

Análise de agrupamento dos gastos públicos em municípios goianos e sua qualidade de vida

Renan Oliveira Firmino ^{1 2}, Everton Batista da Rocha ³

1 Introdução

Não há dúvidas de que alta renda familiar em um município é fator determinante na qualidade de vida dos seus habitantes, mas de acordo com o Programa das Nações Unidas para Desenvolvimento (PNUD, 2018), esta não está atrelada tão somente à questão econômica, mas também ao desenvolvimento humano, em que a renda é apenas um dos fatores que colaboram para o bem-estar social de um indivíduo.

No contexto brasileiro, a qualidade de vida passa diretamente pela ação dos municípios, o que é reforçado pelo Instituto FIRJAN que calcula a medida do IFDM (Índice FIRJAN de Desenvolvimento Municipal), a qual utilizando as mesmas medidas do IDH (Índice de Desenvolvimento Humano) global, quantifica entre zero e um o desenvolvimento humano nos municípios brasileiros, que possuem as seguintes categorias: baixo (de 0 a 0,4), regular (0,4 a 0,6), moderado (de 0,6 a 0,8) e alto (0,8 a 1).

O objetivo deste trabalho é analisar descritivamente, por meio de análise de agrupamentos, como se relaciona o desenvolvimento humano dos municípios do estado de Goiás no Brasil, e seus respectivos gastos públicos em variadas áreas que afetam diretamente a vida de seus cidadãos, buscando semelhanças e diferenças entre os gastos *per capita* feitos por diferentes grupos de municípios e como estes se revestem diretamente na qualidade de vida local.

2 Materiais e métodos

Nesta análise, foram utilizados de cada município a população em 2010, o IFDM (Índice FIRJAN de Desenvolvimento Municipal) em 2010 e gastos públicos municipais em 2011 em seis diferentes áreas: saúde e saneamento, educação e cultura, administração, habitação e urbanismo, assistência social e gastos municipais. A datação dos dados é consequência de sua disponibilidade no sítio eletrônico do Instituto de Pesquisa Econômica Aplicada (IPEA, 2018), de onde foram obtidos.

No tratamento inicial dos dados, feito por meio programa estatístico *R*, foram removidos da análise todos os municípios com alguma ausência de dados, reduzindo o número de elementos de 246 para 189, e após isso, cada município teve seus gastos públicos divididos pela sua população para obtenção de um gasto *per capita*, uma vez que municípios maiores têm naturalmente maiores gastos em absoluto.

Uma vez que é inviável a comparação de todos os municípios individualmente, neste trabalho foi utilizada a análise de *cluster* (ou análise de agrupamento), uma técnica da estatística multivariada em que, segundo Rencher(2002), agrupam-se os vetores de observações em grupos, de tal forma que os grupos são internamente similares e externamente

¹Universidade Federal de Goiás. e-mail: renanoliveira628@gmail.com

²Bolsista PICME/CNPq

³Universidade Federal de Goiás. e-mail: evertonbatista@ufg.br

dissimilares, sendo o processo de comparação realizado por meio de uma medida de distância. Entre os variados métodos de agrupamento existentes, foi utilizado o algoritmo de Ward, um método hierárquico aglomerativo, isto é, inicia-se o algoritmo com n grupos (os vetores amostrais), e, a cada passo, dois deles se unem de forma que ao final tem-se um único *cluster* com os n objetos amostrais, cabendo ao pesquisador decidir o número ideal de partições para análise (RENCHE, 2002).

