO

1.1 Considerações Iniciais

O estado de Mato Grosso, localizado na Região Centro-Oeste do Brasil e integrado à Amazônia Legal, constitui importante fronteira agrícola que vem consolidando-se como área de produção agroindustrial. Dentre as suas principais *commodities* estão a soja, o milho, o algodão e a carne, que advém de modernos sistemas de produção viabilizados por intenso investimento de capital (IBGE – Produção Agrícola Municipal). A combinação de recursos naturais abundantes, empreendedorismo, tecnologia, investimento de capital e políticas públicas contribuíram para um vigoroso processo de crescimento econômico e de desenvolvimento social no estado. Estes elementos lhe valeram importante destaque em nível nacional e internacional como grande produtor agrícola. Porém, se por um lado o estado de Mato Grosso reflete uma nova etapa no contexto agrícola do Brasil, marcado pela dinamicidade e desenvolvimento, por outro, em alguns momentos sobressaem as desigualdades, marcadas pela discrepância econômica e social. Em parte, esse antagonismo deriva das relações entre a população e os elementos naturais, tais como, o solo, o clima, a fauna e a flora. Ou seja, o estado possui regiões com excelente potencial para a produção, e outras regiões não tão favorecidas e que, por esta razão, não comportam atividades intensivas de produção. Algumas regiões ainda compensam a inadequação à produção agrícola com o turismo ou a pecuária, como é o caso do Pantanal Mato-grossense.

Um aspecto importante a ser ressaltado e que certamente influencia no desenvolvimento socioeconômico das mesorregiões do estado, compreende o grande mosaico ecológico no qual o estado de Mato Grosso está inserido. O Mato Grosso possui 903,378 mil km² de área, com 478 mil km² cobertos pelo Bioma Amazônia, 367 mil km² cobertos pelo Bioma Cerrado e 58 mil km² pelo Bioma Pantanal. Seus 141 municípios estão integrados a 22 microrregiões e cinco mesorregiões, a saber: mesorregião norte, mesorregião nordeste, mesorregião sudeste, mesorregião centro-sul e mesorregião sudoeste. Do seu total de municípios, 74 tem a predominância do Bioma Amazônico; 62 do Bioma Cerrado e em apenas cinco municípios há a predominância do Bioma Pantanal (SEMA – MT).

Considerando que este trabalho é voltado à análise de alguns municípios mato-grossenses e seus respectivos biomas, a priori, levantou-se que há 42 municípios nos quais se verifica somente o bioma amazônico e em 44 municípios há a presença concomitante dos biomas Amazônia e Cerrado. Ou seja, estes municípios encontram-se na área de transição, com

ecótonos e encraves. E ainda, de forma principal ou secundária, o bioma amazônico está presente em 93 municípios e o Bioma Cerrado em 92 municípios.

Em meio aos estudos preliminares da pesquisa, averiguou-se que em Mato Grosso verifica-se um processo complexo de interação do homem com a natureza, com especificidades que remetem a um entendimento do funcionamento da dinâmica ecossistêmica regional e local, a fim de garantir um sistema extrativo e produtivo capaz de sustentar a população residente e gerar excedentes exportáveis. Os biomas são então suportes que comportam os métodos e técnicas de apropriação de seus recursos, que podem constranger ou estimular a expansão socioeconômica diante da capacidade técnica e institucional disponível. O mapa presente na Figura 1 mostra a distribuição espacial dos biomas no espaço territorial do estado de Mato Grosso, conforme o Portal Matogrosso Brasil. No referido mapa verifica-se que no extremo norte do estado encontra-se o bioma Amazônia e, no extremo sul, o bioma Pantanal. Na grande porção central do estado encontram-se, quase que igualmente distribuídos, o bioma Cerrado e a área de transição Cerrado-Amazônia.

Figura 1: Biomas do Estado de Mato Grosso



Fonte: <http://www.matogrossobrasil.com.br/ecossistema.asp>

Quanto à densidade demográfica, o estado de Mato Grosso, que se compara a outros estados da federação, ocupa a 25ª posição dentre as 27 unidades que compõem o Estado Brasileiro. Com escore de aproximadamente 3,57 hab./km², verifica-se um disparate quando se compara este valor à realidade dos municípios que o integram. Por exemplo, Cuiabá registra

densidade demográfica de 157,66 hab./km², Várzea Grande com 240,98 hab./km² e Rondonópolis com 47 hab./km². Por outro lado, em oposição a esta realidade encontra-se, por exemplo, Araguainha com ínfimos 1,59 hab./km².

Assim, conclui-se que Mato Grosso seja marcado por desigualdades, pois, apesar da existência de municípios altamente povoados como, por exemplo, Sinop, cidade polo com 28,69 hab./km², verificam-se outros com baixa densidade demográfica. E mais, o próprio estado, por sua grande área territorial (3º maior estado da federação), não possui densidade demográfica expressiva.

Particularmente, no tocante à mesorregião norte, na qual se encontra Sinop, acredita-se que, pela sua grande extensão, destinada à agricultura, tenham surgidos municípios com menor densidade populacional, mas igualmente desenvolvidos, que contribuem para o destaque da região, por exemplo, Sorriso que apresenta 7,13 hab./km² ou Lucas do Rio Verde, com 12,43 hab./km².

Considerando a economia do Estado de Mato Grosso, verifica-se que esta, se caracteriza por um Produto Interno Bruto estimado em R\$ 101,23 bilhões no ano de 2014, em contraposição ao período anterior que fora de R\$ 89,21 bilhões. Em termos de crescimento real, o indicador de volume aponta para um crescimento de 4,4% em relação à 2013 e um volume acumulado de 26,7%, o que representa um crescimento médio estimado de 6,1% ao ano no período de 2010 a 2014 (SEPLAN – MT).

Tal realidade repercute na renda da população. Assim, com uma população estadual residente estimada em no ano de 2014 em 3.224.357 milhões de pessoas, segundo o Instituto Brasileiro de Geografia e Estatística (IBGE), o PIB per capita alcançou em 2014, a cifra de R\$ 31.396,81, com um aumento de 11,99% se considerados os R\$ 28.036,00 referentes ao ano de 2013. Porém, o que se percebe é que, apesar destes números serem animadores, tais benesses não atingem grande parte da população do estado.

Da composição do Produto Interno Bruto (PIB) mato-grossense, 10 (dez) municípios somam 54,6% do montante de R\$ 101,23 bilhões. Juntos representam R\$ 55,29 bilhões de tudo que se produziu em 2014. Segundo dados do Instituto Brasileiro de Geografia e Estatística (IBGE), Cuiabá registrou um PIB de R\$ 20,52 bilhões no ano em que o Brasil foi sede da Copa do Mundo de Futebol. Já os principais municípios do agronegócio mato-grossenses registraram R\$ 21,21 bilhões no mesmo período (IBGE, 2018). Os montantes dizem respeito às informações de PIB dos Municípios no ano de 2014.

Constata-se que mais da metade do PIB de Mato Grosso, conforme o IBGE (2018), advém das cidades de Cuiabá (R\$ 20,52 bilhões), Rondonópolis (R\$ 7,64 bilhões), Várzea

Grande (R\$ 5,90 bilhões), Sorriso (R\$ 4,43 bilhões), Sinop (R\$ 4,28 bilhões), Primavera do Leste (R\$ 2,94 bilhões), Lucas do Rio Verde (R\$ 2,75 bilhões), Nova Mutum (R\$ 2,50 bilhões), Tangará da Serra (R\$ 2,26 bilhões) e Campo Novo do Parecis (R\$ 2,02 bilhões). Dentre os municípios acima citados, verifica-se que Cuiabá, Várzea Grande e Rondonópolis possuem economias diversificadas, aliadas à produção agroindustrial. Dos 10 municípios destacados acima, 05 (cinco) encontram-se na mesorregião norte.

Considerando a qualidade de vida do cidadão mato-grossense, especifica-se que, no ranking nacional do Índice de Desenvolvimento Humano (IDH), Mato Grosso ocupa o 11º lugar, segundo o Atlas 2013, do Programa das Nações Unidas (PNUD/2010). O índice avalia a qualidade de vida (saúde e educação) e o desenvolvimento econômico (renda) das populações. O índice varia de uma escala de 0 a 1, de modo que valores mais próximos de 1, representam maior desenvolvimento humano da localidade. Em 2010, o IDH de Mato Grosso (0,725) foi avaliado como alto, sendo Cuiabá a cidade com melhor índice do Estado (0,785).

Confirmando a tendência de desenvolvimento dos municípios de economia agroindustrial e essencialmente agrícolas, tem-se como destaque os valores de IDH de: Lucas do Rio Verde (0,768), Nova Mutum (0,758), Rondonópolis (0,755), Sinop (0,754) e Primavera do Leste (0,752), entre outros. Tais valores, indicam os excelentes resultados provenientes da organização e da geração de melhorias alcançadas com a colonização. Assim, aqueles que migraram para o estado de Mato Grosso, principalmente para os Municípios do Médio Norte e Norte Mato-grossense, em busca da realização de seus sonhos e melhores condições de vida, possuem uma alta probabilidade de encontrar condições favoráveis aos seus objetivos.

Importante notar que, se o IDH de uma população aumenta em um determinado período histórico, entende-se que sua qualidade de vida atingiu padrões gerais mais exigentes e mais satisfatórios. Logo, investir em melhorias nos produtos e serviços para atender de modo diferenciado a esta população pode ser vantajoso em qualquer ramo de atividade.

Do acima exposto, depreende-se que o estado de Mato Grosso venha evoluindo a passos largos e que tal desenvolvimento seja reflexo do desenvolvimento de seus municípios, que sustentam a cadeia produtiva do estado. Porém, tais observações, apesar de verdadeiras, não expressam objetivamente este desenvolvimento e não permitem uma análise quantitativa da realidade. Assim, buscando quantificar e gerar um índice numérico que expresse esta realidade e permita com que análises estatísticas de sondagem sejam feitas, tem-se o trabalho aqui apresentado. Ou seja, tal trabalho consistiu da criação de um índice de desenvolvimento regional baseado em elementos quantificáveis (indicadores), como, área de município, população de município, densidade demográfica, produto interno bruto (PIB) e índice de

desenvolvimento humano (IDH). A fim de validar estatisticamente o índice, foram utilizados dados de Sinop e de alguns municípios de seu entorno, considerando agrupamentos de municípios de acordo com os biomas a que pertencem.

1.2 Desenvolvimento Regional

O conceito de desenvolvimento, há algum tempo atrás, era utilizado como sinônimo de crescimento econômico. Porém, em decorrência de seu significado e da sua abrangência a expressão deixou de ser utilizada apenas no âmbito da Economia e passou a ser interpretado de forma complementar, ou de forma diferenciada, em outras áreas do conhecimento como, por exemplo, na Sociologia, na Administração, na Geografia, na História etc.

O conceito de desenvolvimento deve abranger múltiplos aspectos, como: aspectos sociais, culturais, econômicos e espaciais. O desenvolvimento regional segue a mesma linha lógica, mas abrange, além do crescimento econômico, fatores culturais, ambientais e também políticos. Isto se deve ao fato de que tais aspectos contribuem incisivamente para a melhoria da qualidade de vida e bem-estar da sociedade.

O desenvolvimento econômico é visto então como sendo o crescimento econômico somado à melhoria do padrão de vida da população e às alterações fundamentais na estrutura econômica e social de uma localidade. Tais, elementos deveriam possibilitar a distribuição mais equânime das riquezas produzidas, a fim de reduzir as desigualdades sociais aberrantes. A conquista de tais efeitos benéficos implica em efetuar profundas transformações estruturais relacionadas com uma concepção coletiva, composta por conhecimentos que permitam hierarquizar problemas e soluções e facilitar a coordenação/convivência entre os atores sociais. Se tal configuração for alcançada, então a distribuição de poder econômico e social seriam facilitadas, o que constituiria uma nova economia política. Desta maneira, a geração de desenvolvimento econômico tem como condição a mudança na distribuição do poder político e, conseqüentemente, associa a produção de recursos à distribuição mais justa, em função da força política dos atores sociais.

Conclui-se que, o desenvolvimento seja um processo social holístico, que envolva aspectos econômicos, políticos, sociais e culturais. A definição de “desenvolvimento” depende de valores historicamente construídos, inerentes a cada sociedade. Porém, independentemente das diferenças particulares das definições, subentende-se sempre a conquista de padrões de vida mais elevados acessíveis à maioria da população. Segundo esta visão, o conceito de desenvolvimento pode ser até antagônico à ideia de progresso econômico, pois sua concepção vai além da oferta de bens e serviços resultantes do aumento de produtividade.

Nestes termos, o desenvolvimento passa a ser visto como um processo de mudanças socioeconômicas. Logo, chega-se à necessidade e ao desafio de mensurar tal desenvolvimento. Assim, historicamente, considerando o aspecto econômico, conjecturou-se o Produto Interno Bruto – PIB, que é um indicador voltado à Economia e que é utilizado como referencial quantitativo do desenvolvimento de uma nação ou região.

Seguindo a mesma vertente e considerando a discrepância entre diferentes aspectos da qualidade de vida, surge o índice de Desenvolvimento Humano – IDH. Este índice, proposto por um Programa das Nações Unidas, nasceu da amostragem de indicadores da qualidade de vida de mais de 1 milhão de pessoas, em mais de 130 países. Assim, partir de 1990 a ONU – Organização da Nações Unidas - apresentou o IDH especificando que este se destinava à quantificação do desenvolvimento humano, sendo capaz de mensurar aspectos relacionados à qualidade de vida.

Dada a complexidade do fenômeno, acredita-se que não há um indicador ou índice suficientemente abrangente para a mensuração do desenvolvimento regional. Há, na verdade, aspectos distintos que podem ou não ser quantificados ou considerados. No entanto, a fim de se poder desenvolver políticas positivas de desenvolvimento, pode-se optar por índices, variáveis e/ou indicadores, que permitam analisar, mesmo que de modo imperfeito, o sucesso ou insucesso de uma nação ou região, pois tais índices podem, de alguma forma, evidenciar pontos fortes e pontos fracos do desenvolvimento de uma sociedade e, conseqüentemente, da estratégia política de desenvolvimento empregada.

Nestes termos, neste trabalho buscou-se desenvolver matematicamente e validar estatisticamente um índice de desenvolvimento regional. Tal índice tem por base 05 (cinco) indicadores (Área do município, População, Densidade Demográfica, PIB e IDH) relativos à Sinop e a alguns municípios da mesorregião norte mato-grossense. Os experimentos para a validação do índice criado foram efetuados considerando-se dois grupos de municípios, os quais foram agregados segundo seus respectivos biomas. Tal opção metodológica se deu por se entender que o desenvolvimento econômico de um município esteja intrinsecamente ligado ao bioma a que pertence. Assim, a divisão em dois grupos visou, além de gerar redundância na experimentação, a obtenção de classificações hierárquicas dos municípios baseadas no índice proposto, considerando afinidades entre os municípios em cada grupo.

1.3 Trabalhos Relacionados

Quando se efetua uma busca por relatos científicos (artigos, dissertações, teses etc..) que guardem alguma semelhança com o tema desenvolvido no trabalho de pesquisa aqui descrito,

pode-se encontrar inúmeros relatos relacionados. Tem-se que a maioria deles trata especificamente de aspectos econômicos e/ou políticos acerca da ocupação, colonização e desenvolvimento da mesorregião norte mato-grossense. Tais aspectos também são abordados aqui, porém estão simultaneamente abarcados em um mesmo índice. Em outros trabalhos encontrados, os autores enfatizam apenas aspectos particulares que contribuíram com a fixação dos inúmeros migrantes nos municípios estudados e em outros municípios do Brasil. Assim, a fim de fornecer uma contextualização ao leitor sobre o tema e gerar uma possibilidade de comparação entre o trabalho aqui apresentado e o que vem sendo cientificamente produzido, são apresentados alguns trabalhos relacionados.

Jorge *et al.* (2010), em seu trabalho, elabora um Índice de Desenvolvimento da Gestão Municipal (IDGM) para o município sergipano de Itabaiana. Itabaiana é um município de médio porte e, segundo o autor, no Estado de Sergipe não existe nenhuma cidade que se utilize de um índice para avaliar o grau de desenvolvimento em relação à educação, saneamento e saúde. O município em questão possui significativa importância na economia estadual, além de configurar como um polo de ensino superior, com uma população de aproximadamente 84 mil habitantes (JORGE *et al.*, 2010).

Lopes (2016) enfatiza em sua Dissertação “Fatores Determinantes no Processo de Desenvolvimento Endógeno de um Território: Um olhar para o Município de Ijuí a partir dos atores da tríplice hélice” fatores que podem ou não gerar desenvolvimento para uma localidade. Embora a pesquisa não objetive o desenvolvimento de um índice, buscou-se constituir num modelo analítico da articulação de um conjunto de fatores e indicadores de um determinado território que, em tese, poderiam influenciar ou inibir seu progresso, seja em termos de taxas de crescimento econômico, seja em termos da qualidade de vida da população. Lopes (2016) especifica que, Ijuí se destaca pela diversidade de colônias de imigrantes, visto que já em 1890 contava com 18 etnias, caracterizando-se como “Terra das Culturas Diversificadas”.

Arruda (2010) vale-se de Análise Multivariada, aplicada ao Índice de Desenvolvimento Sustentável e Agronegócio, para entender as inter-relações entre sustentabilidade ambiental, econômica e social dos municípios do Estado de Goiás, no período entre os anos de 1995 e 2005. Segundo a autora, o Estado de Goiás, diante do modelo de modernização conservadora, passou por transformações e mudanças calcadas no favorecimento de algumas formas de produção em detrimento de outras. A gestão do estado preocupou-se prioritariamente com a expansão da fronteira agrícola e com a ocupação de áreas para o aumento da produção para a exportação, deixando outros aspectos do desenvolvimento em segundo plano.

