

Elias Silva de Medeiros<sup>1</sup> | Jaqueline Andrade<sup>2</sup> | Marcos Antonio Alves<sup>3</sup> | Carolina Cristina Bicalho<sup>4</sup>

# UMA APLICAÇÃO DA ESTATÍSTICA MULTIVARIADA AOS INDICADORES SOCIOECONÔMICOS DO BRASIL

AN APPLICATION OF MULTIVARIATE STATISTICS TO SOCIOECONOMIC INDICATORS IN BRAZIL

UNA APLICACIÓN DE ESTADÍSTICA MULTIVARIADA A INDICADORES SOCIOECONÓMICOS EN BRASIL

## RESUMO

A desigualdade regional no Brasil tem sido investigada com a perspectiva de prover caminhos à administração pública para reduzir as disparidades socioeconômicas. A fim de promover uma pesquisa mais ampla, este artigo teve como objetivo investigar a relação entre as unidades federativas do país e buscar entender em quais aspectos socioeconômicas as regiões se diferem ou se assemelham. Para isso, um conjunto de indicadores da dimensão socioeconômica foi levado em consideração: percentual de analfabetos, renda média mensal, acesso a água potável, domicílios com acesso a instalações de esgoto, escolaridade média, taxa de desemprego, domicílios com acesso a saneamento básico, taxa de mortalidade infantil e esperança de vida ao nascer. Uma análise estatística foi conduzida verificando a normalidade entre as variáveis. Por meio da análise de componentes principais constatou-se o número fatores iguais a dois, uma vez que as duas componentes principais explicaram 78,51% da variabilidade dos dados. Em seguida, com o intuito de investigar a similaridade entre as unidades federativas do Brasil foi realizada uma análise de clusters, em que foi determinada uma construção de três grupos, sendo um deles formado unicamente pelo Distrito Federal, outro grupo com Estados da região Norte e Nordeste e o último com Estados da região Centro-Oeste, Sul e Sudeste. Ao evidenciar a diversidade e complexidade das realidades regionais brasileiras, este estudo destaca-se como uma ferramenta adicional para orientar estratégias eficazes de desenvolvimento regional. Espera-se que os resultados apresentados inspirem novas investigações sobre as desigualdades socioeconômicas.

Palavras-chave: Mortalidade infantil. Saneamento básico. Desigualdade regional.

## ABSTRACT

Regional Inequalities in Brazil has been investigated with the perspective of providing ways to the public administration to reduce the existing socioeconomic inequality. With the purpose of promoting a broader research, this paper aimed to investigate the relationship between the federal units of the country and seek to understand in which aspects the regions differ or resemble each other. For this, a set of indicators of the socioeconomic dimension was taken into account: percentage of illiterates, average monthly income, access to drinking water, households with access to sewage facilities, average schooling, unemployment rate, households with access to basic sanitation, infant mortality rate and life expectancy at birth. A statistical analysis was carried out verifying the normality between the variables. Through the multivariate analysis it was found that the number of factors could be reduced to two main components, being able to explain 78.51% of the variability of the data. The values found in the factor analysis were used to perform a cluster analysis. Three groups were clearly grouped, one of them being formed solely by the Federal District, another group with States of the North and Northeast, and the latter with States of the Midwest, South and Southeast. The results are similar to those reported in the literature and show territorial inequality, as well as provide guidelines for directing public policies aimed at reducing disparities between the regions of the country.

**Keywords:** Infant mortality. Basic sanitation. Regional inequality

## RESUMEN

Las Desigualdades Regionales en Brasil han sido investigadas con la perspectiva de proporcionar formas a la administración pública para reducir la desigualdad socioeconómica existente. Con el propósito de promover una investigación más amplia, este artículo tuvo como objetivo indagar en la relación entre las unidades federales del país y buscar comprender en qué aspectos las regiones se diferencian o se parecen entre sí. Para ello se tuvo en cuenta un conjunto de indicadores de la dimensión socioeconómica: porcentaje de analfabetos, ingreso mensual promedio, acceso a agua potable, hogares con acceso a alcantarillado, escolaridad promedio, tasa de desempleo, hogares con acceso a saneamiento básico, población infantil tasa de mortalidad y esperanza de vida al nacer. Se realizó un análisis estadístico verificando la normalidad entre las variables. Mediante el análisis multivariado se encontró que el número de factores se pudo reducir a dos componentes principales, pudiendo explicar el 78,51% de la variabilidad de los datos. Los valores encontrados en el análisis factorial se utilizaron para realizar un análisis de conglomerados. Se agruparon claramente tres grupos, uno de ellos formado únicamente por el Distrito Federal, otro con los Estados del Norte y Noreste, y el último con los Estados del Medio Oeste, Sur y Sudeste. Los resultados son similares a los reportados en la literatura y muestran la desigualdad territorial, además de brindar lineamientos para direccionar políticas públicas encaminadas a reducir las disparidades entre las regiones del país.

