

UNIVERSIDADE FEDERAL DE ALFENAS

GABRIEL CALIJURI

**AGRUPAMENTO DOS MUNICÍPIOS BRASILEIROS SEGUNDO SUAS CARACTERÍSTICAS
DEMOGRÁFICAS**

VARGINHA

2022

UNIVERSIDADE FEDERAL DE ALFENAS

GABRIEL CALIJURI

**AGRUPAMENTO DOS MUNICÍPIOS BRASILEIROS SEGUNDO SUAS CARACTERÍSTICAS
DEMOGRÁFICAS**

Trabalho de conclusão de Piepex
apresentado ao Instituto de Ciências
Sociais Aplicadas da Universidade Federal
de Alfenas como requisito parcial para a
obtenção do título de Bacharel em Ciência
Economia.
Orientadora: Patrícia de Siqueira Ramos.

VARGINHA

2022

UNIVERSIDADE FEDERAL DE ALFENAS

GABRIEL CALIJURI

**AGRUPAMENTO DOS MUNICÍPIOS BRASILEIROS SEGUNDO SUAS CARACTERÍSTICAS
DEMOGRÁFICAS**

A banca examinadora, abaixo assinada, aprova o Trabalho de conclusão de Piepex apresentado ao Instituto de Ciências Sociais Aplicadas da Universidade Federal de Alfenas como requisito parcial para a obtenção do título de Bacharel em Ciência Economia.

Aprovada em: 7 de abril de 2022

Prof. Patrícia Siqueira Ramos

Universidade Federal de Alfenas – campus Varginha Assinatura: _____

Prof. Luísa Pimenta Terra

Universidade Federal de Alfenas – campus Varginha Assinatura: _____

Walef Machado de Mendonça

Assinatura: _____

Resumo

A transição demográfica ocorre de forma cada vez mais rápida no Brasil, um dos motivos é, que com a tecnologia, o mundo fica mais globalizado, o que faz com que hábitos e tecnologia de países mais desenvolvidos cheguem rapidamente a países menos desenvolvidos. Assim, a população brasileira está passando por um processo de queda na mortalidade e fecundidade, o resultado é uma população cada vez mais envelhecida. Isso impacta vários setores, como o regime geral de previdência brasileiro, a quantidade de mão de obra ofertada e gastos com saúde pública. Esse processo ocorre de forma mais/menos adiantada em cada município, dessa forma necessitam de políticas públicas distintas. Este trabalho visa agrupar os municípios brasileiros de acordo com seu perfil demográfico, a fim de gerar conhecimento sobre a transição demográfica em cada município brasileiro. Para esse fim, foi utilizada a análise de agrupamento, por meio da linguagem *python*. Foram identificados quatro grupos de municípios com localizações distintas no território brasileiro, sugerindo uma diferença nos perfis demográficos, principalmente entre as porções norte e sul do Brasil.

Palavras-chave: Transição Demográfica, Análise Multivariada, *k*-médias

ABSTRACT

The demographic transition is occurring more and more quickly in Brazil, one of the reasons is that with technology, the world becomes more globalized, which makes habits and technology from more developed countries quickly reach less developed countries. Thus, the Brazilian population is undergoing a process of decline in mortality and fertility, the result is an increasingly aging population. This impacts several sectors, such as the general Brazilian social security system, the amount of labor offered and public health expenditures. This process occurs more/less in advance in each municipality, thus requiring different public policies. This work aims to group Brazilian municipalities according to their demographic profile, in order to generate knowledge about the demographic transition in each Brazilian municipality. For this purpose, cluster analysis was used, using the python language. Four groups of municipalities with different locations in the Brazilian territory were identified, suggesting a difference in demographic profiles, mainly between the northern and southern portions of Brazil.