Braga *et al.*, (2004) apresenta metodologia de construção de índices de sustentabilidade e desenvolvimento local, baseada em Análise Multivariada, para os municípios da bacia do rio Piracicaba em Minas Gerais. Segundo Braga *et al.*, (2004), a formulação de uma metodologia baseada em índices/indicadores focados na caracterização de sistemas hídricos, funções de utilidade e de preferências, e baseada na Teoria de Apoio à Decisão, pode contribuir para a melhoria da gestão da realidade vigente.

Outros trabalhos pesquisados e relacionados ao tema aqui abordado, apontam isoladamente para fatores políticos, sociais e econômicos que contribuíram para as transformações socioculturais e espaciais no norte do estado de Mato Grosso. Porém, tais trabalhos não estabelecem índices ou efetuam a matematização do problema. Eles expressam, em sua quase totalidade, elementos condicionantes da evolução socioeconômica de Mato Grosso e da mesorregião norte mato-grossense, apenas citando fatores que contribuíram no processo de ocupação e urbanização da região.

1.4 Objetivos:

1.4.1 Objetivo Geral

O objetivo geral deste trabalho foi desenvolver um índice de desenvolvimento regional baseado em diversos indicadores sociais, econômicos e geográficos, e verificar estatisticamente a sua validade.

1.4.2 Objetivos Específicos

A fim de que o objetivo geral proposto fosse alcançado, os seguintes objetivos específicos estabelecidos e atingidos:

- Angariar, por meio de pesquisa bibliográfica, conhecimentos relativos ao tema da pesquisa e que subsidiem o seu desenvolvimento.
- Coletar, junto ao IBGE, dados relativos aos indicadores Área do município, População, Densidade Demográfica, PIB e IDH. Tais indicadores foram escolhidos por expressarem não apenas o desenvolvimento econômico dos municípios em estudo, mas também a qualidade de vida da população.
- Desenvolver um índice de desenvolvimento regional inédito e de cunho estatístico que permita a mensuração do desenvolvimento regional com base nos indicadores escolhidos;
- Desenvolver um índice de desenvolvimento regional com base em teorias cientificamente já aceitas (Método dos Mínimos Quadrados Aparentados) de modo a, por meio de comparação estatística, validar o índice inédito criado;

- Efetuar experimentos com os índices de desenvolvimento regional criados, considerando os grupos de municípios escolhidos, a fim de validar estatisticamente o índice inédito, que é o cerne deste trabalho;
- Ao final da pesquisa, prover, se possível, os gestores públicos interessados, das diferentes esferas de poder, de um índice de desenvolvimento confiável que permita a quantificação e análise do desenvolvimento regional, com base em indicadores já conhecidos. Espera-se que a metodologia aqui apresentada propicie análises, avaliações e tomadas de decisão quanto ao desenvolvimento socioeconômico local, de forma a embasar o desenvolvimento e a aplicação de políticas públicas de alcance local e regional.

1.5 Estrutura do Trabalho

Este trabalho foi dividido em 5 capítulos. O capítulo 1 contém a introdução do trabalho, destinada a gerar uma breve contextualização ao leitor e fornecer uma ideia básica acerca do trabalho, bem como, expressar os principais objetivos da pesquisa.

O capítulo 2 contém os fundamentos teóricos mínimos necessários à compreensão da metodologia elaborada e executada. Neste capítulo é abordada de forma sucinta a história de Mato Grosso e são apresentados aspectos relacionados ao desenvolvimento socioeconômico da mesorregião norte mato-grossense, que é foco da pesquisa realizada. Ainda neste capítulo, são dadas as definições/características relacionadas às estatísticas que fundamentaram a metodologia deste trabalho, tais como: Distribuição Normal, Medidas de Posição, Medidas de Dispersão e Método dos Mínimos Quadrados Aparados – *LST*.

O capítulo 3 destinou-se à apresentação da metodologia desenvolvida para que os objetivos propostos fossem alcançados. Assim, em um primeiro momento, se explana sobre os dados de entrada considerados e coletados. A partir daí, descreve-se a criação do índice proposto e as formas de avaliação empregadas.

O capítulo 4 contém a apresentação dos resultados decorrentes da aplicação do índice criado e as respectivas análises estatísticas relativas à sua validação, considerando que, os experimentos estão divididos de acordo com o bioma aos quais pertencem os municípios considerados.

Por fim, o capítulo 5 destina-se à apresentação das principais conclusões depreendidas da realização do trabalho aqui exposto, bem como as principais recomendações para a execução de trabalhos futuros.

2. FUNDAMENTAÇÃO TEÓRICA

Os conteúdos apresentados neste capítulo são o resultado de pesquisas em livros, artigos, dissertações, teses e sites. Tais conteúdos versam sobre a história da ocupação e colonização do Estado de Mato Grosso, em especial da mesorregião norte mato-grossense, e sobre definições e conceitos estatísticos/matemáticos que embasam o desenvolvimento metodológico.

Assim, na Seção 2.1 estão presentes os principais elementos que caracterizaram o movimento de ocupação do Estado de Mato Grosso, bem como, da mesorregião norte mato-grossense. Os conceitos estatísticos necessários ao desenvolvimento da pesquisa aqui relatada são apresentados na Seção 2.2. Na seção 2.3, tem-se a apresentação do Método dos Mínimos Quadrados Aparados ou *Least Trimmed Squares – LTS*, que consiste de uma variação do Método dos Mínimos Quadrados – MMQ tradicional.

2.1 Contexto Histórico

A ocupação parcial de Mato Grosso começou por volta do século XVIII, mais especificamente na porção sul do estado, na região cujo núcleo é hoje a cidade de Cuiabá. Inicialmente tal região era acessada por hidrovias pantaneiras, mas, atualmente, o modal mais frequente são as rodovias. Posteriormente, na região do Pantanal, surgiram duas importantes ocupações que originaram os municípios de Cáceres e de Poconé. Na região oeste, o expoente do início da colonização foi a cidade de Vila Bela da Santíssima Trindade. A princípio, a ocupação do estado pouco contribuiu para a geração de uma estrutura produtiva que alavancasse a região. Quando a ocupação Portuguesa, de 1719 até 1796, entrou em colapso, o ouro subsidiou uma estrutura que fortalecia principalmente a fronteira. Posteriormente, até 1870, em virtude do definimento aurífero, do decaimento da navegação e do inexpressivo comércio com a cidade de Belém, a província de Mato Grosso permaneceu isolada e empobrecida, mantendo-se basicamente da produção de subsistência e dos aportes do governo federal.

Com o restabelecimento da navegação no rio Paraguai, após a guerra do Paraguai (1870), novos fatores incrementaram a economia de Mato Grosso, dentre estes: a industrialização da carne bovina, a extração de poaia e a extração de borracha e erva-mate. Tais atividades propiciaram um importante momento para o estado, pois promoveram o acúmulo de riquezas, principalmente pela exportação de produtos extrativos. Porém, no início do século XX, a concorrência comercial da borracha inglesa e a exaustão das reservas de poaia do estado desestruturaram esses dois sistemas extrativos. Estes fatos acabaram por desestruturar também os empreendimentos voltados à erva-mate, em 1930. Em contrapartida, mesmo sem a

implantação de frigoríficos no estado, a exportação de bovinos, que eram transportados pela ferrovia inaugurada em 1915 e que ligava Campo Grande a São Paulo, contribuiu para com a economia de Mato Grosso e fez com que o estado se consolidasse como grande fornecedor de proteína animal até o final da década de 1980.

Com a Política de Integração Nacional, quando foram construídas rodovias como a Transamazônica e a Cuiabá-Santarém, passou a ocorrer a interiorização do desenvolvimento da agricultura. Conseqüentemente, o processo de absorção de excedentes populacionais, iniciado pelo governo de Getúlio Vargas em 1930 e reforçada pelo governo Juscelino Kubistchek em 1950, foi impulsionado. Neste contexto, Mato Grosso foi amplamente beneficiado por programas que aumentaram a oferta de suas terras. Estes programas atraíram milhares de famílias oriundas principalmente da região sul do Brasil, movidas pelo desejo de uma vida melhor. Os programas de colonização instituídos naquela época foram: PROTERRA, POLAMAZONIA, POLOCENTRO, POLONOROESTE e PRODOESTE.

De modo geral, os programas envidavam esforços que compreendiam: incentivos à privatização de terras públicas, a operacionalização de projetos de colonização pública e privada, a construção de rodovias, a adoção de estratégias de modernização agrícola, a implantação de sistemas de comunicação e a destinação de incentivos à migração. Tais ações ocorreram, pois, até o fim da década de 1970, a população e a produção do estado estavam fortemente concentradas na região centro/sul ocupando inicialmente regiões de Cerrado e Pantanal. Logo, houve o intento de homogeneizar o desenvolvimento no estado, fazendo sua região norte progredir.

Com o próspero desenvolvimento da região norte do Estado, ainda na década de 1970, a ideia de separação entre as regiões norte e sul do estado ganhou força. Esta tendência se fortaleceu principalmente no governo Geisel, em 1979, que estava interessado na integração nacional. Cabe ressaltar que, a proposta separatista vinha sendo articulada desde a década de 1950 e já repercutia nacionalmente. Seu lema era “Dividir para Multiplicar”. Assim, no dia 11 de outubro de 1977 o presidente Geisel assinou a lei que dividia o estado de Mato Grosso (Lei Complementar nº 31), por meio da criação do estado do Mato Grosso do Sul.

A partir da divisão, as projeções pessimistas que preconizavam que o estado de Mato Grosso iria estagnar diante do cenário vigente da época não se concretizaram. Pelo contrário, ocorreu um processo de avanço da fronteira agrícola aliado à criação e desenvolvimento de municípios extremamente promissores, os quais contribuíram consideravelmente para a riqueza e desenvolvimento do estado.

Assim, Mato Grosso, em oposição à expectativa de fracasso após a separação em 1977, teve seu desenvolvimento socioeconômico acelerado a partir da década de 1980. Aportaram então no estado empresas agroindustriais nacionais e internacionais que se estabeleceram, principalmente na região centro/sul, sudeste e médio/norte. Esse desenvolvimento ocorreu mediante fomentos e investimentos do governo em infraestrutura, os quais modernizaram as vias de transporte, a base energética e as telecomunicações.

As rodovias Transamazônica e Cuiabá-Santarém, estabeleceram vários núcleos de colonização oficiais ao longo de seu percurso. Estes núcleos foram implantados pelo INCRA – Instituto Nacional de Colonização e Reforma Agrária e, além dos núcleos, vários outros projetos agropecuários extrativos na Região foram viabilizados mediante a concessão de incentivos fiscais e créditos a grandes empresas nacionais e multinacionais. Essas empresas ganharam impulso por estarem aliadas à colonização oficial, principalmente pela implantação privilegiada de suas instalações ao longo da Rodovia Cuiabá-Santarém/BR-163, ainda no Estado de Mato Grosso (HESPANHOL, 2000).

Com a abertura da Rodovia Cuiabá-Santarém, BR-163, surgiram em Mato Grosso muitos projetos de colonização organizados por empreendimentos particulares que mobilizaram grande contingente de migrantes principalmente do Sul do País e também de outras regiões brasileiras. Estes projetos deram origem a vários municípios na porção Médio – Norte do Estado, tais como: Sinop, Sorriso, Nova Mutum, Cláudia, entre outras.

A criação destas cidades, a partir das medidas governamentais e de projetos particulares de colonização, foi de fundamental importância para o desenvolvimento regional, conforme afirma Souza (2002, p. 30):

Estes municípios têm papel preponderante na ocupação, povoamento e desenvolvimento para a mesorregião norte mato-grossense: Importante papel do Governo Federal atraindo, por um lado, o capital privado (Colonizadoras) para a região, através de incentivos fiscais e financeiros, e, por outro, estabelecendo as condições básicas de infraestrutura; também há aspectos externos à região, como elementos de dinamização e de modernização econômica, social e política; e a ocupação econômica da região de forma integrada à economia nacional.

O mapa da Figura 2 mostra a distribuição espacial dos projetos de colonização no estado de Mato Grosso e que deram origem a muitos dos municípios que hoje encontram-se consolidados.

bacias de mão de obra local. Se os objetivos estratégicos de ocupação são em parte alcançados, o controle certamente não, acentuando os conflitos.

A construção da Rodovia Cuiabá-Santarém (BR-163) realmente efetivou o desenvolvimento de cidades e sub-regiões no Norte Mato-grossense. Contudo, sem o seu asfaltamento, por inúmeras vezes faltaram produtos de necessidade básica para a população residente nos municípios, conforme o relato de seus primeiros moradores. Santos (2007, p. 15) destaca que:

O movimento crescente de veículos, principalmente de caminhões [...] deixaram a rodovia em péssimas condições, principalmente na época das chuvas, quando praticamente ficava intransitável, ocasionando o isolamento e a falta de alimentos e combustíveis na região, como ocorreu no final de 1979 e início de 1980, quando o Governo Federal foi obrigado a socorrer a população por via aérea para atenuar o quadro desolador então existente, várias vezes aviões da FAB – Força Aérea Brasileira pousaram no antigo aeroporto de Sinop com produtos de primeira necessidade para atender seus habitantes.

É importante ressaltar que, a ocupação populacional do estado de Mato Grosso, durante o processo de colonização, ocorre de modo modesto. Devido talvez à precária infraestrutura da época (década de 1940 para 1950), ocorre no estado de Mato Grosso um aumento de aproximadamente 10,0 % da sua população, que é um crescimento bem abaixo do registrado à época em nível nacional, algo em torno de 26,0 %. Durante as décadas seguintes, o processo migratório esteve sempre em crescimento excepcional, alcançando 85,38% nas décadas de 1960 e 1970; 85,79% nas décadas de 1970 e 1980, 78,03% na década de 1990, sendo muito superior ao registrado nacionalmente para o período.

Na década de 1990 a 2000 constatou-se diminuição no ritmo de crescimento regional. Acredita-se que isto se deva à extinção da SUDAM (Superintendência de Desenvolvimento da Amazônia) e da SUDECO (Superintendência de Desenvolvimento do Centro-Oeste), ocorrida durante o governo Collor. Neste período foi registrada uma variação de 22,76% na população do estado. Tal percentual está muito abaixo do anteriormente alcançado, considerando-se o período de atuação das instituições no estado.

É bem verdade que, diante de tamanho poder de desenvolvimento e crescimento socioeconômico, a questão ambiental em Mato Grosso, principalmente nos Biomas Cerrado-Amazônia e Amazônia, torna-se alvo de grandes polêmicas e discussões. Pinto (2006, p. 19-20) adverte sobre algumas das principais consequências ambientais das políticas que corroboraram para a colonização do Estado de Mato Grosso, em especial da Região do Médio Norte e Norte, e da Amazônia durante os governos militares:

O histórico de ocupação e “integração” da Amazônia e, mais especificamente, ao longo da rodovia BR-163, [...] mostra que o megaprojeto de abertura da rodovia foi um modelo de desenvolvimento que ignorou a presença das populações nativas e as

especificidades de uma das mais exuberantes florestas tropicais do mundo. Isso trouxe sérias consequências socioambientais [...]. E, ainda, o coração da floresta amazônica é pressionado pela expansão da fronteira agrícola.

Diante do exposto, percebe-se que ocorre um processo intenso de colonização e ocupação do estado de Mato Grosso. Este processo ocorre sob comando do estado, de 1965 (Governo Castelo Branco) a 1990 (Governo José Sarney), e está centrado nas atividades primárias extrativas mineral e vegetal, e na agropecuária. Neste cenário percebe-se um dualismo tecnológico nas atividades do campo. Ou seja, por um lado têm-se formas avançadas de produção agropecuária em médias e grandes propriedades, com forte e farto apoio econômico do estado. Por outro lado, encontra-se a agropecuária de subsistência ou de pouca expressão econômica, privada, de certa forma, dos recursos financeiros limitados do Estado. Estas configurações contribuem fortemente para com o passivo ambiental na região.

A partir da década de 1990, após o extraordinário processo migratório, citado anteriormente, a expansão da ocupação e colonização do espaço territorial passa a ocorrer de uma forma mais natural e dentro da realidade brasileira. A desaceleração acontece, pois, ao mesmo tempo que a oferta de terras na região diminui, os preços das propriedades acabam se elevando, em decorrência da menor oferta e maior demanda, impulsionadas pela forte especulação imobiliária que passou a existir na região.

A execução de projetos de colonização e a descoberta de ouro no início da década de 1980 constituíram fatores para o aumento do fluxo migratório para o norte de Mato Grosso, ao passo que viabilizavam um possível acesso à terra. E mais, estes fatores foram responsáveis, em grande parte, pela urbanização recentemente ocorrida no estado. Assim, rapidamente distritos se transformaram em cidades, em decorrência, principalmente, da vinda de pequenos e médios trabalhadores rurais e migrantes que vislumbravam melhores condições de vida. Mendes (2012, p. 211) enfatiza ainda que: “Essa é uma característica importante do processo de colonização da fronteira agrícola, planejada ao nível das esferas governamentais e pelas companhias colonizadoras, de modo a gerar uma ocupação urbanizada”. Assim, de 34 municípios existentes em 1970, em 2005 o número de unidades municipais foi para 141, sendo que destes, a maioria foi oriunda de projetos de colonização.

A expansão e modernização da agricultura também contribuiu para com o processo rápido de urbanização dos municípios e é evidente e circunstancial que várias cidades, da mesorregião norte mato-grossense, se desenvolveram dependentes de atividades agrícolas e agroindustriais. Nesta relação, imposta pela expansão e modernização da agricultura, ambos os

ambientes, urbano e rural, dependem entre si, contribuindo simultaneamente para o desenvolvimento de ambos.