**Palabras clave:** Mortalidad infantil. Saneamiento básico. Desigualdad regional

## 1. INTRODUÇÃO

A desigualdade territorial no Brasil tem sido um dos tópicos de estudo em políticas públicas. Desigualdade territorial é, de acordo com Araújo e Flores (2017) 'a concentração desigual de indivíduos vulneráveis em unidades territoriais comparáveis'. Esta vulnerabilidade se dá, especialmente, pelas oportunidades dadas à sociedade de usufruir de serviços públicos. Estes, por sua vez, podem estar relacionados com indicadores socioeconômicos tais como renda média, água potável, escolaridade, saneamento e mortalidade infantil.

O desequilíbrio regional é um problema que remete às origens do Brasil (ARAÚJO; FLORES, 2017). Todavia, conforme apontaram Mauro et al. (2018), mudanças significativas ocorridas a partir da década de 1990 levaram ao agravamento das disparidades no desenvolvimento nacional. Neste âmbito, os autores sugeriram criar e concretizar a 'cultura de desenvolvimento' por meio da elevação do estado de bem-estar das pessoas.

Pesquisas relacionadas vêm sendo propostas e elas têm se apoiado nas diferentes dimensões do desenvolvimento e das políticas públicas, como aquelas descritas por Mauro et al., 2018, a fim de analisar em quais aspectos essas regiões se diferem ou se assemelham. Na dimensão econômica, Piacenti C. e Piacenti S. (2018) analisaram um conjunto de 42 variáveis do potencial de desenvolvimento de municípios paranaenses por meio de análise fatorial. O estudo apontou diferentes grupos de municípios com variados potenciais de desenvolvimento. Silva, Sousa e Araujo (2017) conduziram uma pesquisa com o objetivo de apresentar novas perspectivas para a compreensão da pobreza multidimensional na região Norte do Brasil nos anos de 2006 a 2013. As dimensões Alimentos e Água, Comunicação e Informação, Educação, Condições de Moradia, Saúde e Trabalho e Demografia foram analisadas e os resultados revelaram a elevada pobreza no Norte e Nordeste e que os poucos avanços que existem nesta região não são homogêneos entre as unidades federativas.

Comumente as pesquisas que investigam a desigualdade territorial no Brasil ou se apoiam no Índice de Desenvolvimento Humano (IDH) e Produto Interno Bruto (PIB), vide Barbosa et al. (2015), ou focam a análise em apenas algumas regiões ou microrregiões, como Tavares e Porto Junior (2011), Piacenti C. e Piacenti S. (2018) e Silva, Sousa e Araujo (2018). Embora tenham apresentado limitações referentes a abrangência geográfica ou aos fatores investigados, elas forneceram importantes contribuições para as lideranças públicas. Barbosa et al. (2015) apontaram que os elevados índices de IDH e PIB da região Sudeste 'não se traduzem em qualidade de vida para todos os seus habitantes'. Resultados similares para os municípios paranaenses foram apontados por Vidigal, Amaral e Silveira (2012) e Piacenti C. e Piacenti S. (2018). Esses autores indicaram que há uma distribuição de riqueza desigual entre os municípios, embora o estado do Paraná possua uma posição confortável comparado aos demais do país. Também na região Sul, mais especificamente em Santa Catarina, Tavares e Porto Junior (2011) apontaram desigualdades intra e inter-regionais considerando-se dois índices, sendo o primeiro o de desenvolvimento econômico e gestão de finanças públicas e o outro de urbanidade e qualidade de vida. Silva, Sousa e Araujo (2017), por sua vez, indicaram que a pobreza multidimensional, evidenciada na região Norte, pode ser reduzida com melhor distribuição de recursos nas áreas da educação, trabalho e demografia, comunicação e informação e saúde.

Uma vez destacadas que cada região possui suas especificidades, a justificativa para este estudo se alinha aos pressupostos de pesquisas correlatas como as de Tavares e Porto Junior (2011), Vidigal, Amaral e Silveira (2012) e Piacenti C. e Piacenti S. (2018). Os argumentos se baseiam na necessidade de uma análise mais ampla, uma vez que a administração pública deve levar em consideração as disparidades regionais. Ela deve incluir, portanto, diferentes indicadores e levar em consideração todas as unidades territoriais do país.

O objetivo desta pesquisa é, portanto, investigar a relação entre as unidades federativas do Brasil considerando um conjunto de indicadores da dimensão socioeconômica com vistas a fornecer padrões para melhor planejamento de políticas públicas. Com isso, espera-se que os resultados possam captar as possíveis variações de desenvolvimento econômico regional, bem como indicar similaridades e dissimilaridades entre as regiões.