Em particular, o algoritmo de Ward é aquele que, dados dois agrupamentos A e B , minimiza a grandeza I_{AB} , definida por,

$$I_{AB} = SSE_{AB} - (SSE_A + SSE_B) \quad (1)$$

em que SSE representa a soma de erros quadrados,

$$\begin{aligned} SSE_A &= \sum_{i=1}^{n_a} (x_i - \bar{x}_A)^T (x_i - \bar{x}_A), \\ SSE_B &= \sum_{i=1}^{n_b} (x_i - \bar{x}_B)^T (x_i - \bar{x}_B), \\ SSE_{AB} &= \sum_{i=1}^n (x_i - \bar{x}_{AB})^T (x_i - \bar{x}_{AB}), \end{aligned}$$

sendo que n_a e n_b representam os respectivos números de elementos em A e B , \bar{x}_A representa o vetor de médias de A , \bar{x}_B representa o vetor de médias de B e $\bar{x}_{AB} = \frac{n_a \bar{x}_A + n_b \bar{x}_B}{n_a + n_b}$. Rencher, 2002 mostra que (1) pode ser reescrito como,

$$I_{AB} = \frac{n_a n_b}{n_a + n_b} (\bar{x}_A \bar{x}_B)^T (\bar{x}_A \bar{x}_B),$$

significando que minizar SSE é o mesmo que minizar a distância entre os grupos A e B . A vantagem da utilização do método de Ward, está no fato de ele levar em conta o número de elementos dentro de cada agrupamento em cada passo do algoritmo, o que resulta na união de agrupamentos pequenos e de tamanhos próximos, tornando-o também sensível a *outliers* (RENCHE, 2002).

Para a definição do número final de *clusters*, levou-se em conta, além da análise qualitativa dos resultados, três métodos distintos sugeridos por Mingoti(2005): distância de ligação, similaridade e estatística R^2 . No método da distância de ligação, observa-se o número de agrupamentos *vs* distância de união do algoritmo, sendo o número em que ocorre a última queda significativa de distância o número indicado de partições. A observação da similaridade, conforme Mingoti (2005), consiste no cálculo,

$$S_{AB} = \left(1 - \frac{d_{AB}}{\max\{d_{ij}, i, j = 1, 2, \dots, n\}} \right),$$

em que d_{AB} é a distância entre os dois conglomerados A e B unidos em dado passo do algoritmo, e $\max\{d_{ij}, i, j = 1, 2, \dots, n\}$ é a maior distância entre dois vetores amostrais, por consequência, quanto mais próximo de um for S_{AB} , maior a similaridade dentro dos *clusters*, sendo recomendado pelo menos 0,7 de similaridade para a partição final. Já a estatística R^2 é definida em Mingoti(2005) por,

$$R^2 = \frac{SSB}{SST_c},$$

sendo a Soma de Quadrados Total entre os g grupos (SSB), e a Soma de Quadrados Total corrigida para a média global (SST_c),

$$SSB = \sum_{i=1}^g n_i (\bar{\mathbf{X}}_i - \bar{\mathbf{X}})' (\bar{\mathbf{X}}_i - \bar{\mathbf{X}}),$$

$$SST_c = \sum_{i=1}^g \sum_{j=1}^{n_i} (\mathbf{X}_{ij} - \bar{\mathbf{X}})' (\mathbf{X}_{ij} - \bar{\mathbf{X}}),$$

em que $\bar{\mathbf{X}}_i$ denota o vetor de médias do i -ésimo agrupamento, \mathbf{X}_{ij} denota o j -ésimo elemento do i -ésimo cluster, n_i denota o número de membros do i -ésimo *cluster* e $\bar{\mathbf{X}}$ denota o vetor de médias global da amostra. Quanto mais próxima de um for a estatística R^2 , maior a similaridade interna dos grupos (MINGOTI, 2005).

3 Resultados e discussão

Executado o algoritmo de Ward, observa-se indicativo de seis partições finais pelo método número de grupos *vs* distância de ligação, corroborado por uma similaridade 0,825 e estatística R^2 de 0,863. Todavia, ao se levar em conta a sensibilidade do algoritmo a *outliers*, que produziria dois grupos com apenas um membro, foi adotado o número de cinco partições, que está próximo ao indicativo do gráfico, com similaridade de 0,730 e R^2 de 0,832. O dendrograma é apresentado na Figura 1.