Em relação à produção agrícola e à industrialização, em alguns municípios da mesorregião norte vislumbra-se a necessidade de ampliação de atividades produtivas urbanas ligadas aos demais setores da economia. Assim, parece haver demanda por investimentos e pela diversificação de matérias primas produzidas, bem como, pela mão de obra qualificada para o processamento destes insumos. Exemplo de municípios que aparentam ter esta carência são, por exemplo, Alta Floresta, Sinop, Sorriso, Lucas do Rio Verde e Nova Mutum. Assim, a adoção e desenvolvimento de novos agentes econômicos, associados às redes agroindustriais e ao desenvolvimento preexistente destes municípios e da região, requer a ampliação e modernização da infraestrutura modal, elemento que já se apresenta como projeto do poder público em parceria com a iniciativa privada.

Sendo assim, as cidades agrícolas derivadas desse processo intenso de colonização do estado de Mato Grosso, sejam elas oriundas de incentivos governamentais ou geradas por colonizadoras privadas, estão em constante mudança e desenvolvimento, transformando-se a fim de atender aos requisitos e exigências desse processo.

Nestas cidades, diante da visão geral dos colonizadores de que o estado representava um nova oportunidade de vida e sucesso, formou-se uma elite agropecuária constituída por migrantes sulistas, representantes de uma clientela mais capitalizada que dispunha de recursos próprios [...] que por isso tinha como manter-se na aquisição e compra de terras, aquisição de financiamentos para plantio, compra de insumos, aquisição de sementes e maquinários, pagamento por empréstimos bancários, financiamento do plantio, entre outros (MENDES, 2012, grifo do autor).

Atualmente, graças ao processo de formação e excelente desenvolvimento dessas cidades, baseadas no agronegócio, verifica-se ainda uma intensa imigração de muitos outros trabalhadores em busca de nova vida e oportunidades. Estas pessoas são atraídas para estas cidades com excelente planejamento e infraestrutura por motivos que vão, desde a possibilidade da aquisição de terras e aberturas de novas fronteiras, até a oferta de emprego, oportunidades de negócios e excelentes níveis de educação e saúde.

2.2 Tópicos de Estatística

Ao estudar estatística, é bastante usual que se dedique algum tempo às medidas de posição e às medidas de dispersão, pois estes conteúdos são basilares e necessários à compreensão de tópicos mais avançados desta ciência. Assim, devido à sua importância e ao

fato destes conteúdos estarem intimamente ligados ao trabalho de pesquisa aqui descrito, tais medidas são definidas a seguir. Pelos mesmos motivos, conteúdos relacionados à Distribuição Normal e à padronização de variáveis também são apresentados.

2.2.1 Medidas de Posição

As medidas de posição, genericamente denominadas de estatísticas, representam uma série de dados e expressam informações quanto à posição da distribuição da curva de frequências. Assim, estas medidas proporcionam uma descrição resumida sobre o comportamento de um determinado conjunto de dados, além de caracterizar a distribuição como um todo. As principais medidas de posição são a média, a mediana e a moda.

Nas subseções que seguem serão usadas as seguintes notações:

x_i = Valor de cada Variável ou Elemento da Amostra.

\bar{x} = Média Amostral

n = Tamanho da Amostra

μ = Média Populacional

N = Tamanho da População

2.2.1.1 Média Populacional

A média populacional de um conjunto de dados é a razão entre a soma dos valores do conjunto observado e o número total de valores que compõem o conjunto. Assim, considerando uma população com N elementos, a média populacional é dada por:

$$\mu = \frac{x_1 + x_2 + \dots + x_N}{N} \quad \text{ou} \quad \mu = \frac{\sum_{i=1}^N x_i}{N} \quad (01)$$

Equação 1: Média Populacional Simples

2.2.1.2 Média Amostral

Quando se toma um subconjunto propriamente contido de um conjunto de dados tem-se a amostra. Logo, a média amostral consiste do quociente entre a soma de todos os valores da amostra e a cardinalidade da amostra. Logo, a média amostral é dada por:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} \quad \text{ou} \quad \bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (02)$$

Equação 2: Média Amostral Simples

2.2.1.3 Moda

Embora a definição de moda não seja um conceito necessário neste estudo de caso, por uma questão de completeza, define-se a moda como segue. Assim, considerando um conjunto ordenado de valores, a moda (M_o) será o valor mais frequente desse conjunto. Evidentemente, um conjunto de valores pode não apresentar moda, caso em que todos os valores da variável em estudo ocorrem com mesma frequência. Este fato caracteriza um conjunto amodal. Por outro lado, se o conjunto tiver mais de um valor predominante, então diz-se que este conjunto é plurimodal.

2.2.1.4 Mediana

A mediana é o valor que divide uma série ou conjunto de dados ordenados em duas partes, de modo que pelo menos a metade dos elementos sejam iguais ou maiores do que ela, e que pelo menos a outra metade dos elementos sejam menores ou iguais a ela. Desta definição, conclui-se que a mediana é uma separatriz, logo essa medida apresenta um número de ordem. O número que indica a ordem em que se encontra o valor correspondente à mediana é denominado de elemento mediano, cujo símbolo é E_{Md} . (TOLEDO, 1985 p.152).

A determinação do elemento mediano, considerando um rol ou lista ordenada de dados, segue as seguintes possibilidades:

a) Se o número de dados for ímpar:

Considerando um conjunto de dados com cardinalidade ímpar “n”, pode-se encontrar o elemento mediano da seguinte forma:

$$E_{Md} = \frac{n+1}{2} \quad (03)$$

Equação 3: Elemento Mediano para uma Série Ímpar de Dados

b) Se o número de dados for par:

Para a determinação do elemento mediano em um conjunto de dados de cardinalidade par, considerando que haverá dois valores centrais, então a mediana será determinada pela média aritmética destes valores. Desta forma, o elemento mediano será determinado pela fórmula abaixo, tomando-se também o elemento subsequente.

$$E_{Md} = \frac{n}{2} \quad (04)$$

Equação 4: Elemento Mediano para uma Série Par de Dados

2.2.2 Medidas de Dispersão

Entende-se por dispersão a variação, variabilidade ou espalhamento dos dados de uma determinada série. Para se quantificar a dispersão, as seguintes medidas são usadas mais frequentemente: a amplitude total, a variância e o desvio-padrão. A escolha de uma medida de dispersão em detrimento de outra dependerá do objetivo que se pretende alcançar. Especifica-se que, para este estudo de caso, serão consideradas as medidas de dispersão para dados não agrupados.

2.2.2.1 Amplitude Total

De acordo com Toledo (1985), a amplitude total ou intervalo total de um conjunto de números é a diferença entre os valores extremos do conjunto.

Para dados não agrupados, o cálculo da amplitude total (AT) pode ser expresso pela seguinte fórmula:

$$AT = x(\text{máx}) - x(\text{mín}) \quad (05)$$

Equação 5: Amplitude Total

Em que:

- $x(\text{máx})$ é o maior valor presente na série;
- $x(\text{mín})$ é o menor valor verificado na série.

Quanto menor a amplitude total, mais homogêneo será o conjunto, isto é, menos dispersos estarão os dados.

Em relação à variância e ao desvio-padrão, pode-se afirmar que estes elementos estatísticos complementam as informações obtidas por meio das medidas de tendência central e capturam outros aspectos da dispersão dos dados, não abarcados pela amplitude total. O fato das medidas de dispersão serem consideradas bastantes estáveis, faz com que elas sejam frequentemente utilizadas no cotidiano.

Para que o desvio-padrão seja definido é necessário *a priori* definir a variância.

As seguintes notações serão utilizadas nas definições que seguem:

$$\sigma^2 = \text{Variância Populacional}$$

$$s^2 = \text{Variância Amostral}$$

$$\sigma = \text{Desvio – Padrão Populacional}$$

$$s = \text{Desvio – Padrão Amostral}$$

2.2.2.2 Variância Populacional

A variância de uma população ou conjunto de dados $\{x_1, x_2, \dots, x_N\}$, de cardinalidade N , é uma medida de dispersão definida como a média do quadrado dos desvios dos elementos do conjunto de dados em relação à média populacional μ . Ou seja, matematicamente tem-se:

$$\sigma^2 = \sum_{i=1}^N \frac{(x_i - \mu)^2}{N} \quad (06)$$

Equação 6: Variância Populacional

2.2.2.3 Variância Amostral

A variância de uma amostra $\{x_1, x_2, \dots, x_n\}$, com cardinalidade n , é definida como a soma dos quadrados dos desvios dos elementos do conjunto de dados em relação à sua média \bar{x} , dividido por $(n - 1)$ elementos. Ou seja, a variância amostral é dada por:

$$s^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n-1} \quad (07)$$

Equação 7: Variância Amostral

Cabe ressaltar que, ao se utilizar a média amostral como estimador de n elementos no cálculo da variância amostral, perde-se 1 grau de liberdade. Assim, o denominador do termo geral é “ $n - 1$ ”.

2.2.2.4 Desvio-Padrão Populacional

Ao se extrair a raiz quadrada da variância populacional, obtém-se o desvio-padrão de um conjunto de dados. Matematicamente tem-se:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\sum_{i=1}^N \frac{(x_i - \mu)^2}{N}} \quad (08)$$

Equação 8: Desvio Padrão Populacional

2.2.2.5 Desvio-Padrão Amostral

O desvio-padrão amostral é uma medida de dispersão dos em relação à média. De modo análogo ao definido para o desvio-padrão populacional, se obtém o desvio-padrão amostral extraíndo-se a raiz quadrada da variância amostral. Caber ressaltar que, a variância amostral é expressa na unidade de medida dos dados elevada ao quadrado. Assim, o desvio-padrão é dado na mesma unidade de medida dos dados. Logo, esta medida de dispersão é dada por:

$$s = \sqrt{s^2} = \sqrt{\sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n-1}} \quad (09)$$

Equação 9: Desvio Padrão Amostral

2.2.3 A Distribuição Normal e a Padronização de Variáveis

Ao perceber que, à medida que o número de eventos do lançamento de algumas moedas aumentava, a distribuição binomial se aproximava de uma curva suave, o matemático francês, Abraham de Moivre (1667 – 1754) começou a desenvolver o conceito de Distribuição Normal.

Pierre Simon de Laplace aperfeiçoou a teoria e, com base nela, estabeleceu o que hoje constitui o Teorema de Moivre/Laplace ou teorema central do limite¹, que expressa que: quando o tamanho da amostra aumenta, a distribuição amostral da sua média aproxima-se gradativamente de uma distribuição normal. O teorema, que decorre da análise dos erros de experimentos por meio da distribuição normal, foi publicado no livro “Teoria Analítica da Probabilidade”. Este teorema é importante, visto que muitos fenômenos naturais possuem distribuição normal ou, mesmo que não possuam distribuição normal, a média de repetidas amostras coletadas é aproximadamente normal. Logo, quanto maior for o tamanho da amostra, melhor será essa aproximação.

Em 1808, os matemáticos Robert Adrian e Carl Friedrich Gauss desenvolveram a equação da distribuição, demonstrando que esta modelava com eficácia os erros de observações astronômicas. Em 1809, Gauss utilizou-se da equação para justificar o Método dos Mínimos Quadrados – MMQ, introduzido por Adrien Marie Legendre no mesmo ano.

A distribuição normal é então a mais conhecida das distribuições contínuas, sendo comumente representada pela curva de Gauss. Sua importância é visível e comprovada por meio de diversos estudos práticos, nos quais, após testes, se conclui pela normalidade dos dados. Por exemplo, em inúmeros estudos sobre a altura de determinada população, percebe-se que, entre outras características físicas e sociais, a média dos dados convergem sempre para uma distribuição normal.

Nestes termos, se estabelece a função densidade de probabilidade na variável aleatória X que caracteriza a distribuição normal como:

¹ O teorema central do limite é um resultado fundamental com diversas aplicações. Mesmo quando a distribuição da população de eventos não segue uma distribuição Normal, a distribuição da média dos dados converge para a distribuição Normal conforme o tamanho da amostra aumenta. Assim, qualquer que seja a distribuição da amostra, para um número suficientemente grande de observações, a distribuição de probabilidade da média se aproxima da distribuição Normal com média μ e variância $\frac{\sigma^2}{n}$, nos quais μ é a média e σ^2 é a variância da população.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], x \in (-\infty, \infty) \quad (10)$$

Equação 10: Função Densidade de Probabilidade

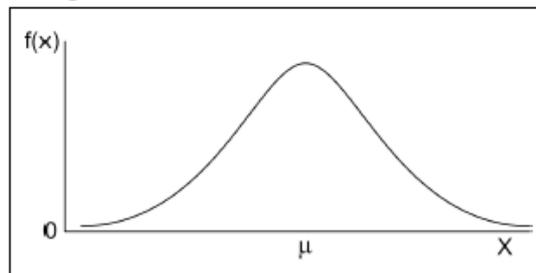
Em que:

$$X \sim N(\mu, \sigma^2)$$

Ou seja, X tem distribuição Normal com média μ e desvio-padrão σ .

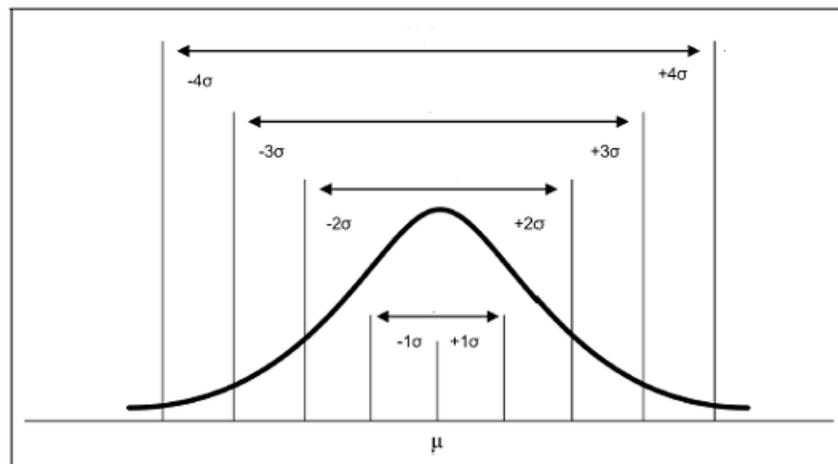
Quando o gráfico da função gaussiana é desenhado, verifica-se que este tem a forma de "sino", conforme o ilustrado na Figura 3:

Figura 3: Gráfico da curva de Gauss.



Para achar a área sob a curva gaussiana e assim acessar uma dada probabilidade, deve-se conhecer dois valores numéricos relativos ao conjunto de dados em questão: a média μ e o desvio-padrão σ populacionais. A Figura 4, a seguir, indica algumas áreas importantes por meio de μ e σ :

Figura 4: Áreas sob a curva Normal em termos da Média (μ) e do Desvio Padrão (σ).



Quando as medidas μ e σ , relativas à população, não são conhecidas (caso usual), então estes valores serão estimados por \bar{x} e s .

Sendo a curva de Gauss uma função densidade de probabilidade, então, para se acessar uma dada probabilidade, dever-se-ia calcular a integral definida (ou imprópria) da curva em relação ao intervalo de domínio desejado. Logo, a área obtida corresponderia à probabilidade desejada. No entanto, tal tarefa não é trivial e, na prática, se utiliza a padronização

(normalização) das variáveis do referido intervalo e se acessa a probabilidade em uma tabela de probabilidades, relativa aos dados padronizados.

Logo, de posse de μ e σ ou \bar{x} e s , pode-se proceder a padronização dos dados. Em Estatística a padronização é bastante utilizada e que consiste em subtrair de uma de uma variável que pertence ao conjunto de dados a sua média e dividir o resultado pelo desvio-padrão deste conjunto de dados.

Verifica-se que, cada valor de μ e σ dá origem a uma curva de distribuição de probabilidade. Assim, ao invés de se trabalhar com curvas particulares, relativas aos possíveis valores de μ e σ que podem ocorrer, trabalha-se com uma curva de distribuição normal padronizada, na qual $\mu = 0$ e $\sigma = 1$. Para obter tal distribuição, deve-se reduzi-la a uma variável padronizada Z , efetuando-se o seguinte cálculo:

$$Z = \frac{x - \mu}{\sigma} \quad (11)$$

Equação 11: Padronização de Varáveis

Desta forma, ao se deslocar o centro de um conjunto de dados por meio da média μ , para a origem do plano cartesiano, realiza-se a padronização destes dados de modo que correspondam à condição necessária, tornando o desvio-padrão do conjunto de dados unitário ($\sigma = 1$), independentemente do desvio-padrão original do conjunto de dados.

Ao se padronizar ou normalizar um conjunto de dados, sua média passa de um valor qualquer para $\mu = 0$, e seu desvio padrão passa a ser $\sigma = 1$. Logo, qualquer que seja o conjunto de dados, após a padronização de seus valores x_i ($i = 1, \dots, N$), estes serão transformados em valores padronizados Z_i ($i = 1, \dots, N$). Nestes termos, qualquer que seja o conjunto de dados, após a padronização, os seus elementos estarão dispersos no entorno de 0 (zero) e possuirão, salvo *outliers*, amplitudes totais semelhantes.

Embora a média e o desvio padrão de todas as variáveis após a padronização sejam iguais, isto não torna os conjuntos de dados todos iguais. Como prova desta afirmação, basta notar que os conjuntos serão distintos, pois eles terão assimetrias e curtoses distintas e são essas medidas que moldam os conjuntos. Ou seja, estas medidas definem sua aparência e seu comportamento. Tanto a assimetria quanto a curtose não se alteram quando a variável é padronizada. Logo, verifica-se que estas medidas são invariantes a transformações desse tipo.