O restante deste artigo é organizado da seguinte forma: A Seção 2 apresenta os dados utilizados e a fonte da busca, bem como os métodos de análise empregados. A Seção 3 descreve os resultados obtidos e associa-os com outras pesquisas na área. A Seção 4 faz as considerações finais promovendo uma breve contextualização sobre as regiões estudadas e dá indicações para a promoção de ações de políticas públicas. As referências são listadas em seguida.

## 2. MATERIAL E MÉTODOS

Os indicadores utilizados foram: percentual de analfabetos, renda média mensal, acesso a água potável (identificada ao longo do trabalho apenas por água), percentual de domicílios com acesso a instalações de esgoto (AIE), escolaridade média, taxa de desemprego, proporção de domicílios com acesso a saneamento básico adequado (PMASB), taxa de mortalidade infantil (TMI) e esperança de vida ao nascer (EVN). Estes indicadores foram utilizados em outras pesquisas, como aquelas propostas por Barbosa et al. (2015) e Piacenti C. e Piacenti S. (2018).

Os dados utilizados são referentes ao ano de 2014. A escolha desse período se deve ao fato de que alguns dados são divulgados em períodos fixos e outros em datas aleatórias. Para que todas estas variáveis pudessem ser analisadas em conjunto, o ano de 2014 foi aquele que contém os dados completos mais recentes. O conjunto de dados é proveniente do Instituto de Pesquisa Econômica Aplicada (IPEA) e do Instituto Brasileiro de Geografia e Estatística (IBGE).

Alguns indicadores foram retirados da Pesquisa Nacional por Amostra de Domicílios (PNAD), cujos dados são de abrangência nacional, permitindo uma constante investigação sobre o desenvolvimento socioeconômico do Brasil por meio de indicadores como emprego e rendimento. Desta pesquisa foram retirados os seguintes indicadores: Analfabetos (Taxa de Analfabetismo), cujo valor é dado por uma porcentagem de pessoas de um grupo etário em relação ao total de pessoas do mesmo grupo etário, e mostra o percentual de pessoas de 15 anos ou mais de idade que não sabem ler nem escrever um bilhete simples; Renda (Renda mensal média por habitante) e Escolaridade, sendo esta variável operacionalizada como sendo a média de anos de estudo no período estabelecido em função da série e do grau mais elevado alcançado pela pessoa, considerando a última série concluída com aprovação. Os indicadores Água Potável, AIE e PMSAB são obtidos por meio de investigação de órgãos públicos, entidades privadas, prefeituras e todo e qualquer órgão prestador de serviços de saneamento básico por todo o país e os dados são levantados por meio de questionários, verificando a abrangência da prestação de serviço na data da pesquisa em cada município. Por fim tem-se os indicadores TMI e EVN. O primeiro indica a frequência com que ocorrem os óbitos infantis entre 0 (zero) e 1 (um) ano de vida em uma população em relação ao número de nascidos vivos no mesmo período. Esse valor é dado para cada mil crianças nascidas vivas. Já o segundo indicador, EVN, representa a expectativa de vida ao nascer, ou esperança de vida ao nascer, sendo este o número aproximado de anos que um grupo de indivíduos nascidos no mesmo ano irá viver se mantidas as mesmas condições desde o seu nascimento. Esse valor é calculado com base em uma média aritmética das idades das pessoas que morreram naquela região.

### **As etapas executadas para a análise dos dados são apresentadas a seguir:**

- a) Teste de Normalidade: Para verificação da existência ou não de correlações entre as variáveis investigadas. Esse procedimento é essencial para que a matriz de correlação, base da análise fatorial, possa ser gerada com o coeficiente de correlação adequado. No caso de dados com um comportamento normal, utiliza-se o coeficiente de correlação de Pearson. Já para distribuições amostrais que não atendam a essas características utiliza-se o teste de correlação de Spearman. Apesar de similar ao coeficiente de Pearson, o método de Spearman se difere por transformar em postos os escores originais (ASTIVIA; ZUMBO, 2017).