Keywords: Demographic Transition, Multivariate Analysis, k-means

SUMÁRIO

1. INTRODUÇÃO	5
2. REFERENCIAL TEÓRICO	5
2.1 TRANSIÇÃO DEMOGRÁFICA	5
2.2 ANÁLISE MULTIVARIADA	7
3. METODOLOGIA	8
4. RESULTADOS E DISCUSSÃO	12
5. CONSIDERAÇÕES FINAIS	16
6. REFERÊNCIAS BIBLIOGRÁFICAS	16

1. INTRODUÇÃO

As famílias vêm tendo cada vez menos filhos, e isso se deve a fatores, como métodos anticonceptivos mais eficientes e a inserção das mulheres no mercado de trabalho. Soma-se a esse cenário um avanço significativo na medicina e um aumento na adoção de hábitos saudáveis, o que proporciona uma maior longevidade levando de uma forma geral, a uma população mais envelhecida.(SOUZA; RAMOS; FRIAS, 2017)

Isso impacta muitos setores da sociedade, principalmente o Regime Geral de Previdência Social (RGPS), o mercado de trabalho, e gastos com saúde pública. Esse fenômeno ocorre no mundo inteiro e é um assunto estudado pela demografia, conhecido como teoria das transições demográficas, e ocorre de forma diferente mesmo dentro de um país.(SOUZA; RAMOS; FRIAS, 2017)

Tendo em vista isso, este trabalho pretende traçar os perfis demográficos dos municípios brasileiros, a fim de gerar conhecimento sobre a transição demográfica em cada município brasileiro. Como a identificação desses perfis requer a análise de algumas variáveis, a análise multivariada, por meio da análise de agrupamento, fornece os meios necessários para realizar essa identificação de grupos com comportamentos semelhantes.

2. REFERENCIAL TEÓRICO

2.1 TRANSIÇÃO DEMOGRÁFICA

Segundo Carvalho, Sawyer e Rodrigues (1998), o termo “Demografia (dêmos=população, graphein=estudo) refere-se ao estudo das populações humanas e sua evolução temporal no tocante a seu tamanho, sua distribuição espacial, sua composição e suas características gerais.”. Este estudo é feito através da estatística e eventos históricos que possam contribuir para a variação da estrutura etária da população, isto é, eventos que modifiquem a quantidade de pessoas em cada faixa etária. A transição demográfica é uma teoria que surgiu com o estudo da demografia, que indica a presença de uma tendência de queda na taxa de mortalidade e natalidade conforme a medicina evolui e a cultura se altera. Esse fenômeno ocorre em vários lugares em estágios diferentes. Até o momento foram observadas quatro fases. (CASTIGLIONI, A. H., 2020)

A primeira fase é caracterizada por uma grande quantidade de mortes e de nascimentos, o que faz o tamanho da população se manter estável. As pessoas geralmente morrem cedo, assim a população possui poucos idosos.

A segunda fase é caracterizada por um avanço na medicina e no saneamento básico, que faz com que a quantidade de mortes diminua consideravelmente, as pessoas passam a morrer mais velhas, e a população passa a ter mais idosos.

Na terceira fase a taxa de natalidade começa a diminuir devido a vários fatores, tais como métodos anticoncepcionais mais eficientes e maior participação feminina no mercado de trabalho, enquanto a taxa de mortalidade continua a cair. Com isso, o crescimento vegetativo da população diminui e a população passa a ter uma proporção cada vez maior de idosos.

Na quarta fase, as taxas de mortalidade e natalidade começam a se estabilizar novamente, com uma baixa quantidade de mortes e de nascimentos e a proporção de idosos na população é alta.

No Brasil, a segunda fase começou na segunda metade do século XX, e a terceira em meados de 1960 (SOUZA; RAMOS; FRIAS, 2017).

Os países que foram pioneiros nas transições demográficas tiveram mais tempo para se adequar à nova realidade demográfica, comparados aos países que passaram por esse processo posteriormente. A tecnologia de forma geral, faz com que hábitos, costumes e outras tecnologias de países em fases mais avançadas, chegue mais facilmente a países em fases anteriores. O que faz com que a transição demográfica ocorra de forma mais rápida nesses países do que nos países pioneiros. (REIS, C.; BARBOSA, L. M. L. H.; PIMENTEL, V. P., 2016, p.104)

Assim, é esperado um envelhecimento rápido da população do Brasil, e isso impacta seriamente a economia brasileira, em diversos fatores, como o mercado de trabalho, RGPS e gastos com saúde pública.(SOUZA; RAMOS; FRIAS, 2017,p.28)

Com menos pessoas trabalhando, o mercado brasileiro como um todo terá uma queda na produção. Para reduzir essa queda, é importante que a mão de obra seja qualificada, e que a maioria das pessoas em idade produtiva trabalhem. Assim, políticas públicas que visem integrar as mulheres ao mercado de trabalho e políticas públicas que melhorem a educação básica, e que incentivem o ingresso desses alunos no ensino superior são importantes, para o país estar preparado para os efeitos da transição demográfica. (PAIVA, P. T. A.; WAJNMAN, S, 2006)