2.3 O Método dos Mínimos Quadrados Aparados

O método dos mínimos quadrados aparados (em inglês: *Least Trimmed Squares - LTS*) foi criado por Rousseeuw e Leroy (1987) e consiste de uma variação do Método dos Mínimos

Quadrados (MMQ) tradicional. Cabe lembrar que, o MMQ consiste de uma técnica de otimização matemática que objetiva a determinação do melhor ajuste para um conjunto com n observações (dados), tendo por base a minimização da soma dos quadrados dos resíduos (e^2). Assim, matematicamente a condição imposta no MMQ é:

$$\min \sum_{i=1}^n e_i^2 \quad (12)$$

Equação 12: Soma dos Quadrados dos Resíduos (e^2)

Onde, os resíduos e_i ($i = 1, \dots, n$) são as diferenças entre os valores das observações presentes no conjunto de dados e os seus respectivos valores estimados.

Nestes termos, a mudança proposta por Rousseeuw e Leroy (1987) consiste em não se somar todos os n erros quadrados, mas apenas uma quantidade h destes termos. Assim, no processo de escolha dos h termos, são eliminados os maiores termos da soma quadrática. Para que a escolha seja feita, são calculadas todas retas possíveis, considerando as n observações disponíveis, tomando-as duas a duas.

A partir da construção das retas, são calculados e ordenados, em ordem crescente, os quadrados dos resíduos e^2 em cada reta. Assim tem-se, na expressão abaixo, os quadrados dos resíduos, relativos a uma das possíveis retas calculadas e ordenados do menor para o maior:

$$(e^2)_{1:n} \leq (e^2)_{2:n} \leq (e^2)_{3:n} \leq \dots \leq (e^2)_{n:n}$$

Cabe explicar que, na expressão acima, os índices subscritos $1:n$, $2:n$, ..., $n:n$ representam, respectivamente, o menor quadrado de resíduo de n resíduos, o segundo menor de n , o terceiro menor de n e assim por diante.

Uma vez que os resíduos de uma dada reta sejam conhecidos e ordenados, escolhe-se apenas os h primeiros destes valores ordenados, com $h < n$, para gerar a soma dos quadrados dos resíduos. Logo, os demais termos da soma são desprezados ou, conforme especifica o nome da técnica, aparados.

O processo segue de modo análogo, sendo repetido para uma nova reta, relativa a outros dois pontos quaisquer do conjunto de observações. Cada reta construída dará origem à uma nova soma de quadrados aparada. Assim, será escolhida, dentre o total de retas geradas, aquela que originar o menor valor de soma aparada.

Fazendo uma analogia ao apresentado para o MMQ, matematicamente, pode-se expressar o critério de Rousseeuw e Leroy (1987) pela seguinte função objetivo:

$$\min \sum_{i=1}^h e_i^2$$

Onde, $h = \left(\frac{n}{2}\right) + 1$, que é conhecida como “mediana alta”.

Rousseeuw e Leroy (1987), demonstram que o estimador *LTS* atinge o máximo ponto de quebra de 50%.

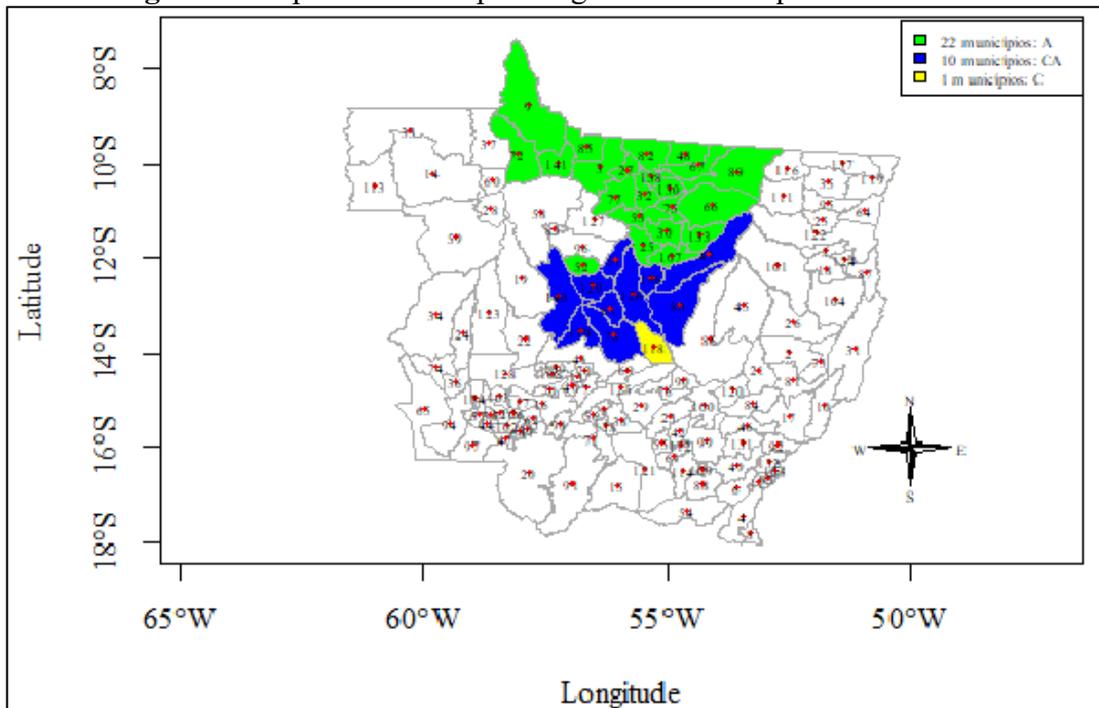
3. METODOLOGIA

Para atender aos objetivos propostos neste trabalho, desenvolveu-se, após a criação de índices de desenvolvimento, uma pesquisa exploratória dos dados relativos aos indicadores físicos e socioeconômicos de uma amostra dos municípios do estado de Mato Grosso. Os municípios tomados como amostra situam-se ao longo do Rio Teles Pires, os quais serão afetados pela construção das usinas hidrelétricas no norte do estado de Mato Grosso. Desta forma, esta dissertação se insere no contexto do projeto de pesquisa “Estudo das formas de relação das atividades econômicas (Pecuária e Produção de grãos) com a água na Amazônia Norte Matogrossense e suas consequências”, projeto este em andamento, após aprovação via edital de Redes de Pesquisa em Mato Grosso nº 037/2016. Ou seja, após a criação e utilização dos índices para a geração de valores descritivos do desenvolvimento dos municípios, efetuou-se uma análise comparativa de cunho estatístico entre índices, a fim de validar o índice inédito criado.

Assim, na seção 3.1 tem-se a descrição dos dados/indicadores relativos aos municípios que compõem a amostra, com a indicação das variáveis utilizadas no estudo (Área, População, Densidade Demográfica, Produto Interno Bruto e Índice de Desenvolvimento Humano). Na seção 3.2 são apresentados os pormenores para a geração do índice proposto através da Normalização dos Dados, o qual buscou-se que, todos os dados se apresentassem num mesmo *range* e, posteriormente, o cálculo da Média Aritmética Simples dos dados normalizados. Os procedimentos para a validação estatística do índice estão descritos na seção 3.3 onde por meio de uma regressão resistente, baseado em medianas, foi gerado um segundo índice cujos valores foram comparados com o índice criado.

3.1 Dados

O presente trabalho, teve como objetivo principal a geração de uma classificação hierárquica para 33 municípios do norte do estado de Mato Grosso, situados no entorno da cidade polo de Sinop e que margeiam o Rio Teles Pires, considerando subconjuntos de municípios agrupados segundo os seus respectivos biomas de localização (A: Amazônia, CA: Cerrado-Amazônia - Área de transição entre os dois biomas e C: Cerrado). O mapa presente na Figura 5 mostra os trinta e três municípios, objetos deste estudo separados por seus respectivos biomas.

Figura 5: Mapa dos Municípios Segundo Seus Respectivos Biomas.

Fonte: Resultados da pesquisa.

Tal classificação se deu a partir do desenvolvimento de um índice de desenvolvimento regional baseado em indicadores (dados de entrada) que refletissem os aspectos do desenvolvimento de cada município. Estes indicadores são:

1. Área Territorial no ano de 2016 em km² (ÁREA),
2. População estimada no censo de 2016 (POPULAÇÃO),
3. Densidade demográfica (DENSIDADE),
4. Produto Interno Bruto per capita no ano de 2014 (PIB); e
5. Índice de desenvolvimento humano (IDH).

O conjunto destes dados/indicadores agrupados por município, considerando os respectivos biomas, pode ser visualizado na Tabela 1.

Tabela 1: Dados/indicadores que refletem o desenvolvimento de municípios do Norte de Mato Grosso: Divisão por Bioma.

MUNICÍPIOS	BIOMA	ÁREA (km ²)	POPULAÇÃO	DENSIDADE	PIB	IDH
ALTA FLORESTA	A	8.953,19	49.164	5,48	20.853,81	0,71
APIACÁS	A	20.488,47	8.567	0,42	12.978,53	0,68
CARLINDA	A	2.416,14	10.990	4,59	11.626,21	0,67
COLÍDER	A	3.103,96	30.766	9,95	21.645,43	0,71
GUARANTÃ DO NORTE	A	4.734,59	32.216	6,80	16.236,84	0,70
ITAÚBA	A	4.529,58	4.575	1,01	27.664,32	0,69

MARCELÂNDIA	A	12.273,55	12.006	0,98	19.129,38	0,70
MATUPÁ	A	5.219,03	14.174	2,71	27.482,08	0,72
NOVA MONTE VERDE	A	5.150,56	8.093	1,54	15.110,04	0,69
NOVA BANDEIRANTES	A	9.593,27	11.643	1,21	12.713,91	0,65
NOVA CANAÃ DO NORTE	A	5.955,41	12.127	2,03	23.921,71	0,69
NOVA GUARITA	A	1.114,13	4.932	4,43	13.879,57	0,69
NOVA SANTA HELENA	A	2.375,58	3.468	1,47	18.624,77	0,71
NOVO MUNDO	A	5.791,05	7.332	1,27	14.152,02	0,67
PARANAÍTA	A	4.796,01	10.684	2,23	81.433,04	0,67
PEIXOTO DE AZEVEDO	A	14.400,64	30.812	2,16	10.643,67	0,65
TERRA NOVA DO NORTE	A	2.432,07	11.291	4,41	18.176,83	0,70
CLÁUDIA	A	3.849,99	11.028	2,86	23.945,17	0,70
ITANHANGÁ	A	2.898,07	5.276	1,82	29.837,47	0,71
SANTA CARMEM	A	3.855,36	4.085	1,06	47.565,45	0,72
SINOP	A	3.942,23	113.099	28,69	33.807,26	0,75
UNIÃO DO SUL	A	4.581,91	3.760	0,82	24.126,02	0,67
FELIZ NATAL	CA	11.679,00	10.933	0,95	23.601,40	0,70
IPIRANGA DO NORTE	CA	3.467,05	5.123	1,48	78.912,83	0,73
LUCAS DO RIO VERDE	CA	3.683,59	45.556	12,43	49.953,90	0,77
NOVA MARINGÁ	CA	11.557,30	6.590	0,57	45.166,01	0,66
NOVA MUTUM	CA	9.544,57	31.649	3,31	65.493,85	0,76
NOVA UBIRATÃ	CA	12.490,21	9.218	0,73	65.145,82	0,67
SÃO JOSÉ DO RIO CLARO	CA	4.536,20	17.124	3,77	29.627,02	0,68
SORRISO	CA	9.329,60	66.521	7,13	57.087,49	0,74
TAPURAH	CA	4.491,05	10.392	2,30	53.550,07	0,71
VERA	CA	2.962,69	10.235	3,45	33.853,15	0,68
SANTA RITA DO TRIVELATO	C	4.733,92	2.491	0,53	132.591,36	0,74

Fonte: IBGE.

Da análise da tabela, pode-se verificar que no bioma Amazônia estão presentes 22 municípios, na área de transição CA encontram-se 10 municípios e há apenas um município relativo ao bioma Cerrado (município de Santa Rita do Trivelato). Assim, a classificação que considere este último bioma é desnecessária.

3.2 Índice de Desenvolvimento Proposto

Considerando os municípios pertencentes a cada um dos biomas remanescentes (Amazônia e Cerrado-Amazônia), procedeu-se o cálculo do índice proposto como segue.

→ Padronização dos Dados

Para geração do índice de desenvolvimento proposto foi feita a normalização/padronização dos vetores de dados/indicadores (ÁREA, POPULAÇÃO, DENSIDADE, PIB e IDH), com intuito de transformá-los em variáveis que, mesmo representando grandezas (ou indicadores) diferentes, estivessem normalizadas em um mesmo intervalo (ou *range*). A normalização se deu conforme a equação 13.

$$Z_{ij} = \frac{X_{ij} - \bar{X}_j}{\sqrt{S_{ij}}} \quad (13)$$

Equação 13: Normalização dos Dados

Na qual:

- X_{ij} : representa cada observação dos vetores de dados/indicadores;
 - $i = \{\text{ÁREA, POPULAÇÃO, DENSIDADE, PIB, IDH}\}$: indexa cada um dos vetores de dados/indicadores;
 - j : indexa os municípios; este contador variará de 1 até 22 quando o bioma Amazônia for considerado e de 1 até 10 quando o bioma Cerrado-Amazônia estiver em foco;
 - \bar{X}_j : Média de cada vetor de dados/indicadores (ÁREA, POPULAÇÃO, DENSIDADE, PIB e IDH), nos biomas Amazônia e Cerrado-Amazônia considerados separadamente.
 - $\sqrt{S_{ij}}$: Desvio-padrão de cada vetor de dados/indicadores (ÁREA, POPULAÇÃO, DENSIDADE, PIB e IDH), nos biomas Amazônia e Cerrado-Amazônia considerados separadamente.
- Média dos Dados Padronizados

Em seguida, para cada um dos municípios, considerando os biomas em análise separadamente, foi construída uma variável-índice de desenvolvimento (ÍNDICE), a partir das variáveis Z_{ij} . A equação abaixo expressa a variável-índice criada:

$$\widehat{ÍNDICE}_j = \frac{Z_{ÁREA_j} + Z_{POPULAÇÃO_j} + Z_{DENSIDADE_j} + Z_{PIB_j} + Z_{IDH_j}}{5} \quad (14)$$

Equação 14: Índice Proposto

Cabe explicar que, para cada município (indexado pelo contador j na equação 14) foi calculado um valor para a variável-índice (ÍNDICE) e tal valor representa o nível de desenvolvimento de cada município, pois a variável-índice foi construída considerando-se os indicadores de desenvolvimento (ÁREA, POPULAÇÃO, DENSIDADE, PIB e IDH) de cada município. Assim, pôde-se hierarquizar o desenvolvimento de cada município por meio dos valores obtidos para a variável-índice.

Nestes termos, uma vez calculada a variável ÍNDICE de cada município, foi possível ordenar a lista de municípios em ordem crescente, para cada bioma, considerando tal variável.

Considerando que, o índice construído (ÍNDICE) é fruto de uma modelagem matemática particular e inédita, desenvolvida no âmbito desta pesquisa, fez-se necessária a validação de tal teoria por meio de comparação com métodos tradicionalmente aceitos.

3.3 Validação Estatística do Índice Proposto

Com o intuito de respaldar os resultados obtidos a partir do índice gerado, procedeu-se a sua validação estatística. Assim, buscando evitar possíveis problemas de multicolinearidade das variáveis e o não atendimentos dos demais pressupostos relativos aos métodos estatísticos paramétricos como, por exemplo, aqueles pertinentes ao método dos mínimos quadrados, foram utilizados métodos de regressão não paramétricos (robustos) nas variáveis padronizadas, para os dois biomas. Isto posto, para fins de comparação, foi gerado um segundo índice ($\widehat{ÍNDICE}$) por meio de uma regressão resistente (método não paramétrico), que se baseia em medianas e não em médias, cujos valores foram comparados com o índice inédito criado e apresentado na seção anterior. O segundo índice possui a seguinte forma funcional:

$$\widehat{ÍNDICE}_j = f(ÁREA_j, POPULAÇÃO_j, DENSIDADE_j, PIB_j, IDH_j, \beta_0, \dots, \beta_5)$$

Ou

$$\widehat{ÍNDICE}_j = \beta_0 + \beta_1 \cdot ÁREA_j + \beta_2 \cdot POPULAÇÃO_j + \beta_3 \cdot DENSIDADE_j + \beta_4 \cdot PIB_j + \beta_5 \cdot IDH_j + e_j \quad (15)$$

Equação 15: Fórmula funcional por LTS

Em que:

- $\beta_1, \beta_2, \beta_3, \beta_4$ e β_5 são os coeficientes que constituem as formulações das funções que aproximam $\widehat{ÍNDICE}_j$ para os municípios (j) dos biomas em estudo. Estes coeficientes, ao serem estimados, revelaram os impactos de cada uma das variáveis sobre valor do $\widehat{ÍNDICE}_j$, em cada bioma;
- e_j são os erros ou discrepâncias entre os valores da função discreta e sua respectiva aproximação dada pela função a ser estimada para cada bioma;
- j indexa os municípios envolvidos na pesquisa e variará de 1 até 22 quando o bioma Amazônia for considerado, ou de 1 até 10 quando o bioma Cerrado-Amazônia estiver em foco.

O erro, ou discrepância que se estabelece entre os índices, pode ser definido pela diferença, $e_j = \left(\widehat{ÍNDICE}_j - ÍNDICE_j \right)$ e não necessariamente possuirá distribuição normal $N(0, \sigma^2)$. Tal fato decorre da abordagem não paramétrica utilizada. Assim, considera-se que os erros e os dados possuirão distribuição livre em cada uma das estimativas.

O método não paramétrico utilizado no presente trabalho foi o dos Mínimos Quadrados Aparados (ou, em inglês, *Least Trimmed of Squares - LTS*), que consiste em:

- 1) Capturar os resíduos e_j^2 , gerados a partir das estimativas do método;
- 2) Ordenar tais resíduos em ordem crescente: $e_1^2 < e_2^2 < \dots < e_q^2 < \dots < e_{j-1}^2 < e_j^2$.