- b) Avaliar se a distribuição Normal é adequada para modelar cada variável: utilizou-se o teste Shapiro-Wilk, cuja hipótese nula é a de normalidade dos dados (SEN, 2002). Após a realização deste teste e constatada a existência de variáveis que não podem ser modeladas por uma distribuição Normal, procedeu-se à análise de correlação utilizando o método de Spearman.
- c) Análise fatorial: Os seguintes testes foram realizados para confirmação da adequação de uma análise fatorial: 1) estatística Kaiser-Meyer-Olkin (KMO), uma medida para verificar se as variáveis estão fortemente correlacionadas entre si. Nesta estatística, foi utilizado um limiar de 0,50, sugerindo que KMO acima de 0,50 tem-se que as variáveis estão significativamente correlacionadas (KUMAR, 2014) 2) teste de esfericidade de Bartlett que permite, em um único teste, verificar, conjuntamente, o nível de correlação entre as variáveis. Para isto, a hipótese nula do teste é a de que a matriz de correlação é igual a matriz de identidade. Caso essa hipótese não seja rejeitada, tem-se que não há correlação (HORZUM; DEMIRHAN, 2017). Espera-se que neste teste, esta hipótese seja rejeitada para que se possa prosseguir com a análise fatorial.
- d) Extração dos fatores: após a confirmação da adequação em se realizar a análise fatorial, a extração dos fatores foi baseada na técnica de componentes principais (PCA, do inglês *Principal Component Analysis*). Esta técnica tem como objetivo a síntese, em alguns fatores, da maior parte da informação original. A decisão referente ao número de fatores a serem extraídos foi baseada no critério de Kaiser, que considera que a quantidades de fatores na análise corresponde aos autovalores maiores do que 1 na PCA (QURESHI et al., 2017). Posteriormente os fatores foram rotacionados ortogonalmente utilizando o critério VARIMAX, gerando valores que são compreendidos mais facilmente. Quanto maior o valor absoluto da carga fatorial, maior a importância da variável dentro do fator. Estas cargas foram classificadas como forte ( $> 0,75$ ), moderada (0,50 a 0,75), e fraca (0,30 a 0,50) (WANG et al., 2017). Por último foram geradas as comunalidades que visam identificar a porcentagem explicada pelos fatores para cada variável.
- e) Agrupamento: O agrupamento (também conhecido na literatura como clusterização) foi utilizado para dividir as UF em diferentes grupos. Por meio desta análise, as UF pertencentes ao mesmo grupo indicam que possuem características semelhantes. Para criação dos grupos foi utilizado o método não hierárquico, empregando-se as distâncias euclidianas como medida de similaridade.

Todas as análises estatísticas foram realizadas no programa R (R CORE TEAM, 2018). Para a construção do correlograma foi utilizada a biblioteca corrgram (WRIGHT, 2017). Para a análise multivariada utilizou-se a biblioteca psych (REVELLE, 2018).

### 3. RESULTADOS E DISCUSSÕES

#### 3.1 Análise de Normalidade das Variáveis

Na Tabela 1 são apresentadas medidas descritivas e os valores p do teste de normalidade Shapiro-Wilk. Estes resultados dão evidências de que as variáveis AIE, Escolaridade, Desemprego, PMASB, TMI e EVN se mostraram não significativas ao nível de confiança de 5%, indicando a não rejeição da hipótese de normalidade. Já em relação as variáveis Analfabetos, Renda e Água, a hipótese nula de normalidade deve ser rejeitada ao nível de 5% de significância.

**Tabela 1:** Valor do teste de normalidade de Shapiro-Wilk e outras estatísticas descritas para cada variável investigada

Variáveis	Valor p	Mínimo	Máximo	Média	Desvio Padrão	Coefficiente de Variação (%)
Analfabetos	0,024 **	2,73	21,97	10,25	5,93	57,89
Renda	0,002 *	593,00	2.279,70	1033,00	381,98	36,98
Água	0,001 *	73,00	100,00	92,22	7,23	7,84
AIE	0,104	37,00	97,00	65,26	19,07	29,22
Escolaridade	0,775	5,65	10,08	7,43	1,07	14,43
Desemprego	0,812	3,13	13,13	7,85	2,49	31,71
PMASB	0,271	2,41	95,47	48,42	28,13	58,09
TMI	0,132	10,12	24,75	16,76	4,47	26,68

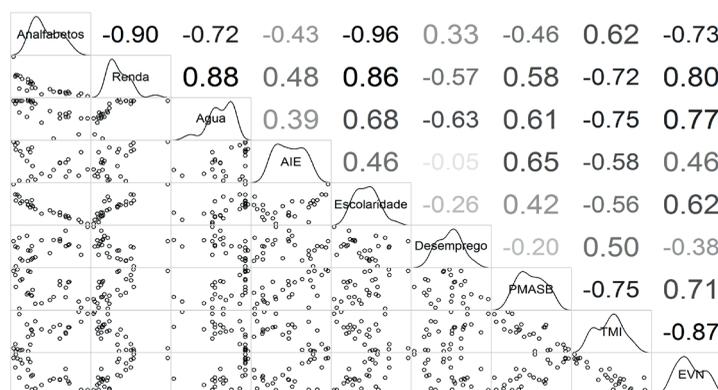
\* Significativo a 1%; \*\* Significativo a 5%

Em relação às medidas de distribuição das variáveis analisadas, destacam-se Analfabetos, Renda, Desemprego e PMASB. Elas mantiveram os maiores níveis de variação em relação a suas médias. Essa discrepância dos dados pode ser confirmada ao se analisar o coeficiente de variação dessas variáveis que mantiveram valores acima de 30%.