O RGPS é embasado no modelo de repartição simples, o que significa que as aposentadorias pagas são provenientes das contribuições dos atuais trabalhadores. Como a tendência é aumentar a quantidade de idosos, e diminuir a de jovens, os impostos coletados para esse fim não serão suficientes para suportar uma aposentadoria de qualidade para todos os idosos.(GUMINIAK, S. C., 2017, p.6)

Os gastos com saúde pública tendem a aumentar significativamente, visto que idosos gastam muito mais com saúde do que jovens, e a tendência é que passe a haver mais idosos do que jovens. (SOUZA; RAMOS; FRIAS, 2017)

Dado todo esse contexto, cada município brasileiro vai carecer de políticas públicas condizentes à sua realidade demográfica. Municípios com mais idosos necessitam de mais investimentos em saúde pública, enquanto municípios com mais jovens necessitam de um investimento proporcional em escolas e creches. Para que o Brasil como um todo esteja preparado para essas mudanças, vindas com a transição demográfica. (PAIVA, P. T. A.; WAJNMAN, S, 2006)

2.2 ANÁLISE MULTIVARIADA

A estatística estuda modelos multivariados (que dependem de mais uma variável) muito antes de surgir a Análise Multivariada (AM), porém tal tarefa não era muito eficiente, visto que se torna muito difícil de ser feita sem o auxílio de uma máquina, pois envolvem dados com muitas observações. Com o avanço na tecnologia, e com sua acessibilidade, surgiram várias técnicas computacionais para esse fim.

A análise multivariada reúne essas técnicas computacionais que possibilitam obter informações de um conjunto com mais de uma variável. Que possibilitam filtrar e agrupar dados ou até mesmo prever o valor para uma variável que seja dependente dos dados. Alguns desses métodos são derivados da própria análise univariada já utilizadas pela estatística, como o coeficiente de correlação e a regressão linear (VICINI, L., 2005).

Dentro da AM existem vários procedimentos, que possibilitam agrupar os dados segundo suas características e constituem a análise de agrupamento. Existem métodos de agrupamento hierárquicos e não hierárquicos. A aglomeração hierárquica não depende de uma quantidade de grupo pré definida, ela gera uma sequência de grupos cada vez maior, sendo os novos grupos contidos nos grupos anteriores, tomando visualmente o formato de árvore. (VICINI, 2005).

De maneira geral, esses métodos de agrupamento utilizam alguma medida de distância de vetores para calcular a semelhança entre as observações. Segundo Vicini (2005), a medida mais utilizada é a euclidiana, que é calculada através da seguinte expressão:

$$D_{AB} = \sum_{j=1}^p \sqrt{(X_{A,j} - X_{B,j})^2} \quad (1)$$

em que p é a quantidade de variáveis a serem utilizadas, A e B são as observações a medir a distância (ou uma observação e algum outro ponto) e X é o valor da variável na amostra.

Segundo Vicini(2005), é necessária a padronização dos valores, quando estes não estão na mesma escala de medida. A padronização é feita subtraindo-se a média de cada variável para todas as observações e dividindo o resultado pelo desvio padrão da variável. O valor da variável X padronizada para Z é

$$z = \frac{X - \bar{x}}{S}, \quad (2)$$

em que \bar{x} é a média amostral da variável X e S é o desvio padrão amostral dessa variável.

O método de agrupamento não hierárquico das k -médias agrupa os dados através da distância euclidiana. Para a aplicação desse método, o número k de grupos é definido, são sorteadas as coordenadas dos centros dos grupos e, em seguida, é calculada a distância euclidiana entre as observações e os centros dos grupos. Cada observação irá pertencer ao grupo cujo centro esteja mais próximo da mesma. Em seguida, são calculadas novamente as posições dos centros e as novas posições serão as médias dos vetores de coordenadas das

observações pertencentes a cada grupo. Neste ponto, o processo é repetido, porém os centros que antes foram escolhidos aleatoriamente agora serão os centros obtidos anteriormente. A quantidade de vezes que esse processo é repetido varia caso a caso. (PALMA, L. F., 2018)

Existem diversas formas para interpretar a melhor quantidade de grupos para um agrupamento, é possível utilizar métodos hierárquicos para essa análise, é um dos mais conhecidos é o método de Ward. Segundo Mingoti (2005), nesse método, conhecido como de mínima variância, inicialmente, cada elemento é considerado um grupo único, ou seja, há n grupos. A cada passo do algoritmo, o procedimento cria um novo grupo de forma a minimizar a soma de quadrados entre os grupos. Ao final, tem-se um único grupo formado. A decisão sobre o número de grupos é uma análise visual do dendrograma formado, buscando-se identificar qual a maior distância para fazer o corte.