A posição de e_q^2 é normalmente maior que a metade de j e este é o número de observações de resíduos utilizados do ajuste;

3) Minimizar, via método *LTS*, a soma dos menores resíduos quadráticos, dados pela posição mediana e_q^2 . Cabe esclarecer que, a mudança proposta em relação ao Método dos Mínimos Quadrados é que, ao invés de se somar todos os erros quadrados de regressão, toma-se apenas alguns destes termos, eliminando os maiores termos da soma quadrática. Dessa forma, pode-se encontrar assim os coeficientes β_0, \dots, β_5 , com base na equação abaixo:

$$\frac{\partial(e_j^2)}{\partial\beta_i} = 0, \quad i = \{0,1,2,3,4,5\}$$

A aplicação do método *LTS*, não fornece as medidas de ajustes de bondade do índice para os coeficientes, como os testes t e F fazem. O R^2 também não é fornecido.

Uma vez criado o segundo índice (via *LTS*) e de posse do índice proposto (seção 3.2), procedeu-se a comparação entre ambos. Logo, Para a comparação e descrição de ambos os

índices criados ($\widehat{ÍNDICE}_j$ e $\widehat{ÍNDICE}_j$), foram inicialmente construídos histogramas pela fórmula clássica de Sturges, da seguinte maneira:

- 1) Os valores obtidos por meio dos índices criados foram ordenados em ordem crescente, dando origem, respectivamente, a duas listas de valores ordenados;
- 2) Calculou-se o número ideal de intervalos de classes k para cada lista de valores ordenados de índices, pela fórmula de Sturges:

$$k = 1 + 3.33\log(n) \quad (16)$$

Equação 16: Cálculo da Quantidade de Classes

Na qual n é o número de municípios envolvidos em cada bioma analisado.

- 3) Determinou-se ainda a largura ideal h de cada intervalo de classe, da seguinte forma:

$$h = \frac{(X_{\max} - X_{\min})}{k} \quad (17)$$

Equação 17: Cálculo do Intervalo de Classe

Onde, X_{\max} e X_{\min} , são os valores máximos e mínimos de cada índice calculado, em cada lista de valores ordenados gerada.

- 4) Calculados k e h , pode-se construir para ambos um gráfico de barras, relativos a cada uma das listas. A partir de tais gráficos, pode-se ter noção da variabilidade dos índices criados em suas respectivas classes.

Ainda com a finalidade de comparar descritivamente os índices criados, foram construídos intervalos de confiança, com nível de confiança de 1%, ou seja, 99% de certeza das estimativas de cada índice. Os intervalos de confiança foram dados por:

$$IC(\mu, 99\%) = \overline{ÍNDICE} \pm t_{99\%} \frac{S_{ÍNDICE}}{\sqrt{n}} \quad (18)$$

Equação 18: Cálculo do Intervalo de Confiança

Em que:

- $\overline{ÍNDICE}$: média dos índices criados. Logo, haverá duas médias, uma relativa à lista de valores de índices gerados por meio do índice proposto, e outra média referente à lista de valores calculados a partir do índice construído via *LTS*;
- $S_{ÍNDICE}$: desvio-padrão amostral dos índices criados. Da mesma forma como no caso anterior, haverá dois valores de desvio-padrão;
- n : número de municípios de cada bioma;

- $t_{99\%}$: valor tabelado da distribuição t-Student, relativo à precisão almejada nas estimativas. Este valor é utilizado em medidas amostrais.

Cabe esclarecer que, no presente trabalho também foi analisado o comportamento de associação das duas variáveis-índices criadas $(\widehat{ÍNDICE}_j, \widehat{ÍNDICE}_j)$ por meio do coeficiente de correlação de Pearson, com o intuito de identificar comportamentos parecidos e válidos entre os índices. O coeficiente de correlação de Pearson é dado por:

$$r = \frac{n \cdot \sum_{i=1}^j (\widehat{ÍNDICE}_j \cdot \widehat{ÍNDICE}_j) - (\sum_{i=1}^j \widehat{ÍNDICE}_j) \cdot (\sum_{i=1}^j \widehat{ÍNDICE}_j)}{\sqrt{\left(n \cdot \sum_{i=1}^j \widehat{ÍNDICE}_j^2 \right) - (\sum_{i=1}^j \widehat{ÍNDICE}_j)^2} \cdot \sqrt{\left(n \cdot \sum_{i=1}^j \widehat{ÍNDICE}_j^2 \right) - (\sum_{i=1}^j \widehat{ÍNDICE}_j)^2}} \quad (19)$$

Equação 19: Correlação de Pearson

com $-1 \leq r \leq 1$.

O teste de hipóteses associado ao coeficiente de correlação de Pearson consiste em pôr a prova as seguintes hipóteses:

- $H_0: r = 0$: não há relação entre as variáveis $(\widehat{ÍNDICE}_j, \widehat{ÍNDICE}_j)$
- $H_1: r \neq 0$: há relação entre as variáveis $(\widehat{ÍNDICE}_j, \widehat{ÍNDICE}_j)$

O teste de hipóteses foi efetuado por meio do teste t, cuja estatística é:

$$t_c = \frac{r \cdot \sqrt{j-2}}{\sqrt{1-r^2}} \quad (20)$$

Equação 20: Teste de Confiança

Deve-se considerar que a estatística acima possui distribuição t-Student, com $j - 2$ graus de liberdade. A interpretação do teste de hipóteses se dá da seguinte maneira: Rejeita-se H_0 ao nível de significância α se $|t_c| > t_{\alpha, j-2}$, sendo que os valores críticos $t_{\alpha, j-2}$, da estatística t (t-Student) são tabelados.

4. RESULTADOS E DISCUSSÃO

Neste capítulo encontram-se os principais resultados e análises advindos da execução da pesquisa, bem como são especificados os recursos computacionais utilizados para o processamento dos dados. Inicialmente, na seção 4.1 é apresentado o *software* utilizado para a geração dos resultados aqui apresentados. Na seção 4.2 os principais resultados são apresentados, detalhados e discutidos.

4.1 Aspectos Computacionais:

O processamento dos dados, previsto na metodologia, ocorreu no *software* RStudio. O *software* livre RStudio é um ambiente de desenvolvimento integrado (ou, em inglês, IDE - *Integrated Development Environment*) que trabalha com a linguagem de programação de mesmo nome (R). O R é uma linguagem de programação indicada àqueles que desejam gerar gráficos e efetuar cálculos estatísticos. No entanto, suas potencialidades não se restringem apenas aos conteúdos estatísticos. Pode-se fazer o *download* gratuito de “pacotes” que contém ferramentas computacionais voltadas às mais diferentes áreas do conhecimento.

O funcionamento do RStudio depende inicialmente da instalação do *software* R, que materializa a linguagem R. Posteriormente, a fim de se agregar recursos suplementares, o programa RStudio deve ser instalado. O programa R é escrito em linguagem de programação C++ e está disponível para *download* no *site* do CRAN (*The Comprehensive R Archive Network*). Por sua vez, o RStudio, que possui uma versão *desktop* gratuita, é disponibilizado em seu *site* (<https://www.RStudio.com/products/RStudio/#Desk>).

4.2 Análise dos resultados

A apresentação dos resultados e discussões relativas a eles vão no sentido de:

1. Gerar ordenações coerentes das listas de municípios de cada bioma, considerando os índices de desenvolvimento;
2. Comparar estatisticamente os índices apresentados a fim de validar o índice proposto na pesquisa aqui relatada.

Cabe esclarecer que, nesta seção são apresentados resultados relativos aos municípios de cada bioma em separado: Amazônia e Cerrado-Amazônia. Assim, pôr-se-á a prova os índices relativos a cada bioma.

4.2.1 Bioma Amazônia

Para o bioma Amazônia obteve-se os seguintes resultados das análises descritivas das variáveis em foco, considerando-se 22 municípios.

Tabela 2: Valores das estatísticas de cunho descritivo das variáveis dos municípios do bioma Amazônia.

AMAZÔNIA	ÁREA (Km ²)	POPULAÇÃO	DENSIDADE DEMOGRÁFICA	PIB	IDH
MÉDIA	6.020,67	18.185,82	4,00	23.888,80	0,69
DESVIO PADRÃO	4.598,74	24.188,39	5,98	15.477,28	0,02

Fonte: Resultados da pesquisa.

Na Tabela 2, verifica-se que as variáveis (indicadores) com maior dispersão são a “População” e o “PIB”. Por outro lado, o “IDH” e a “Densidade Demográfica” variam menos, considerando o desvio-padrão calculado.

As variáveis $\widehat{INDICE}_j (I)$, $\widehat{INDICE}_j (\hat{I})$ e a ordem crescente dos municípios (O), foram construídas a partir da normalização das variáveis e são apresentadas na Tabela 3, considerando o bioma Amazônia.

Tabela 3: Ordenação de Municípios que compõem o bioma Amazônia considerando os valores dos índices calculados:

$\widehat{ÍNDICE}_j (I)$ e $\widehat{ÍNDICE}_j (\hat{I})$									
MUNICÍPIOS	BIOMA	ÁREA	POPULAÇÃO	DENSIDADE	PIB	IDH	I	\hat{I}	O
CARLINDA	A	2.416,14	10.990	4,59	11.626,21	0,67	-0,52	-0,57	1°
NOVO MUNDO	A	5.791,05	7.332	1,27	14.152,02	0,67	-0,48	-0,52	2°
NOVA BANDEIRANTES	A	9.593,27	11.643	1,21	12.713,91	0,65	-0,47	-0,52	3°
UNIÃO DO SUL	A	4.581,91	3.760	0,82	24.126,02	0,67	-0,45	-0,51	4°
NOVA GUARITA	A	1.114,17	4.392	4,43	13.879,57	0,67	-0,44	-0,49	5°
NOVA MONTE VERDE	A	5.150,56	8.093	1,54	15.110,04	0,69	-0,32	-0,33	6°
NOVA SANTA HELENA	A	2.375,58	3.468	1,47	18.624,77	0,71	-0,27	-0,27	7°
ITAÚBA	A	4.529,58	4.575	1,01	27.664,32	0,69	-0,23	-0,28	8°
TERRA NOVA DO NORTE	A	2.432,07	11.291	4,41	18.176,83	0,70	-0,19	-0,22	9°
NOVA CANAÃ DO NORTE	A	5.955,41	12.127	2,03	23.921,71	0,69	-0,12	-0,15	10°
CLÁUDIA	A	3.849,99	11.028	2,86	23.945,17	0,70	-0,11	-0,14	11°
PEIXOTO DE AZEVEDO	A	14.400,64	30.812	2,16	10.643,67	0,65	-0,10	-0,11	12°
ITANHAGÁ	A	2.898,07	5.276	1,82	29.837,47	0,65	-0,07	-0,11	13°
GUARANTÃ DO NORTE	A	4.734,59	32.216	6,80	16.236,84	0,70	0,14	0,13	14°
MARCELÂNDIA	A	12.273,55	12.006	0,98	19.129,38	0,70	0,14	0,12	15°
MATUPÁ	A	5.129,02	14.174	2,71	27.482,08	0,70	0,18	0,18	16°
APIACÁS	A	20.488,47	8.567	0,42	12.978,53	0,68	0,20	0,15	17°
SANTA CARMEM	A	3.855,36	4.085	1,06	47.565,45	0,72	0,25	0,19	18°
COLÍDER	A	3.103,96	30.766	9,95	21.645,43	0,71	0,31	0,27	19°
PARANAÍTA	A	4.796,01	10.684	2,23	81.433,04	0,67	0,40	0,21	20°
ALTA FLORESTA	A	8.953,19	49.164	5,48	20.853,81	0,71	0,56	0,61	21°
SINOP	A	3.942,23	113.099	28,69	33.807,26	0,75	2,14	2,12	22°

Fonte: Resultados da pesquisa.

O índice criado pela regressão resistente (Método *LTS*), para cada um dos municípios deste estrato, possui a seguinte forma funcional.

$$\hat{I} = -7,34 + (4,07 \cdot 10^{-5}) \cdot \text{ÁREA}_j + (1,12 \cdot 10^{-5}) \cdot \text{POPULAÇÃO}_j + (2,08 \cdot 10^{-2}) \cdot \text{DENSIDADE}_j + (1,06 \cdot 10^{-5}) \cdot \text{PIB}_j + 9,44 \cdot \text{IDH}_j + e_j \quad (21)$$

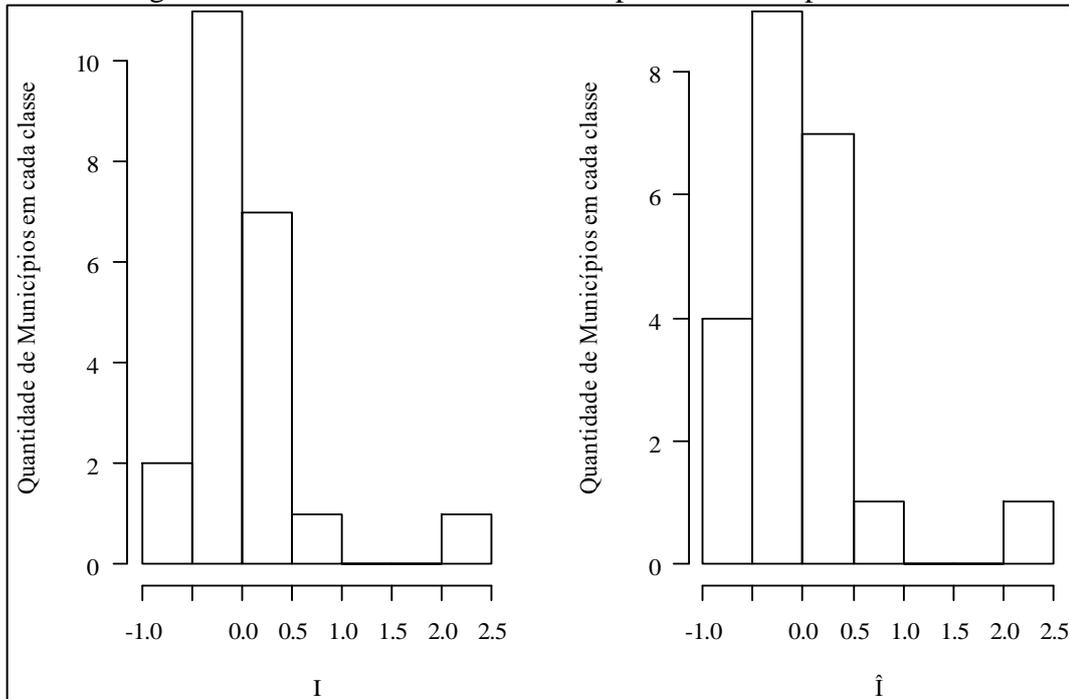
Equação 21: Fórmula/Índice Funcional para o Bioma Amazônia

Os valores de índices presentes na Tabela 3, para cada município, foram calculados através da substituição dos valores normalizados de cada uma das variáveis que os compõem nas respectivas fórmulas que definem os índices, desprezando a componente do erro e_j . Observa-se na Tabela 3 que, os dois índices, apesar de possuírem formulação distinta, geraram quase que a mesma ordenação de municípios e possuem valores numéricos bastante aproximados. Para efeitos de comparação a ordenação apresentada foi feita com base no índice I, desenvolvido na metodologia.

Pôde-se verificar também que os municípios que apresentaram pior e melhor desempenho, respectivamente, foram Carlinda e Sinop. E mais, verifica-se que Sinop apresenta um índice superior ao das demais cidades consideradas. Sinop possui índice cerca de 3,5 vezes maior que o penúltimo município da lista, que é Alta Floresta. Verificou-se que todas as variáveis apresentaram impactos positivos na totalização do índice estimado via *LTS* e que a variável mais impactante é o IDH.

O comportamento das distribuições dos índices criados, podem ser avaliados pela construção de seus respectivos histogramas, como mostra a Figura 6. Depreende-se da semelhança dos histogramas, a similaridade entre os valores gerados a partir de ambos os índices.

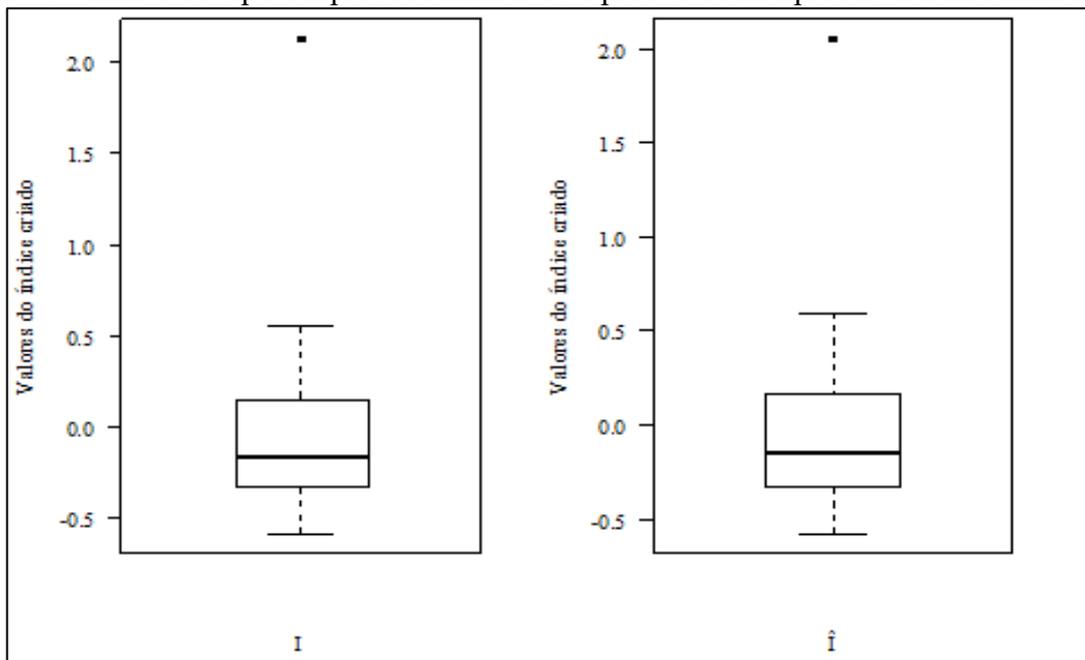
Figura 6: Histogramas dos valores dos índices I e \hat{I} para os municípios do bioma Amazônia.