A partir dos resultados do teste de normalidade, para algumas variáveis, foi observado o valor p inferior de nível de significância de 5%. Este valor sugere que o método de Spearman é o mais indicado para calcular as correlações entre as variáveis e para a realização da PCA.

A Figura 1 aponta os resultados dos coeficientes de correlação de postos de Spearman entre as variáveis. Por meio desta figura é possível perceber que as variáveis com as maiores relações diretamente proporcionais são: Água e Renda (0,88), Renda e Escolaridade (0,86), EVN e Renda (0,80) e EVN e Água (0,77). Já as variáveis com maiores relações inversamente proporcionais são: Escolaridade e Analfabetos (-0,96), Analfabetos e Renda (-0,90), TMI e Água (-0,75), EVN e Analfabetos (-0,73), Analfabetos e Água (-0,72) e TMI e Renda (-0,72). De forma geral, identificam-se altas correlações entre as variáveis analisadas, o que justifica a relevância do emprego de uma análise multivariada.

**Figura 1:** Coeficientes de correlação de Spearman (diagonal superior) e gráficos de dispersão (diagonal inferior) entre as variáveis no estudo, seguida do gráfico de densidade (diagonal principal) para cada uma destas variáveis.



### 3.2 Análise Multivariada

Os resultados para os testes KMO e Esfericidade de Bartlett são apresentados na Tabela 2. Para aplicação da análise fatorial na matriz de correlação, foi averiguada inicialmente a adequabilidade dos dados pelo teste individual MSA e geral KMO. Identificou-se que apenas a variável Desemprego apresentou um valor de 0,40 e, de acordo com a literatura, evidencia uma baixa correlação dessa variável com as demais, podendo ser excluída da Análise Fatorial. Contudo, como o número de variáveis analisadas é de apenas nove e como o valor alcançado pela variável Desemprego não foi muito inferior ao 0,50 antes estipulado, optou-se pela permanência de todas elas para análise.

**Tabela 2:** Adequação dos resultados para o modelo de Análise Fatorial

<b>Kaiser-Meyer-Olkin (KMO)</b>		<b>0,700</b>
Teste de Esfericidade de Bartlett	Qui-quadrado aproximado	270,919
	Valor p	$8 \times 10^{-30}$
	Graus de liberdade	36

A possibilidade de se realizar a análise fatorial também pode ser confirmada por meio do teste de Esfericidade de Bartlett, o qual demonstrou que a hipótese nula deve ser rejeitada, ao nível de 5% de significância, uma vez que o valor p inferior a 1%, evidenciando assim a existência de correlação entre as variáveis estudadas. Em relação ao valor de KMO do conjunto das variáveis verificou-se que este manteve um valor de 0,70, evidenciando a adequação do uso da Análise Fatorial para esse estudo.

A Tabela 3 sumariza os valores de cada componente na PCA seguida do percentual da variância explicada por cada componente e o valor acumulado. Os dados ainda revelam que as duas primeiras componentes apresentaram autovalores maiores do que 1, sugerindo a redução dos fatores para apenas dois. Estas duas componentes, provenientes da PCA, são capazes de explicar 78,51% da variabilidade dos dados.

**Tabela 3:** Autovalores e o percentual da variância explicada por cada componente na PCA.

Componente	Autorvalor	% Variância	% Acumulado
1	2,43	65,53	65,53
2	1,08	12,98	78,51
3	0,97	10,44	88,96
4	0,67	5,03	93,99
5	0,49	2,71	96,70
6	0,38	1,65	98,35
7	0,32	1,17	99,51
8	0,17	0,32	99,84
9	0,12	0,16	100,00

Após a extração dos fatores é realizada a matriz de cargas fatoriais a fim de obter uma pré-identificação de possíveis grupos. Essas informações estão organizadas na Tabela 4. Por meio desses dados é possível perceber a criação de dois grupos distintos. Eles estão relacionados ao desempenho de cada variável. O primeiro grupo é composto pelas variáveis que possuem cargas negativas. Ele indica, nesse caso, que quanto menor o valor obtido pela UF nessas variáveis, melhor será seu desempenho. O segundo grupo é formado pelas variáveis com cargas positivas. Ele indica que quanto maior o valor obtido pela unidade federativa para essas variáveis, melhor será seu desempenho.