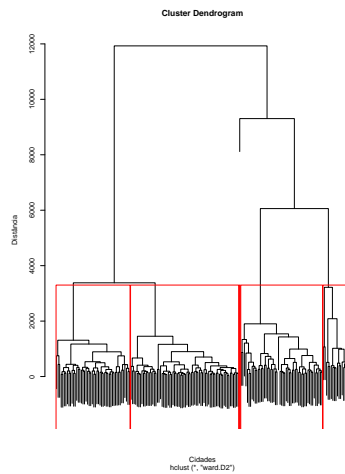


Figura 1: Dendrograma dos agrupamentos

Fonte: Elaborado pelos autores

Devido ao número de cidades presentes na análise, é difícil notar com exatidão o número de cidades em cada grupo, o que é expresso na Tabela 1.

Tabela 1: Número de municípios de cada agrupamento

Nº do <i>cluster</i>	Nº de municípios do <i>cluster</i>
1	48
2	70
3	17
4	1
5	53

Fonte: Elaborado pelos autores

Objetivando a comparação entre os 189 municípios, agora alocados em 5 agrupamentos representativos, foi extraída a média de cada agrupamento, tanto dos gastos, quanto dos IFDM's, além da média dos agrupamentos (Tabela 2).

Tabela 2: IFDM e gastos médios *per capita* de cada agrupamento

Nº do cluster	1	2	3	4	5	Média global
IFDM	0,69	0,67	0,68	0,72	0,66	0,68
Administração	277,06	208,36	812,12	3806,34	517,64	1124,30
Edu. e cultura	514,39	407,89	841,70	2854,99	658,40	1055,48
Hab. e urbanismo	233,74	171,31	438,17	542,16	261,07	329,29
Saúde e san.	509,44	650,61	869,32	1453,54	589,07	754,40
Municipal	1453,24	1061,16	3097,97	7534,96	2009,22	3031,31
Assis. social	182,17	139,98	366,68	1036,09	264,04	397,80

Fonte: Elaborado pelos autores

Graficamente, as informações da Tabela 2 podem ser resumidas na Figura 2.

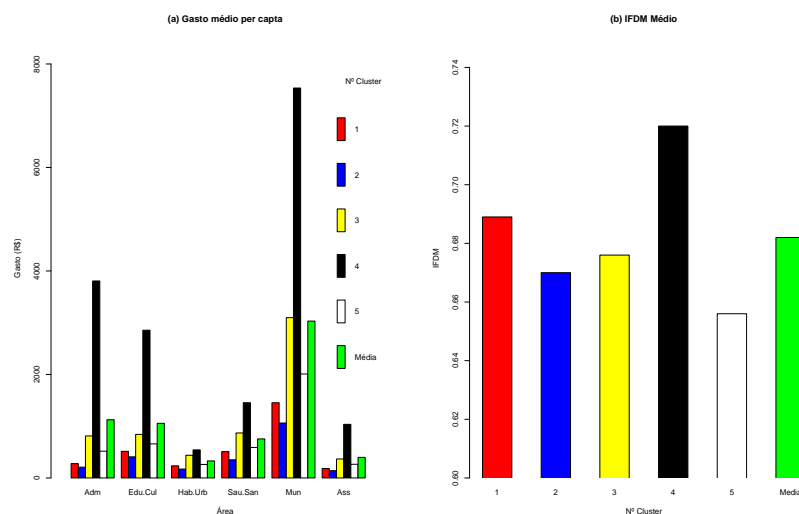


Figura 2: (a) Despesa média *per capita* de cada agrupamento. (b) IFDM médio de cada agrupamento

Fonte: Elaborada pelos autores

Nota-se, inicialmente, que todos os *clusters* têm IFDM entre 0,6 e 0,8, logo têm desenvolvimento humano médio moderado segundo FIRJAN(2018), mas notam-se diferenças (em absoluto) quanto aos gastos de cada agrupamento. O agrupamento 4, representado pelo município de Alto Horizonte, tem o maior IFDM médio, entretanto possui também gastos consideravelmente maiores que os demais agrupamentos, indicando que este município precisa gastar mais recursos para prover desenvolvimento humano.