Fonte: Resultados da pesquisa.

O comportamento de ambos os índices construídos pode ser melhor visualizado por meio do gráfico do tipo boxplot, apresentado na Figura 7.

Figura 7: Gráfico do tipo boxplot dos Índices I e \hat{I} para os Municípios do Bioma Amazônico

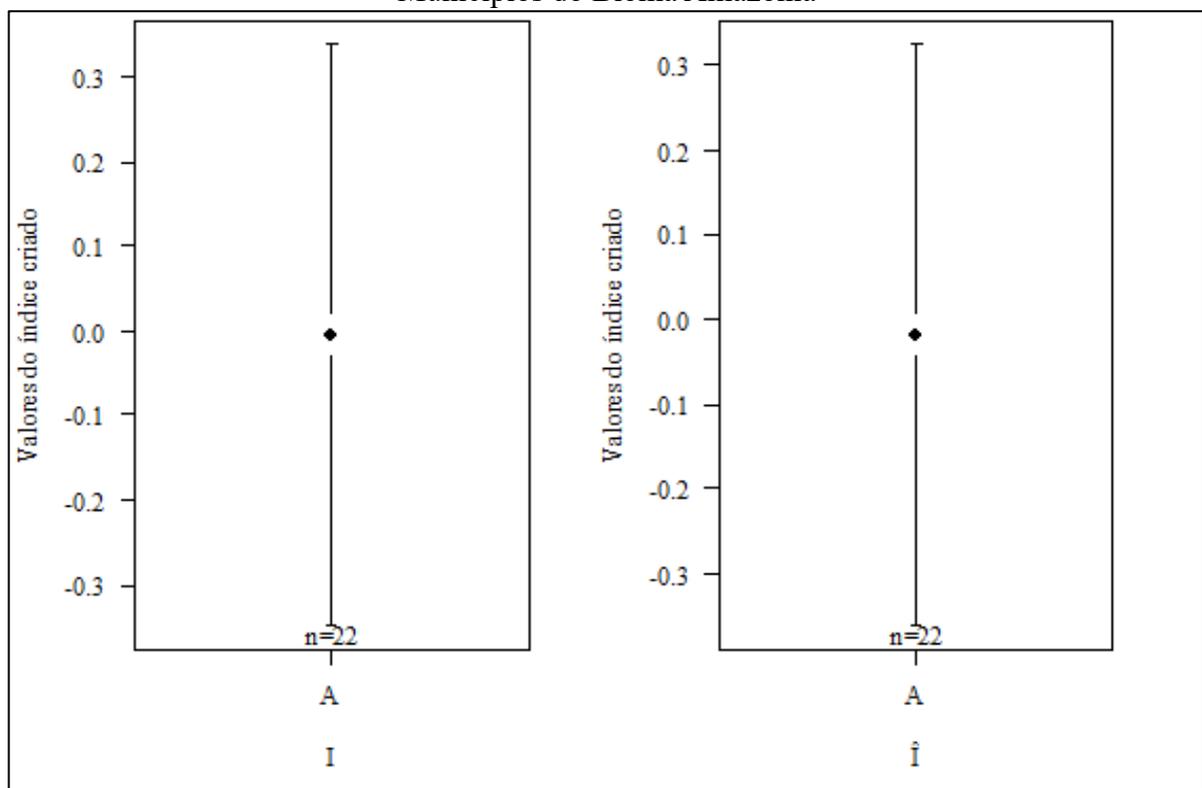


Fonte: Resultados da pesquisa.

Ao visualizar os gráficos boxplot pode verificar um ponto em sua porção superior. Tal ponto é relativo ao município de Sinop. Assim, infere-se que o município de Sinop se apresente como um *outlier* (ponto fora da distribuição dos dados) em termos de desenvolvimento quando comparado com a distribuição dos outros municípios. Verifica-se também que as distribuições dos dois índices são praticamente idênticas.

Com o intuito de ainda avaliar a distribuição dos dados dos índices criados, construiu-se também um gráfico do intervalo de confiança, com 1% de erro, para os índices criados (Figura 8).

Figura 8: Gráfico dos Intervalos de Confiança dos Índices I e \hat{I} , com erro de 1%, para os Municípios do Bioma Amazônia

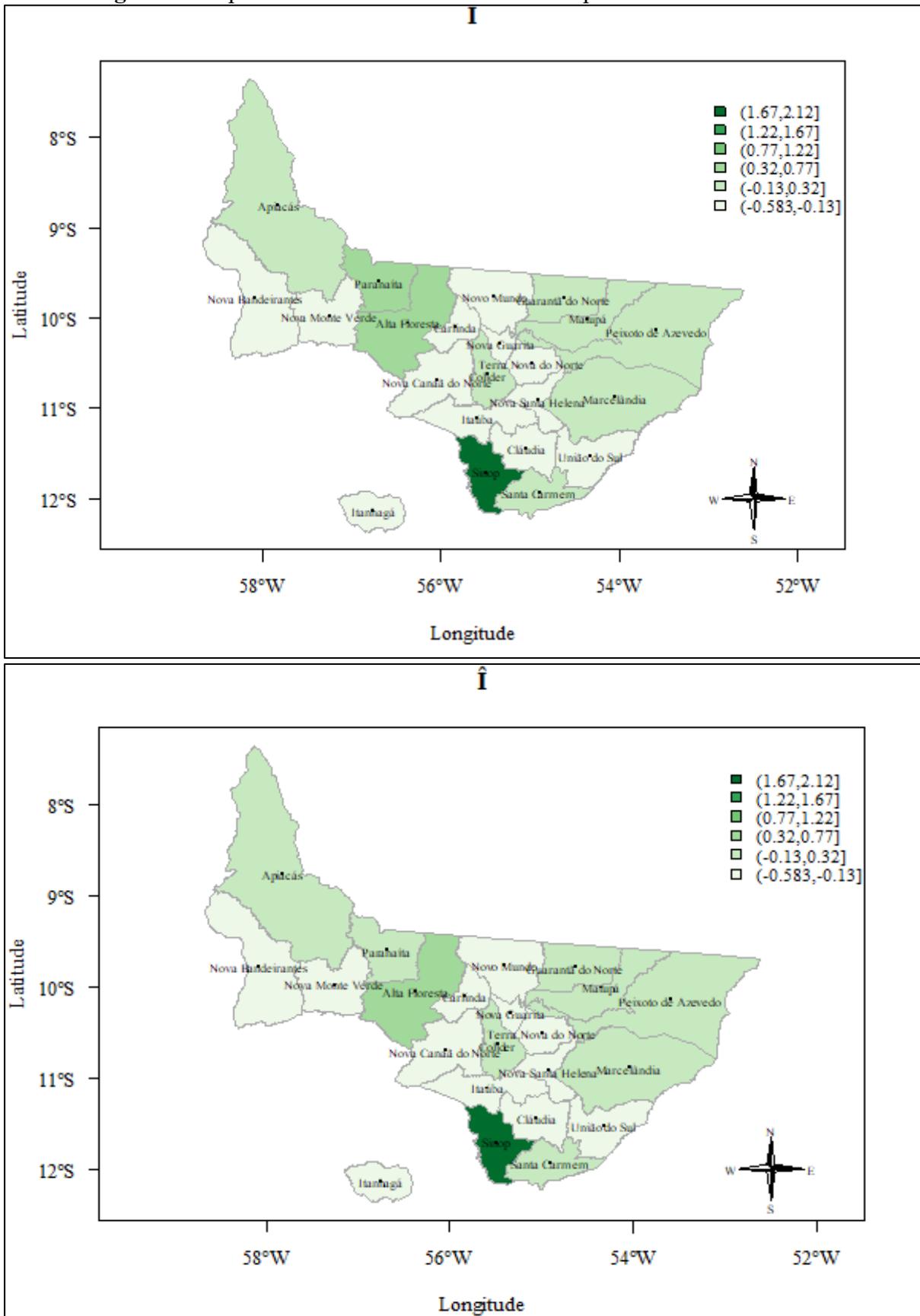


Fonte: Resultados da pesquisa.

Também nos gráficos dos intervalos de confiança (Figura 8) verifica-se a semelhança de comportamento dos índices gerados, uma vez que estes são quase idênticos.

Ainda com o intuito de analisar a semelhança e influência espacial dos índices no bioma amazônico, foram criados os mapas da Figura 9.

Figura 9: Mapas dos índices I e \hat{I} criados: Municípios do Bioma Amazônia.

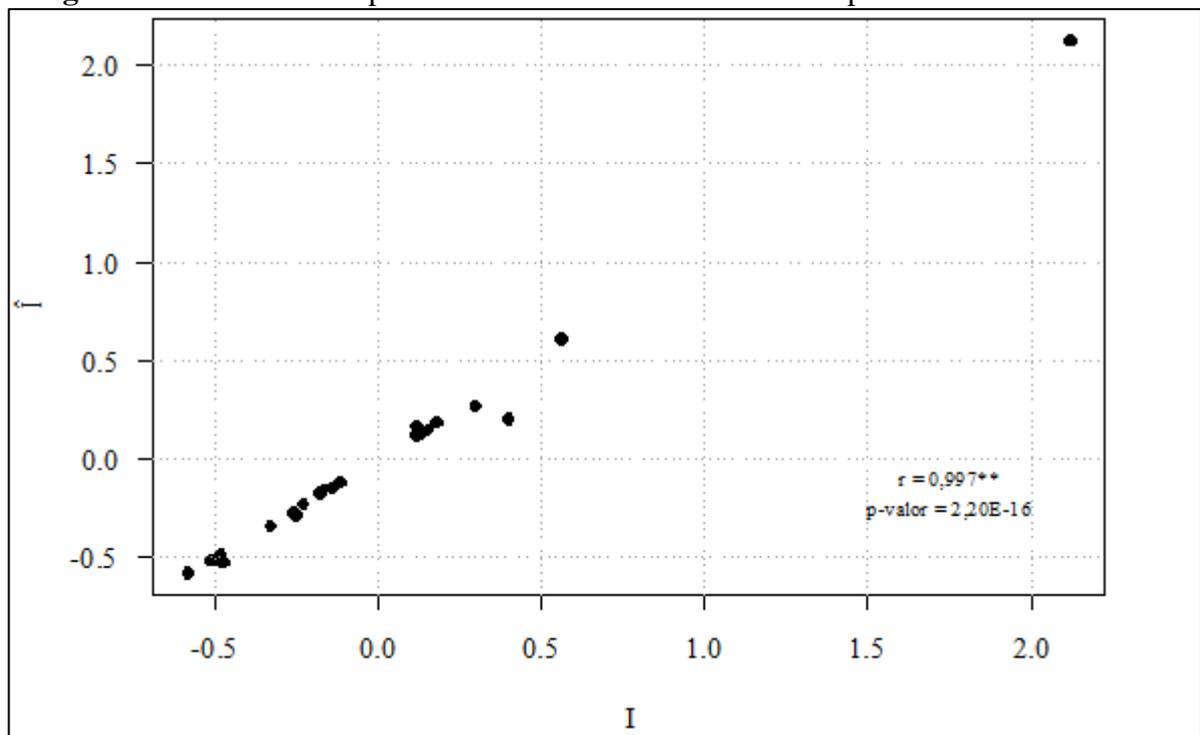


Fonte: Resultados da pesquisa.

Os mapas dos índices criados são praticamente os mesmos para o bioma Amazônia. Ou seja, ao se comparar as cores presentes em posições homólogas em ambos os mapas, infere-se que, ambos os índices geraram uma mesma estratificação especializada, relativa ao desenvolvimento dos municípios.

Com o intuito de verificar a similaridade dos índices ainda se recorreu ao gráfico de dispersão e ao coeficiente de correlação de Pearson. O gráfico de dispersão abaixo (Figura 10) revela a relação de associação entre os índices criados ($\widehat{ÍNDICE}_j (I)$ e $\widehat{ÍNDICE}_j (\hat{I})$) para o bioma Amazônia.

Figura 10: Gráfico de dispersão entre os Índices I e \hat{I} – Municípios do Bioma Amazônia



Fonte: Resultados da pesquisa.

Por meio do gráfico presente na Figura 10, pode-se verificar que os índices estão diretamente correlacionados. A realidade expressa pelo gráfico é corroborada pelo valor do coeficiente de correlação de Pearson, $r = 0,997$. Tal valor, quando da realização do teste de hipóteses, indicou que as séries apresentam alta correlação positiva e que os vetores de índice são bastante semelhantes e correlacionados ($p = 2,2 \times 10^{-16}$). Depreende-se de tais resultados que, tanto o índice I, como o \hat{I} , são válidos para a geração de uma hierarquização relativa ao desenvolvimento dos respectivos municípios do bioma Amazônia.

4.2.2 Área de transição Cerrado-Amazônia

Para o estrato Cerrado-Amazônia, os resultados das estatísticas descritivas podem ser visualizados na Tabela 4, considerando-se os 10 municípios selecionados.

Tabela 4: Valores das estatísticas de cunho descritivo das variáveis dos municípios do bioma Cerrado-Amazônia

CERRADO E AMAZÔNIA	ÁREA	POPULAÇÃO	DENSIDADE	PIB	IDH
MÉDIA	7.374,13	11.334,10	3,61	50.239,15	0,71
DESVIO PADRÃO	3.879,46	20.360,00	3,67	17.523,65	0,04

Fonte: Resultados da pesquisa.

Da análise dos resultados da Tabela 4 verifica-se que as maiores variações são alcançadas pelas variáveis População e PIB. Por outro lado, as menores dispersões ficam por conta do IDH e da Densidade Demográfica. A ordenação dos municípios, gerada a partir dos índices criados, pode ser vista na Tabela 5.

Tabela 5: Ordenação de Municípios que compõem o bioma Cerrado-Amazônia considerando os valores dos índices calculados:

\widehat{INDICE}_j (I) e \widehat{INDICE}_j (Î)									
MUNICÍPIOS	BIOMA	ÁREA	POPULAÇÃO	DENSIDADE	PIB	IDH	I	Î	O
VERA	CA	2.962,69	10.235	3,45	33.853,15	0,68	-0,69	-0,72	1°
SÃO JOSÉ DO RIO CLARO	CA	4.536,20	17.124	3,77	29.627,02	0,68	-0,57	-0,61	2°
NOVA MARINGÁ	CA	11.557,30	6.590	0,57	45.166,01	0,66	-0,41	-0,40	3°
FELIZ NATAL	CA	11.679,00	10.933	0,95	23.601,40	0,70	-0,38	-0,28	4°
TAPURAH	CA	4.491,05	10.392	2,30	53.550,07	0,67	-0,29	-0,30	5°
NOVA UBIRATÃ	CA	12.490,21	9.218	0,73	65.145,82	0,67	-0,05	-0,07	6°
IPIRANGA DO NORTE	CA	3.467,05	5.123	1,48	78.912,83	0,73	-0,05	-0,06	7°
NOVA MUTUM	CA	9.544,57	31.649	3,31	65.493,85	0,76	0,63	0,61	8°
LUCAS DO RIO VERDE	CA	3.683,59	45.556	12,43	49.953,90	0,77	0,83	0,74	9°
SORRISO	CA	9.329,60	66.521	7,13	57.087,49	0,74	0,97	0,75	10°

Fonte: Resultados da pesquisa.

A forma funcional do índice, estimada pelo método de regressão resistente *LTS*, é dada por:

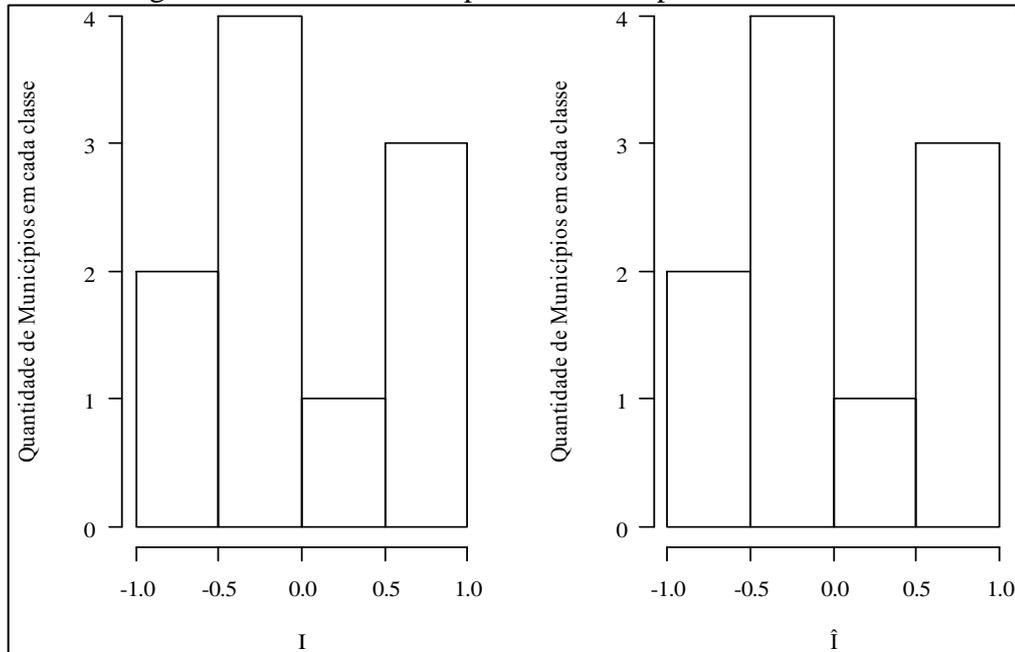
$$\hat{I} = -6,09 + (6,09 \cdot 10^{-5}) \cdot \text{ÁREA}_j + (4,54 \cdot 10^{-6}) \cdot \text{POPULAÇÃO}_j + (5,51 \cdot 10^{-2}) \cdot \text{DENSIDADE}_j + (9,21 \cdot 10^{-6}) \cdot \text{PIB}_j + 6,83 \cdot \text{IDH}_j + e_j \quad (22)$$

Equação 22: Fórmula/Índice Funcional para o Bioma Cerrado-Amazônia

Analogamente ao que ocorreu para o bioma Amazônia, com ordenação baseada no índice obtido na metodologia (I), observa-se que todas as variáveis impactam positivamente na construção do índice. Como no caso anterior, o destaque fica por conta da variável IDH, que apresenta o maior impacto na geração do índice Î. As cidades com maior e menor desempenho, respectivamente, por ambos os índices, são Sorriso e Vera. Da análise da penúltima linha da tabela verifica-se que o desempenho de Lucas do Rio Verde é muito parecido com o do município de Sorriso.