Na próxima seção são apresentados os dados utilizados, bem como os passos da análise empregada.

3. METODOLOGIA

A base de dados utilizada neste trabalho foi construída a partir de dados demográficos de todos os municípios brasileiros no ano de 2019. Tais dados foram retirados do site DataSUS através da plataforma TABNET (DATASUS, 2019).

As taxas calculadas são embasadas em estimativas para o ano de 2019, assim podem haver diferenças da realidade (já que não houve censo populacional em 2019). Além disso, algumas inconsistências podem ocorrer, tais como pessoas que podem não ter nascimento ou morte registrados em um cartório, o que também pode afetar a precisão dos indicadores calculados.

As variáveis utilizadas para o cálculo dos indicadores são: população residente por faixa etária, população de mulheres em idade fértil (15 a 49 anos), número de nascidos vivos por faixa etária da mãe, número de mortes por faixa etária. Esses dados formam a base para o cálculo dos indicadores que serão usados inicialmente para o agrupamento. Os indicadores a serem calculados são apresentados a seguir.

A Razão de dependência de idosos/jovens (RDI/RDJ), calcula a razão de pessoas que não são aptas a trabalhar e as que são aptas (jovens com menos de 15 anos e adultos a partir de 65 anos). Esse indicador pode ser calculado para jovens e idosos, de forma conjunta (razão de dependência total) ou separada. Para esse trabalho, como a amplitude das faixas etárias ser de 10 anos, serão considerados os idosos a partir de 60 anos. A fórmula é dada por:

$$RDI = \frac{N^{\circ} \text{ pessoas com } 60+ \text{ anos}}{N^{\circ} \text{ de pessoas com } 15 \text{ a } 59 \text{ anos}}, \quad RDJ = \frac{N^{\circ} \text{ pessoas com } 15- \text{ anos}}{N^{\circ} \text{ de pessoas com } 15 \text{ a } 59 \text{ anos}}, \quad RDT = RDI + RDJ \quad (3)$$

A Taxa de Fecundidade Total (TFT) é a expectativa da quantidade de filhos que uma mulher terá do começo ao fim de seu período fértil (15 a 49 anos). Ela é calculada

somando-se as taxas específicas de fecundidade, que são as proporções de filhos que cada faixa etária de mulheres teve. (MATUDA, NIVEA S., 2009) A TFT é calculada por:

$$TFT = \sum_{t=15}^{49} \frac{\text{n}^\circ \text{ de nascimentos com mães com idade } t}{\text{População feminina com idade } t} \quad (4)$$

A equação (4) leva em consideração que a amplitude das faixas etárias seja igual a 1, ou seja leva em consideração que as faixas etárias sejam contadas de 1 em 1 ano. Neste trabalho os conjuntos de dados obtidos para o cálculo desse indicador encontram-se em uma amplitude de 5 anos, e portanto correspondem a 15 a 19 anos, 20 a 24 anos, etc. Dessa forma deve-se alterar os limites do somatório e multiplicar pela amplitude das faixas etárias, como apresentado a seguir:

$$TFT = 5 \cdot \sum_{t=1}^{10} \frac{\text{n}^\circ \text{ de nascimentos com mães com idade } t}{\text{População feminina com idade } t} \quad (5)$$

A Taxa Específica de Mortalidade (TEM): proporção entre as pessoas que morreram de uma certa faixa etária e as que estavam expostas a morrer em um período da mesma faixa etária.

$$nTEM_x = \frac{nO_x}{nP_x} \quad (6)$$

Onde nO_x é a quantidade de óbitos de pessoas com x a $x+n$ anos de determinada região, e nP_x a população residente dessa mesma faixa etária e região.