**Tabela 4:** Matriz de cargas fatoriais de cada variável e estimativas de comunalidade.

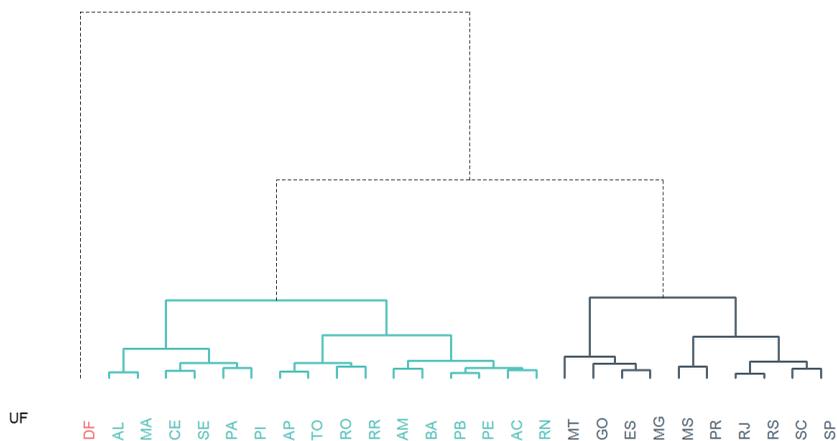
Variáveis	Fator 1	Fator 2	Comunalidade
Analfabetos	-0,350	-0,867	0,875
Renda	0,532	0,742	0,834
Água	0,552	0,503	0,558
AIE	0,583	0,342	0,456
Escolaridade	0,264	0,962	0,995
Desemprego	-0,431	-0,141	0,205
PMASB	0,759	0,252	0,639
TMI	-0,910	-0,364	0,961
EVN	0,811	0,434	0,846

Uma das explicações para o baixo poder explicativo desses fatores está relacionada ao fato de que essas variáveis não demonstraram correlações significativas com as demais. Isso sugere que certos Estados brasileiros podem exibir indicadores de desenvolvimento satisfatórios, porém com níveis relativamente altos de pobreza, como observado nas variáveis Desemprego e AIE. Nesse contexto, nossa pesquisa corrobora os achados de Silva, Sousa e Araujo (2017), que evidenciaram a coexistência de desenvolvimento e pobreza multidimensional na região Norte do Brasil durante o período de 2006 a 2013.

### 3.3 Análise de Agrupamento

Para realizar a análise de clusters foram empregados os valores encontrados anteriormente na análise fatorial. Desta forma, as observações são agrupadas conforme a proximidade entre elas. Conforme ilustrado na Figura 2, optou-se em fazer um corte no dendograma para que se possa visualizar os três grupos existentes: o primeiro grupo (à esquerda do dendograma) composto unicamente pelo Distrito Federal, o segundo grupo (no centro) composto pelos Estados das regiões Norte e Nordeste, e o terceiro grupo (à direita) composto pelos Estados das regiões Centro-Oeste, Sul e Sudeste do país.

**Figura 2:** Dendograma da análise de Cluster com a divisão em três grupos principais, sendo o primeiro com a legenda DF, o segundo de AL à RN e o terceiro de MT à SP.



O primeiro grupo é formado somente pelo Distrito Federal, representado pela cor vermelha na figura, o qual apresentou o maior índice nas variáveis Renda, Escolaridade e Proporção de domicílios com acesso a Saneamento Básico em relação aos demais estados.

O segundo grupo é constituído por Estados com menor Renda, na cor azul, sejam eles Alagoas (AL), Maranhão (MA), Ceará (CE), Sergipe (SE), Pará (PA), Piauí (PI), Amapá (AP), Tocantins (TO), Rondônia (RO), Roraima (RR), Amazonas (AM), Bahia (BA), Paraíba (PB), Pernambuco (PE), Acre (AC) e Rio Grande do Norte (RN). As variáveis Analfabetos, Taxa de Desemprego e a TMI justificam a relação deste grupo, corroborada também pela pesquisa de Silva, Sousa e Araujo (2017). Esses autores, em suas pesquisas, apresentam evidências sobre a situação na região Norte e Nordeste. Os resultados são similares aos apresentados nesta presente pesquisa, o que é notório como essas variáveis fazem jus com a realidade do Brasil. Os autores supracitados afirmaram ainda que ‘a pobreza é mais elevada no Norte e Nordeste’ e os avanços existentes na região não são homogêneos, o que reforça ainda mais as disparidades existentes tanto a curto quanto a longo prazo.

Por fim, o terceiro grupo formado por regiões um pouco mais desenvolvidas. Os estados que compõem este grupo são aqueles do Sudeste e Centro-Oeste: Mato Grosso (MT), Goiás (GO), Espírito Santo (ES), Minas Gerais (MG), Mato Grosso do Sul (MS), Paraná (PR), Rio de Janeiro (RJ), Rio Grande do Sul (RS), Santa Catarina (SC) e São Paulo (SP). As variáveis Renda, Água Potável, AIE e Escolaridade distinguiram este grupo. Além disso, Tavares e Porto Junior (2011) relataram anteriormente que esta região possui bons índices de desenvolvimento, embora exista limitações para a boa qualidade de vida de todos os habitantes.