Cabe dizer que, uma vez mais, os valores dos índices criados se revelam bastante similares, conseqüentemente, a ordenação de municípios gerada, considerando-se tais índices, é a mesma. O comportamento das distribuições dos índices criados pode ser avaliado pela análise dos histogramas presentes na Figura 11.

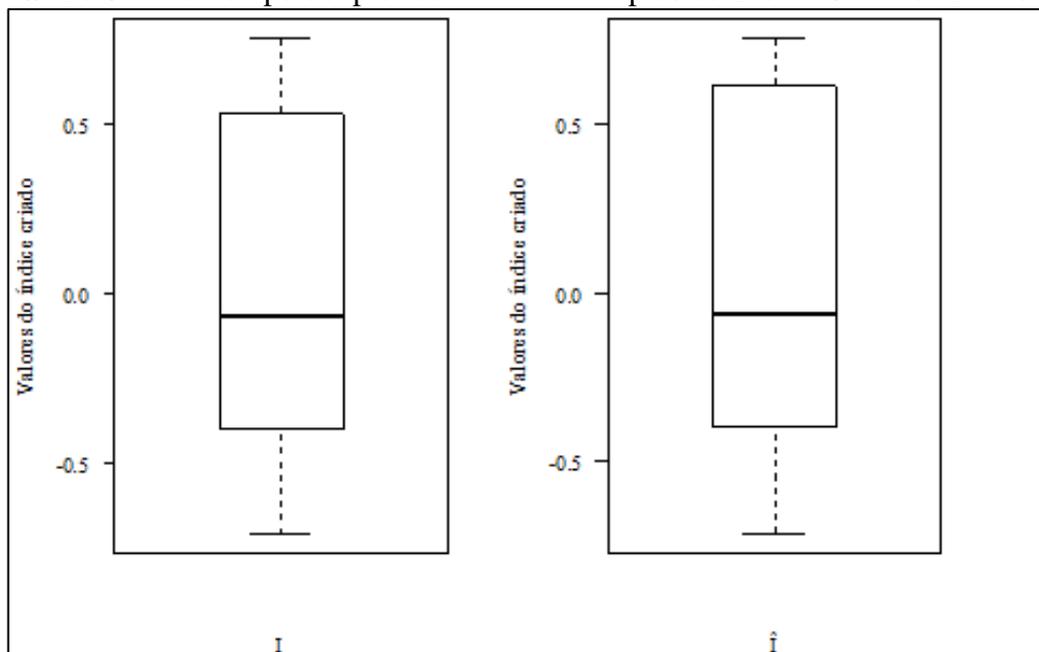
Figura 11: Histogramas dos Índices I e \hat{I} para os Municípios do Bioma Cerrado-Amazônia



Fonte: Resultados da pesquisa.

Ao se comparar os histogramas da Figura 11, verifica-se que estes são basicamente idênticos. Tal fato respalda a afirmação sobre a semelhança entre os valores gerados por ambos os índices, semelhança também revelada no gráfico tipo boxplot (Figura 12).

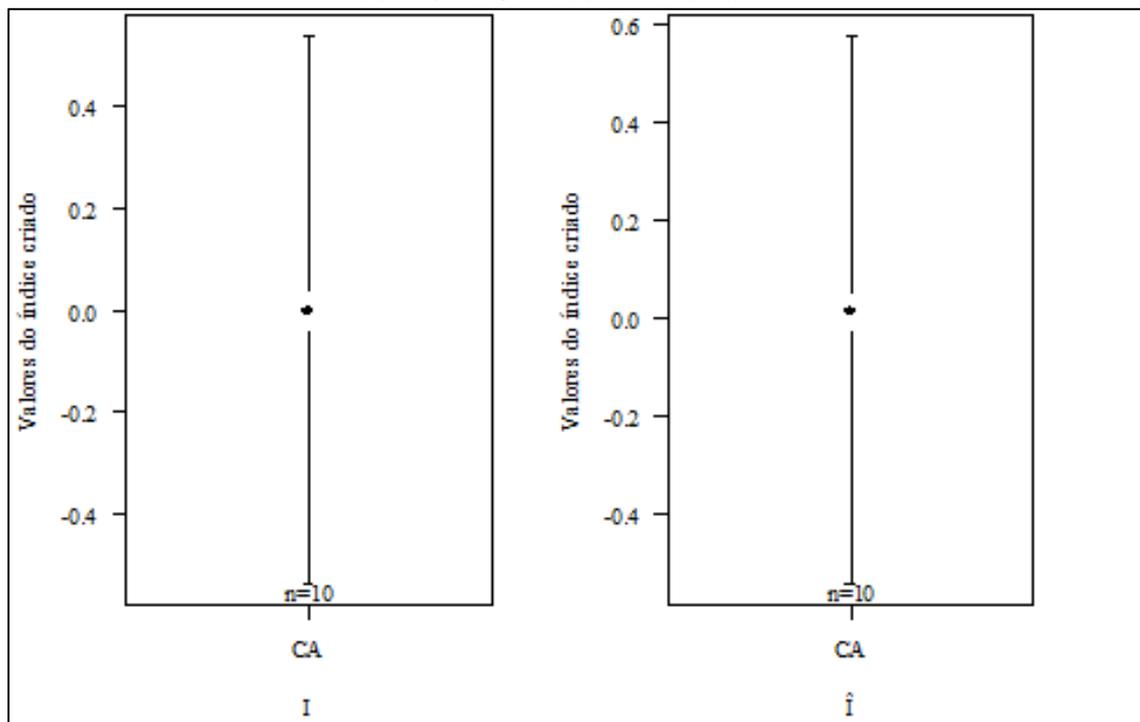
Figura 12: Gráfico do tipo boxplot de I e \hat{I} - Municípios do Bioma Cerrado-Amazônia



Fonte: Resultados da pesquisa.

Diferentemente do verificado para o bioma Amazônia, neste caso o desenvolvimento dos municípios avaliados é mais homogêneo, apesar da dispersão. Ou seja, verifica-se que a distribuição de valores relativos a ambos os índices se assemelha e não se verifica a ocorrência de *outliers*. Da mesma forma, com o intuito de avaliar a distribuição dos índices criados, construiu-se um gráfico do intervalo de confiança, com 1% de erro, para os índices criados (Figura 13).

Figura 13: Gráfico dos Intervalos de Confiança dos Índices I e \hat{I} , com erro de 1%: Municípios do Bioma Cerrado-Amazônia

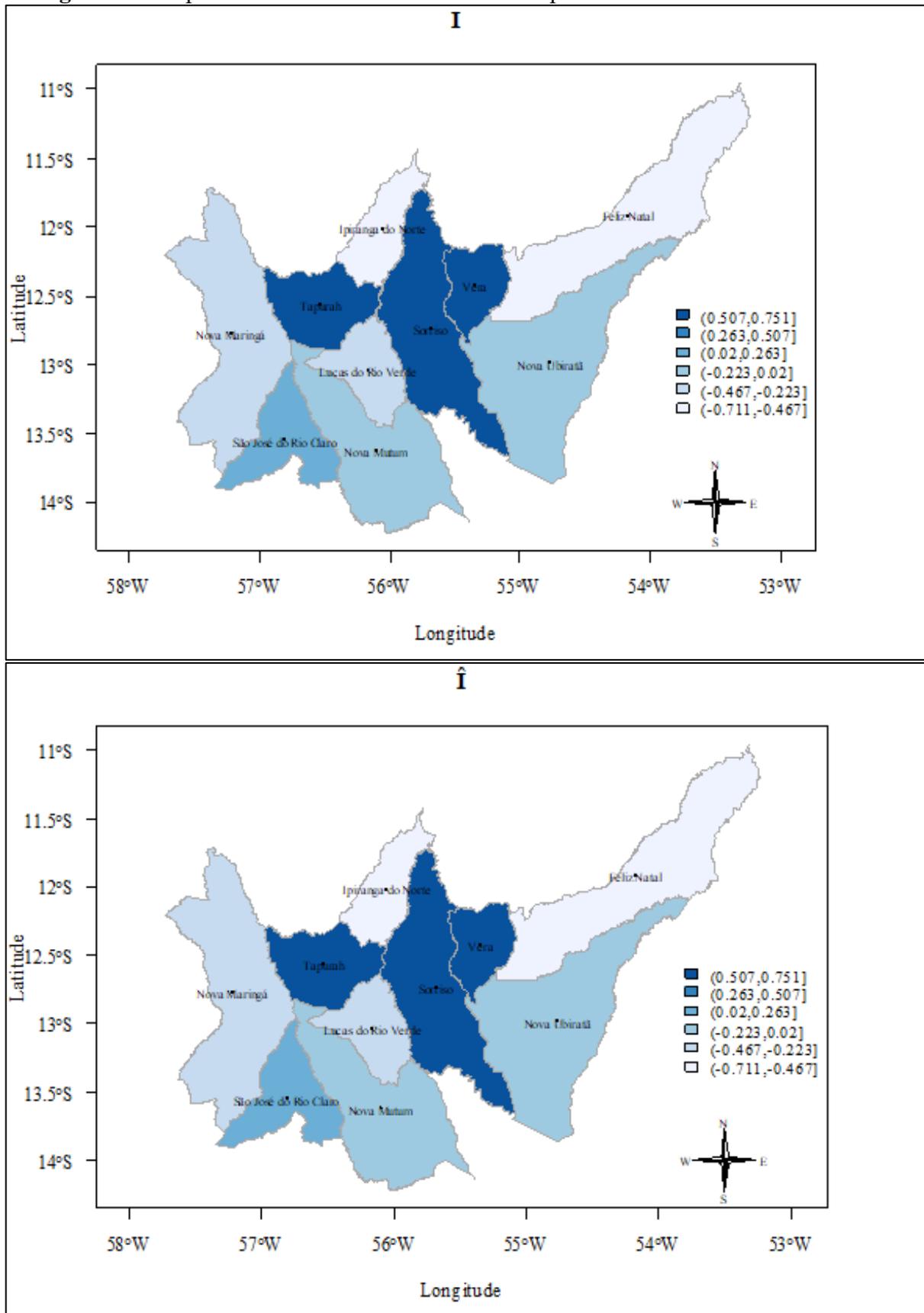


Fonte: Resultados da pesquisa.

De modo análogo ao verificado para os dados relativos ao bioma Amazônia, constatou-se, por meio da análise e comparação dos gráficos presentes na Figura 13, que os índices possuem comportamento bastante semelhante. Isto é, da semelhança dos gráficos, depreende-se que ambos os conjuntos de índices são igualmente confiáveis. Logo, por transitividade, a semelhança dos índices fica evidente.

Com a finalidade de verificar a semelhança na distribuição espacial dos valores gerados índices, foram criados a partir de tais valores, os mapas presentes na Figura 14.

Figura 14: Mapas dos Índices I e \hat{I} criados - Municípios do Bioma Cerrado-Amazônia.

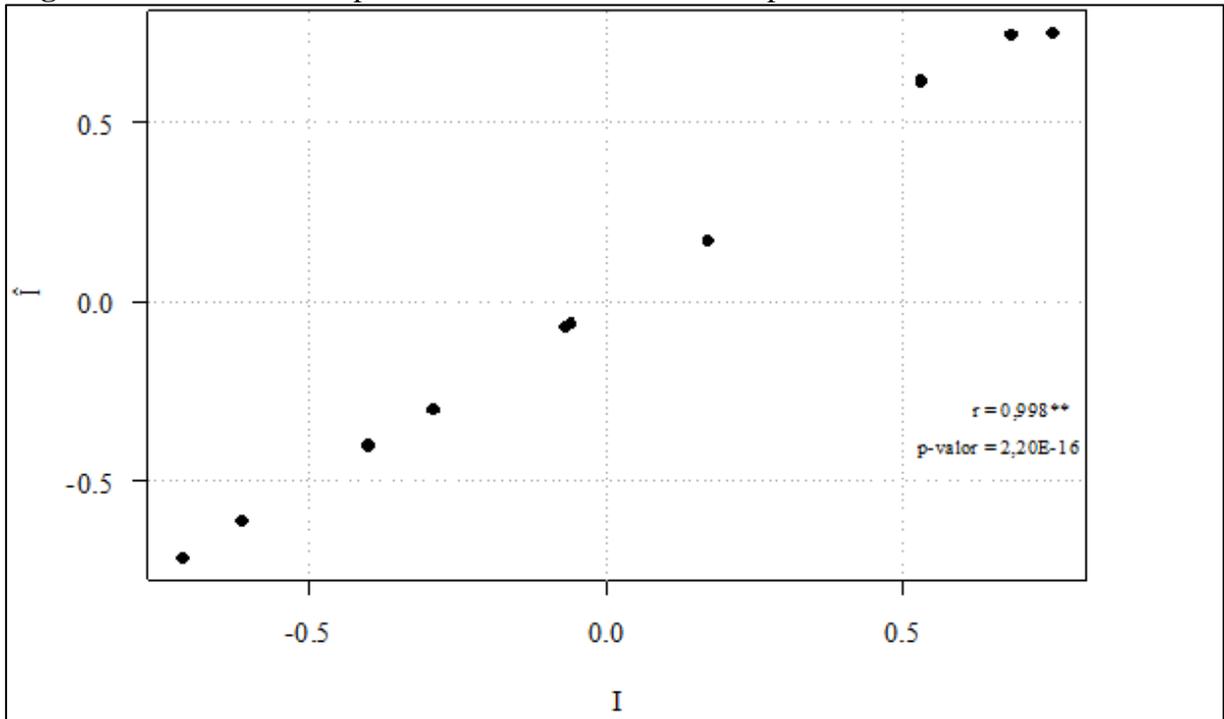


Fonte: Resultados da pesquisa.

Da comparação visual dos mapas apresentados na Figura 14, pode-se afirmar que, a semelhança entre os índices, constatada de modo numérico, mais uma vez é confirmada, agora do ponto de vista da distribuição espacial dos valores.

O gráfico de dispersão abaixo revela a relação de associação entre os índices criados para o bioma Cerrado-Amazônia.

Figura 15: Gráfico de Dispersão entre os Índices - Municípios do Bioma Cerrado-Amazônia



Fonte: Resultados da pesquisa.

Da análise do gráfico presente na Figura 15, verifica-se que os índices I e Î estão diretamente correlacionados. O índice de correlação de Pearson é bastante significativo, sendo de $r = 0,998$. Tais resultados, mostram a semelhanças dos índices construídos e validam estatisticamente o índice I criado. Quando da realização do teste de hipóteses, constatou-se correlação entre os vetores de índices, uma vez que $2,2 \times 10^{-16}$.

4.2.3 Bioma Cerrado

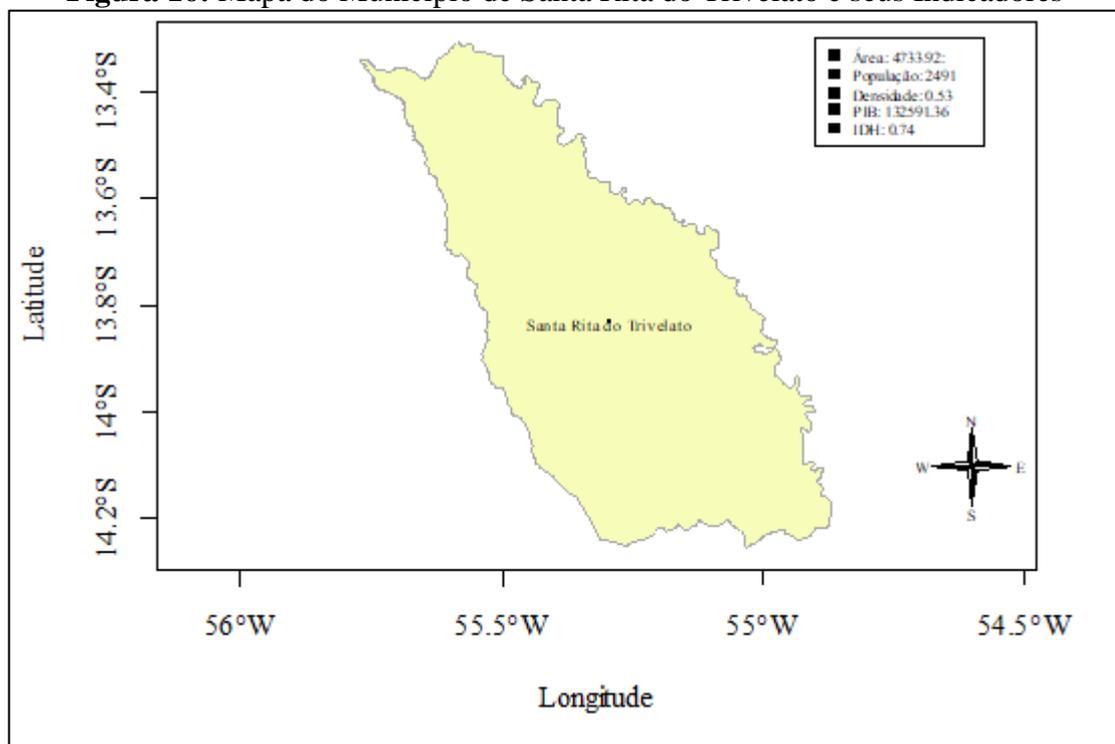
Cabe esclarecer que, como o citado anteriormente, houve uma única cidade no bioma Cerrado, Santa Rita do Trivelato. Tal cidade, possuía variáveis PIB e IDH próximos aos de Sorriso e Lucas do Rio Verde, que são as melhores cidades do bioma Cerrado-Amazônia. A Tabela 6 mostra os valores das variáveis em estudo.