O grupo 1 aparece com o segundo maior IFDM médio, e apenas a quarta média de gastos, sendo destaque positivo, uma vez que representa municípios que dispenderam menos recursos em relação aos outros, mas conseguiram IFDM ligeiramente mais alto. Vale notar que neste agrupamento constam cinco dos 15 maiores municípios do estado de Goiás, como Rio Verde e a capital Goiânia. O grupo 2 também aparece com importantes membros do estado, contém seis dos 15 maiores municípios do estado, como Anápolis e Aparecida de Goiânia, e representa os municípios com a menor média de gastos públicos *per capita*, mas que apresenta um IFDM pouco menor que a média, indicando ainda boa relação IFDM/gastos.

O grupo 3 que com 17 municípios contém 10 dos 15 menores municípios de Goiás, como Cachoeira de Goiás e Lagoa Santa, descreve a relação gastos/IFDM dos municípios

pequenos que apresentaram a segunda maior média de gastos e o terceiro IFDM, bem próximo à média, indicando que estes municípios de menor porte precisam dispendir mais recursos para atingir IFDM semelhante ao dos grandes municípios, cabendo aqui a análise do papel da iniciativa não-governamental, muito mais presentes em municípios de grande porte, e se também grandes municípios possuem melhor capacidade gestora. Por fim, o agrupamento 5 contém dois dos 15 menores municípios do estado, mas representa majoritariamente os municípios de médio porte, que de acordo com os números absolutos, possuem o terceiro maior gasto médio nas seis áreas analisadas e a pior média em IFDM, o que levanta a questão se a qualidade de vida é mais dificilmente alcançada no médio porte em comparação aos de maior e menor respectivamente.

4 Conclusão

Apesar da datação dos dados e de ser uma análise puramente descritiva, no contexto multivariado, este trabalho compara os diversos municípios goianos, apresentando indicativos a respeito do IFDM e dos gastos. A partir dos cinco grupos representativos construídos por meio da análise de *cluster* hierárquica aglomerativa, percebeu-se que os gastos públicos se distribuem de maneira similar nas seis áreas analisadas, isto é, a ordem dos agrupamentos no que se refere a gastos realizados, é exatamente a mesma nas seis áreas analisadas. Por meio da Análise de Agrupamento dos municípios também foi possível formular questões relevantes, como o papel da iniciativa privada e de ONG's nos municípios, e como o porte de um município se relaciona com a sua capacidade de transformar gastos públicos em qualidade de vida.

5 Agradecimentos

Agradecimento ao Programa de Iniciação Científica e Mestrado - PICME da Universidade Federal de Goiás - UFG, pelo conhecimento adquirido. E ao CNPq pelo apoio financeiro.

6 Referencias Bibliográficas

FEDERAÇÃO DAS INDÚSTRIAS DO ESTADO DO RIO DE JANEIRO. *Índice FIRJAN de Desenvolvimento Municipal (IFDM)*. 2018. Disponível em: <http://www.firjan.com.br/ifdm/>. Acesso em: 19 ago. 2018.

INSTITUTO DE PESQUISA ECONÔMICA APLICADA. *ipeadata*. 2018. Disponível em: <http://www.ipeadata.gov.br>. Acesso em: 19 ago. 2018.

MINGOTI, S. A. *Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada*. 1.ed. Belo Horizonte: Editora UFMG. 2005.

PROGRAMA DAS NAÇÕES UNIDAS PARA O DESENVOLVIMENTO. *O que é Desenvolvimento Humano*. 2018. Disponível em: <http://www.br.undp.org/content/brazil/pt/home/idh0/conceitos/o-que-e-desenvolvimento-humano.html>. Acesso em: 19 ago. 2018.

R CORE TEAM. *R*: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2018. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.

RENCHEER, A.C. *Methods of Multivariate Analysis*. 2.ed. Canada: Wiley-Interscience. 2002.