## CONSIDERAÇÕES FINAIS

O Brasil é dividido em cinco regiões, sendo que os principais fatores utilizados para esta divisão são as características naturais, sociais, econômicas e as político-administrativas. Entre estas características, as variáveis percentual de analfabetos, renda média mensal, acesso a água potável, AIE, escolaridade média, taxa de desemprego, PMASB, TMI e EVN retirados do ano de 2014 do IPEA, IBGE e Pnad foram utilizadas para este estudo. Uma análise multivariada por meio de PCA foi conduzida e os resultados apontaram dois novos fatores que representam 78,51% da variabilidade dos dados. Por meio de uma análise de cluster, três grupos ficaram bem explícitos no dendograma, um com apenas o DF, outro grupo com Estados da região Norte e Nordeste e o último grupo com Estados da região Centro-Oeste, Sul e Sudeste.

Os resultados desta pesquisa destacam o alto nível de desigualdade existente no Brasil. Isso pode ser confirmado por meio do agrupamento realizado na análise de cluster. Bem como apontado em pesquisas similares, fica evidente que há uma maior distribuição de recursos a um grupo seleto de Estados, com destaque para o Distrito Federal, o qual alcançou níveis de desenvolvimento muito superiores aos dos demais Estados. Nota-se ainda que os Estados do Sul, Sudeste e Centro-Oeste mantiveram bons indicadores, apresentando um desempenho razoável e similar dentre os grupos formados. A região Sudeste é pioneira em industrialização, o que contribui para o maior acúmulo de capital, importação e exportação, sendo a região mais populosa e com maior densidade demográfica, na qual há maior concentração de renda média do país. No tocante à região Sul, ela é menos populosa, mas possui bons indicadores sociais como menor índice de TMI, menor percentual de analfabetos e elevado IDH. A região Centro-Oeste recebe bons investimentos em infraestrutura e no setor agrícola. Já os Estados que compõem as regiões Norte e Nordeste obtiveram os piores índices de desenvolvimento, indicados especialmente pelas suas taxas de analfabetismo, sendo estas muito superiores às demais regiões. Estes dados apontam para uma carência desses Estados quanto a investimentos na área da educação. Além disso, sua alta taxa de mortalidade infantil revela a necessidade de melhorias na área da saúde. A região Norte se destaca pela sua abrangência territorial, possuindo uma grande área florestal e pouca área urbana, o que justifica seu baixo índice populacional. Como relatado em outras pesquisas, os principais serviços, como o de saúde, acabam sendo indisponibilizados por falta de área de atuação e o deslocamento não usual faz com que

a população viva em condições precárias. Ao contrário da região Norte, a região Nordeste possui grande população, porém mesmo que seu Produto Interno Bruto tenha aumentado nos últimos anos, isso não tem impactado positivamente na qualidade de vida da população, como em infraestrutura e saneamento básico. Além disso, algumas cidades desta região constantemente sofrem com as secas devido às características climáticas. Adicionalmente, as desigualdades regionais e intrarregionais acompanhada do baixo nível de renda per capital encontradas deem destaque a problemas sociais ligados aos altos índices de extrema pobreza.

A matriz de cargas fatoriais foi construída nesta pesquisa. Por meio dela foi possível verificar a criação de dois grupos distintos entre as variáveis, um deles composto por cargas negativas indicando que quanto menor o valor dessa variável melhor o desempenho do estado analisado por ela. O outro grupo foi composto por cargas positivas que indica que quanto maior o valor desse indicador, melhor é o desempenho desse Estado em sua análise individual. Esse resultado condiz com a realidade, pois ao verificar como exemplo o grupo de carga negativa composto pelas variáveis Analfabetos, Desemprego e TMI é possível concluir que quanto menor os valores encontrados para essas variáveis, menores serão as divergências sociais.

Dessa forma, conclui-se que esta pesquisa identificou as relações entre as variáveis socioeconômicas do Brasil e que representa dados reais. Os resultados também relevaram a prevalência das desigualdades entre as unidades federativas, corroboradas por outras pesquisas. Vale salientar que os dados utilizados neste estudo refletem dados reais e que por este motivo eles podem sofrer alterações. Este trabalho pode ser utilizado como instrumento para auxiliar a administração pública no desenvolvimento de políticas focadas no combate à pobreza e aceleração e melhoria do desenvolvimento do país.

## REFERÊNCIAS

ARAÚJO, V.; FLORES, P. Redistribuição de renda, pobreza e desigualdade territorial no Brasil. **Revista de Sociologia e Política**, v. 25, n. 63, p. 159–182, set. 2017.

ASTIVIA, O. L. O.; ZUMBO, B. D. Population models and simulation methods: The case of the Spearman rank correlation. **British Journal of Mathematical and Statistical Psychology**, v. 70, n. 3, p. 347–367, nov. 2017.