Taxa Bruta de Mortalidade (TBM): similar à TEM, porém envolve todas as faixas etárias. A TBM em sua forma original é calculada dividindo-se a quantidade de mortos pela população total de cada região. Porém, dessa forma, ela é sensível à estrutura etária da população de cada região, de forma que, se uma população tem mais pessoas pertencentes a uma faixa etária com mais risco de morrer, a TBM como um todo será elevada, e quando comparada à mortalidade dessa região com outra região com estrutura etária diferente, o resultados obtidos não exprimirão a verdade. Para resolver esse empecilho foi utilizada a padronização, que considera os municípios como tendo a mesma proporção de pessoas em cada faixa etária, transformando a TBM em um indicador comparável (MATUDA, N. S., 2009). Matematicamente tem-se:

$$TBM_{pd} = \sum_{fe=0}^w TEM_{fe} \cdot \frac{\hat{P}_{fe}}{\hat{P}}, \quad (7)$$

em que $\frac{\hat{P}_{fe}}{\hat{P}}$ é a proporção de pessoas em faixa etária fe na população padrão, que é obtida através da seguinte expressão, e w é a quantidade de faixas etárias:

$$\frac{\hat{P}_{fe}}{\hat{P}} = \frac{\sum_{local=0}^{5569} \frac{p_{fe}^{local}}{p^{local}}}{5570}, \quad (8)$$

É importante ressaltar que a padronização nesse caso apenas minimiza o erro, pois a amplitude dos dados é de 5 anos para menores de 19 e de 10 anos para os demais até 79, e a diferença de 10 anos, dependendo da faixa etária, pode possuir uma expressiva diferença de probabilidade de morte.

Taxa de mortalidade infantil (TMI): proporção de recém nascidos vivos que morreram antes de completarem 1 ano entre todos os nascidos vivos. Geralmente, essa quantidade é multiplicada por 1.000 e é interpretada como a quantidade de nascidos que morrem a cada mil nascimentos. Matematicamente tem-se:

$$TMI = \frac{\text{Óbitos de 0 a 1 ano}}{\text{Nascidos vivos}} \times 1000, \quad (9)$$

Na Tabela 1 são apresentadas as siglas das variáveis utilizadas neste trabalho.

Tabela 1- Significados das sigla das variáveis usadas no trabalho

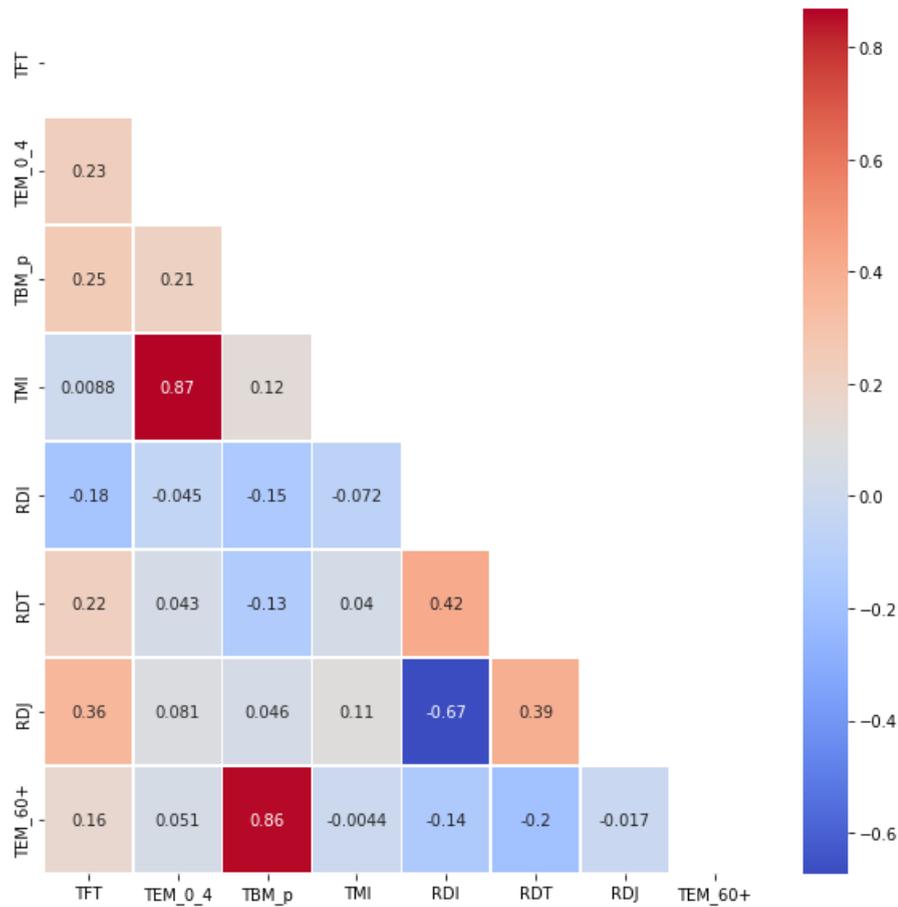
Sigla	Significado
TEM_0_4	Taxa Específica de Mortalidade de 0 a 4 anos
TEM_60+	Taxa Específica de Mortalidade de 60 anos ou mais
RDI	Razão de Dependência de Idosos
RDJ	Razão de Dependência de Jovens
RDT	Razão de Dependência Total
TMI	Taxa de Mortalidade Infantil
TBM_p	Taxa Bruta de Mortalidade padronizada
TFT	Taxa de Fecundidade Total