Tabela 6: Análise Descritiva das variáveis do município do Cerrado.

MUNICÍPIOS	Nº	BIOMA	ÁREA	POPULAÇÃO	DENSIDADE	PIB	IDH
SANTA RITA DO TRIVELATO	118	C	4.733,92	2.491	0,53	132.591,36	0,74

Fonte: Resultados da pesquisa.

Ao se comparar os dados do município com os dos demais, conclui-se que este seja um município pequeno, mas que apresenta um desenvolvimento relevante e próximo àquele das cidades de maior porte do estado de Mato Grosso. A sua representação espacial se encontra no mapa da Figura 16.

Figura 16: Mapa do Município de Santa Rita do Trivelato e seus Indicadores

Fonte: Resultados da pesquisa.

Por ser Santa Rita do Trivelato a única cidade da amostra, relativa ao bioma Cerrado, não se aplicou a metodologia a ela.

5. CONCLUSÕES E RECOMENDAÇÕES

Nas seções seguintes são apresentadas as conclusões referentes ao trabalho executado e as recomendações para futuras pesquisas, as quais podem ser realizadas a partir do estágio já alcançado.

5.1 Conclusões

A pesquisa desenvolvida e aqui descrita teve como objetivo a criação de um índice que permitisse quantificar de modo holístico o desenvolvimento de municípios. Logo, tal índice de desenvolvimento foi criado considerando indicadores que não apenas os econômicos. Os indicadores que integraram o índice foram: área territorial do município, população, densidade demográfica, PIB e IDH. Quando da experimentação, por meio do cálculo do índice individualizado para cada município da amostra, foi possível construir a classificação hierárquica de 33 municípios do norte do Estado de Mato Grosso (amostra). E mais, considerando as afinidades entre os municípios, associadas aos biomas a que pertencem, e a necessidade de redundância de experimentação, resolveu-se dividir a amostra em três grupos de municípios, de acordo com o bioma predominante em seu território. Assim, gerou-se:

1. Um grupo relativo ao bioma Amazônia, com 22 municípios;
2. Um grupo inerente à área de transição (Cerrado-Amazônia), com 10 municípios; e
3. Um grupo com apenas um município, relativo ao bioma Cerrado.

O novo índice consistiu, inicialmente, da normalização/padronização das diferentes grandezas, ou indicadores, de modo que todas se apresentassem num mesmo intervalo ou *range*. Posteriormente, para cada município estudado, por média aritmética simples dos valores padronizados/normalizados dos indicadores (Área, População, Densidade, PIB e IDH) de cada município, chegou-se ao índice proposto. Considerando a modelagem desenvolvida e a busca feita por trabalhos relacionados, pode-se afirmar que o índice criado seja inédito. Em adição a isso, buscou-se validar o índice por meio de métodos estatisticamente tradicionais e já aceitos pela comunidade científica.

Assim, a fim de se evitar problemas de multicolinearidade das variáveis e inconsistências quanto aos pré-requisitos inerentes aos métodos estatísticos paramétricos, foram utilizados métodos de regressão não paramétricos nas variáveis padronizadas, ou seja, o Método dos Mínimos Quadrados Aparados (*Least Trimmed of Squares - LTS*), e, posteriormente, por meio de uma regressão resistente, foi gerado um segundo índice. Assim, a

discrepância, ou erro, entre os índices obtidos, foi utilizada para a análise comparativa de valores e a consequente validação do índice criado.

De posse das discrepâncias entre dos índices obtidos, e utilizando o *software* RStudio, foram construídos histogramas a fim de que se pudesse comparar o comportamento das distribuições dos índices criados. As distribuições dos valores dos índices foram verificadas também por meio da geração e análise de gráficos Boxplot e de intervalo de confiança.

Por outro lado, a análise comparativa da distribuição espacial dos valores dos índices calculados pôde ser efetuada de modo visual por meio de mapas coroplédicos construídos para ambos os índices.

A análise da dispersão relativa dos valores dos índices foi efetuada com o auxílio de gráficos de dispersão. E mais, por meio do coeficiente de correlação de Pearson, foi possível quantificar e averiguar a similaridade entre os vetores de índices gerados.

Após a geração dos resultados, estes foram confrontados a fim de se determinar o nível de similaridade entre os índices e, conseqüentemente, a validade do índice inédito criado.

Após as análises comparativas dos gráficos, mapas e valores numéricos, chegou-se às seguintes conclusões:

1. Apesar de formulação distinta dos índices, estes deram origem a classificações hierárquicas muito próximas aos municípios de cada subconjunto da amostra. Ou seja, os valores numéricos obtidos para os índices, considerando os mesmos municípios em cada subconjunto, foram bastante próximos ou coincidentes.
2. A análise dos histogramas gerados e dos gráficos tipo Boxplot, indicaram uma distribuição dos índices quase idêntica. Em adição a isso, os gráficos do tipo Boxplot relativos aos municípios do bioma Amazônia, construídos para cada um dos os índices, apresentam um município como *outlier* (Sinop). Ou seja, um ponto, relativo ao município de Sinop, encontra-se discrepante na distribuição dos dados no bioma Amazônia. No bioma Cerrado-Amazônia não houve a ocorrência de *outlier*, o que indica uma maior homogeneidade no tocante ao desenvolvimento dos municípios que compõem o subconjunto.
3. Depreende-se dos gráficos de intervalo de confiança conclusões semelhantes àquelas mencionadas até o momento. A distribuição espacial idêntica observada em cada par de mapas, inerentes a cada subconjunto de municípios, corrobora com a conclusão de semelhança entre os índices.

4. Por fim, do gráfico de dispersão, infere-se que os índices obtidos estão diretamente correlacionados e o coeficiente de correlação de Pearson quantificou a semelhança dos vetores de dados. Ou seja, o coeficiente de correlação alcançou $r = 0,997$, para o bioma Amazônia, e $r = 0,998$, para área de transição (bioma Cerrado-Amazônia). Em ambos os casos, o p-valor foi $p = 2,2 \times 10^{-16}$, o que confirma a correlação dos índices.

Em virtude do subconjunto que compõem o bioma Cerrado ser unitário, não se aplicou a metodologia para este bioma.

Em suma, com base nos resultados e análises estatísticas aqui apresentadas, conclui-se que o índice de desenvolvimento criado, apesar de sua formulação relativamente simples e concisa, atinge os objetivos almejados, se equiparando ao índice gerado via *LTS*. Assim, verifica-se que tal formulação, além de eficiente, é relativamente simples e pode ser utilizada por qualquer pesquisador ou interessado, sem que cálculos complexos e extenuantes sejam efetuados. Entende-se que esta característica da metodologia seja positiva, pois, mesmo usuários com conhecimento mínimo em Estatística, poderiam se valer do índice criado.

5.2 Recomendações

Como principais recomendações para trabalhos futuros, tem-se:

- Sugere-se que a metodologia seja aplicada a uma quantidade maior de municípios, a fim de se ter subconjuntos mais numerosos e que seja possível verificar o comportamento dos índices em situações adversas àquelas aqui apresentadas;
- Considerando a formulação atual do índice proposto, poder-se-ia reproduzir o experimento previsto na metodologia, inserindo em sua formulação uma quantidade maior de indicadores. Assim, após o processo de validação, ter-se-ia uma nova avaliação do comportamento dos índices diante das alterações efetuadas.
- Uma vez que o índice proposto possui em sua formulação a média aritmética simples, poder-se-ia utilizar a média ponderada em lugar desta. Assim, seria possível ponderar indicadores que eventualmente sejam considerados como mais relevantes na avaliação de desenvolvimento de um município e, conseqüentemente, ter um índice de desenvolvimento mais realístico.
- Considerando a maleabilidade do índice proposto e do índice gerado via *LTS*, sugere-se que a metodologia seja aplicada ao ambiente escolar, por meio da substituição dos atuais indicadores por indicadores inerentes àquele ambiente.

REFERÊNCIAS BIBLIOGRÁFICAS

- ASSIS, J, P. et al. Estatística Descritiva. 1ª ed. Piracicaba – São Paulo: FEALQ, 2016.
- BARBETTA, P, A. Estatística Aplicada às Ciências Sociais. 7ª ed. Editora da UFSC: Série Didática, 2007.
- BARBETTA, P, A; REIS, M, M; BORNIA, A, C. Estatística para Cursos de Engenharia e Informática. 3ªed. São Paulo: Editora Atlas, 2010.
- BECKER, Bertha K. **Amazônia**. 5ª ed. São Paulo: Ática, 1997.
- _____. Geopolítica da Amazônia. *Estudos Avançados*. v .19, n. 53, 2005.
- _____. Logística e nova configuração do território: que geopolítica será possível? In: DINIZ, C. C. (Org.). **Políticas de desenvolvimento regional: desafios e perspectivas à luz das experiências da União Europeia e do Brasil**. Brasília: Editora da UNB, 2007, p. 267-299.
- BECKER, Olga Maria Schild. Mobilidade espacial da população: conceitos, tipologia, contextos. In: CASTRO, Iná Elias (Org.) **Explorações geográficas: percursos no fim do século**. Rio de Janeiro: Bertrand Brasil, 1997.p. 319- 367.
- BUSSAB, W; MORETTIN, P.A. Estatística Básica. 6ª ed. São Paulo: Saraiva. 2002.
- CASELLA, G; BERGER, R. L. Statistical Inference. 2nd ed. USA: Thomson Learning. 2002.
- COSTA NETO, P. L. O. Estatística. 3ªed. São Paulo: Blucher, 2002.
- CUNHA, U. S; MACHADO, S. A; FILHO, A. F. Uso de análise exploratória de dados e de regressão robusta na avaliação do crescimento de espécies comerciais de terra firme da Amazônia. *Revista Árvore*, Viçosa-MG, v.26, n.04, p.391-402, 2002.
- CRAWLEY, M. J. The R book. John Wiley & Sons, 1ed, Chichester. 2007.
- DE GROOT, MORRIS H. - Probability and statistics. 2nd ed Massachusetts Addison Wesley, 1986.
- DEVORE, J. L. Probabilidade e estatística: para engenharia e ciências. 6ª ed. São Paulo. Cengage Learning, 2011.
- DRAPER, N. R.; SMITH, H. Applied regression analysis. 2.ed. New York: John Wiley, 1981. 709 p.

EVERITT, B.; HOTHOR, B. An Introduction to Applied Multivariate Analysis with R. Use R! Springer. 2011.

GEOGRAFOS. Extensão Territorial dos Estados Brasileiros. Informações Disponível em: <http://www.geografos.com.br/extensao-territorial/> . Acesso em 18/04/2018.

GERVINI, D.; YOHAJ, V. J. A class of robust and fully efficient regression estimates. Mimeo, Universidad de Buenos Aires, 1999.

HAIR, J. F. Jr; BLACK, W. C; BABIN, B. J; ANDERSON, R. E. (2009). Multivariate Data Analysis. 7th ed. Prentice Hall: Upper Saddle River, NJ.

HESPAHOL, Antônio Nivaldo. A expansão da agricultura moderna e a integração do Centro-oeste brasileiro à economia nacional. **Caderno Prudentino de Geografia**. Presidente Prudente, AGB, n° 22, julho 2000.

HOFFMANN, R. Estatística para Economistas. 4ed. São Paulo: Pioneira Thomson Learning, 2006.

HOTHORN, T.; EVERITT, B.S. A Handbook of Statistical Analyses using R, 3rd. Chapman & Hall/CRC, New York. (2010).

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Brasil/ Mato Grosso. Disponível em: <https://cidades.ibge.gov.br/brasil/mt/panorama> . Acesso em 20/02/2018

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Brasil/ MT/ Araguainha. Disponível em: <https://cidades.ibge.gov.br/brasil/mt/araguainha/> . Acesso em 18/04/2018.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Brasil/MT/Cuiabá. Disponível em: <https://cidades.ibge.gov.br/brasil/mt/cuiaba/>. Acesso em 18/04/2018.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Brasil/ MT/ Rondonópolis. Disponível em: <https://cidades.ibge.gov.br/brasil/mt/rondonopolis/>. Acesso em 18/04/2018.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Brasil/ MT/ Sorriso. Disponível em: <https://cidades.ibge.gov.br/brasil/mt/sorriso/>. Acesso em 18/04/2018.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Brasil/ MT/ Sinop. Disponível em: <https://cidades.ibge.gov.br/brasil/mt/sinop/>. Acesso em 18/04/2018.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Brasil/MT/Várzea Grande. Disponível em: <https://cidades.ibge.gov.br/brasil/mt/varzea-grande/>. Acesso em 18/04/2018.

- LANDIM, P. M. B. Análise estatística de dados geológicos multivariados. São Paulo: Oficina de Textos, 2011.
- MANLY, B. F. J. Métodos estatísticos multivariados: uma introdução (tradução Sara Ianda Correa Carmona). 3ed. Porto Alegre: Bookman, 229p, 2008.
- MARONNA, R. A; YOHAI, V. J. Robust regression with both continuous and categorical predictors. *Journal of Statistical Planning and Inference* 89, 197-214, 2000.
- MENDES, M. A. *História e geografia de Mato Grosso*. Cuiabá: Cafarnaum, 2012.
- MEYER, P, L. Probabilidade: aplicações à estatística. Rio de Janeiro: LTC, 2009.
- MONTGOMERY, D. C; RUNGER, G. C. Estatística Aplicada e Probabilidade para Engenheiros. 5ª ed. Rio de Janeiro. LTC, 2012.
- OLHAR DIRETO. Soma do PIB de Cuiabá e Municípios do Agronegócio. Disponível em: <http://www.olhardireto.com.br/agro/noticias/exibir.asp?id=24537¬icia=cuiaba-e-municipios-do-agronegocio-somam-546-do-pib-de-mato-grosso> . Acesso: 17/04/2018.
- SARTORIS, A. Estatística e Introdução à Econometria. São Paulo: Saraiva, 2003.
- SPIEGEL, M, R. Estatística. 3ªed.São Paulo: Pearson Makron Books, 1993.
- TRIOLA, M, F. Introdução à Estatística. 7ed. Rio de Janeiro: LTC, 1999.
- PIAIA, Ivane Inês; Geografia de Mato Grosso – 3. Ed. Ver. Amp. Cuiabá: EDUNIC, 2003.
- PIMENTEL GOMES, F. Curso de Estatística Experimental. 14 ed. Piracicaba: Edição do autor, 477p, 2000.
- PINTO, M. O. **Entre o sonho e a realidade:** desafios da política ambiental na Amazônia. 170 p. Dissertação de Mestrado – Universidade de Brasília. Centro de Desenvolvimento Sustentável, 2006.
- PORTAL ACTION. Estatística Básica – Medidas de Dispersão. Disponível em: <http://www.portalaction.com.br/estatistica-basica/21-medidas-de-posicao>. Acesso em 24/03/2018.
- PORTAL ACTION. Estatística Básica – Medidas de Dispersão. Disponível em: <http://www.portalaction.com.br/estatistica-basica/22-medidas-de-dispersao>. Acesso em 24/03/2018.

YOHAI, V.; STAHEL, W. A.; ZAMAR, R. H. A procedure for robust estimation and inference in linear regression. Springer-Verlag, 1991.

R Development Core Team (2005). R: A Language and Environment for Statistical Computing R Foundation for Statistical Computing, Viena, Austria. ISBN 3-900051-07-0, URL: <http://www.R-project.org>

ROUSSEEUW, P. J. Least median of squares regression. Journal of the American Statistical Association, n. 79, p. 871-880. 1984.

SANTOS, Luiz Erardi. Raízes da História de Sinop. Grafitec: Sinop, 2007.

SEPLAN-MT. Secretaria de Planejamento do Estado de Mato Grosso - **Anuário 2007**. Disponível em: <<http://www.zsee.seplan.mt.gov.br/anuario2007/index.html>

SEPLAN – MT. Secretaria de Planejamento – Mato Grosso. Produto Interno Bruto de Mato Grosso e dos Municípios. Disponível em: <http://www.seplan.mt.gov.br/-/3934855-pib-mato-grosso-e-municipios> . Acesso em 20/02/2018.

SILVA, A. R. Métodos de análise multivariada em R. Piracicaba: FEALQ, p167, 2016.

SOUZA, E. A. **Processo de colonização da Região Norte de Mato Grosso e a construção de suas práticas culturais**. Sinop: CEACD/ UNEMAT, 2002. Produção gráfica FESPE.

TOLEDO et all, Geraldo Luciano; Ivo Izidoro Ovalle. **Estatística básica**. 2ª Edição – São Paulo: Atlas, 1985.