BARBOSA, M. S. M. et al. A RELAÇÃO ENTRE POBREZA E DEGRADAÇÃO AMBIENTAL NO BRASIL SOB A ÓTICA DOS INDICADORES PIB/ IDH. **Semioses**, v. 9, n. 1, 24 mar. 2015.

HORZUM, M. B.; DEMIRHAN, E. The role of chronotype on Facebook usage aims and attitudes towards Facebook and its features. **Computers in Human Behavior**, v. 73, p. 125–131, ago. 2017.

KUMAR, P. J. S. Evolution of groundwater chemistry in and around Vaniyambadi Industrial Area: Differentiating the natural and anthropogenic sources of contamination. **Chemie der Erde - Geochemistry**, v. 74, n. 4, p. 641–651, dez. 2014.

MAURO, R. DE A. et al. Dimensões do Desenvolvimento Territorial e Políticas Públicas: Perspectivas e Desafios a Partir da Constituição Federal de 1988. **Rev. Gest. Ambient. Sustentabilidade**, v. 7, p. 489–506, 2018.

PIACENTI, C. A.; PIACENTI, S. C. V. Economic Development Indicator of Paraná's Cities in an Expanded Way. **Revista Paranaense de Desenvolvimento - RPD**, v. 39, p. 195–216, 2018.

QURESHI, N. A. et al. Application of Principal Component Analysis (PCA) to Medical Data. **Indian Journal of Science and Technology**, v. 10, n. 20, p. 1–9, 1 fev. 2017.

- R CORE TEAM. **R: A language and environment for statistical computing****R: A Language and Environment for Statistical Computing**, 2018. Disponível em: <<https://www.r-project.org/>>
- REVELLE, M. W. psych: Procedures for Psychological, Psychometric, and Personality Research. **R Cran Project**, 2018.
- SEN, P. K. Shapiro-Wilk Type Goodness-of-Fit Tests for Normality: Asymptotics Revisited. In: HUBER-CAROL, C. et al. (Eds.). . **Goodness-of-Fit Tests and Model Validity**. Boston: Birkhäuser Boston, 2002. p. 73–88.
- SILVA, A. F. DA; SOUSA, J. S. DE; ARAUJO, J. A. Evidências sobre a pobreza multidimensional na região Norte do Brasil. **Revista de Administração Pública**, v. 51, n. 2, p. 219–239, mar. 2017.
- TAVARES, J. M.; JUNIOR, S. DA S. P. Análise das desigualdades inter e intraestaduais na região Sul do Brasil por meio da análise de componentes principais. **Perspectiva Econômica**, v. 7, p. 1–14, 2011.
- VIDIGAL, V. G.; AMARAL, I. DE C.; GLAUBER FLAVIANO SILVEIRA. Desenvolvimento Socioeconômico nas microrregiões do Paraná: uma análise multivariada. **Revista de Economia**, v. 38, p. 51–72, 2012.
- WANG, J. et al. Multivariate statistical evaluation of dissolved trace elements and a water quality assessment in the middle reaches of Huaihe River, Anhui, China. **Science of The Total Environment**, v. 583, p. 421–431, abr. 2017.
- WRIGHT, K. corrgram: Plot a Correlogram. **R Cran Project**, 2017.

---

1. Doutorado em Estatística e Experimentação Agropecuária, Universidade Federal da Grande Dourados. ORCID: <https://orcid.org/0000-0002-9694-4019>, E-mail: [eliasmedeiros@ufgd.edu.br](mailto:eliasmedeiros@ufgd.edu.br)

2. Psicóloga, mestre em Psicologia e Tecnóloga em Gestão Financeira, Instituto Federal de Minas Gerais. ORCID: <https://orcid.org/0000-0002-1303-2867>, E-mail: [jaquelinecondemelo@gmail.com](mailto:jaquelinecondemelo@gmail.com)

3. Doutorado em Engenharia Elétrica, Universidade Federal de Minas Gerais. ORCID: <https://orcid.org/0000-0001-6934-6745>, E-mail: [m.voicer@gmail.com](mailto:m.voicer@gmail.com)

4. Doutorado em Estatística e Experimentação Agropecuária, Universidade Estadual de Mato Grosso do Sul. ORCID: <https://orcid.org/0000-0002-4832-1909>, E-mail: [carolinabicalho@gmail.com](mailto:carolinabicalho@gmail.com)

---

---

Recebido em: 18 de Julho de 2022

Avaliado em: 21 de Março de 2024

Aceito em: 15 de Agosto de 2024

---



[www.periodicos.uniftc.edu.br](http://www.periodicos.uniftc.edu.br)

---



Periódico licenciado com Creative Commons  
Atribuição-NãoComercial 4.0 Internacional.