Fonte: elaboração própria

Após o cálculo dos indicadores foram calculados os coeficientes de correlação de Pearson entre as variáveis. A obtenção das correlações permite verificar se pares de variáveis apresentam comportamento muito similar, não sendo necessária a manutenção de ambas ao realizar a análise de agrupamento. A alta correlação pode ocorrer de forma positiva (valores próximos de +1) ou negativa (valores próximos de -1).

O gráfico da Figura 1 ilustra graficamente os valores dessas correlações entre os pares de variáveis. Quanto mais vermelho o valor, mais alta a correlação positiva entre os pares de variáveis e, quanto mais azul, mais alta a correlação negativa.

Figura 1 - Gráfico das correlações entre as variáveis



Fonte: Elaboração própria

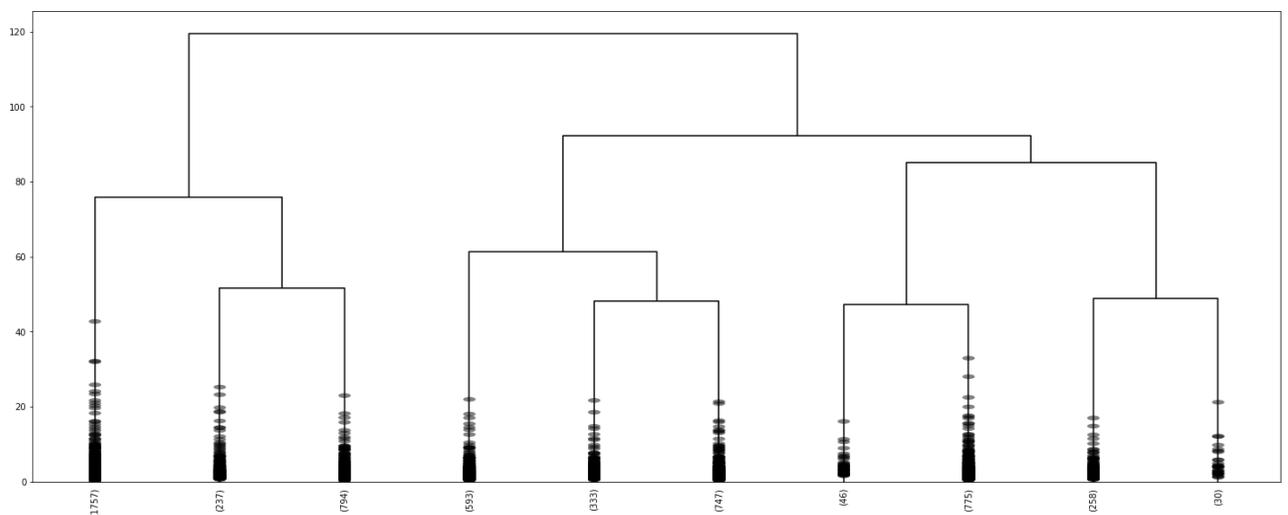
A partir da análise das correlações, foram considerados altos os valores acima de 0,7 que incluem as variáveis TEM_60+ e TBM_p (0,86) e TMI e TEM_0_4 (0,87). A TBM padronizada analisa a mortalidade como um todo, e descarta a necessidade de incluir a TEM de várias faixas etárias no modelo, e isso é benéfico, visto que muitas variáveis podem atrapalhar a classificação e a interpretação dos resultados. Dessa forma, embora a mortalidade de idosos seja pertinente ao trabalho, será utilizada apenas a TBM_p que possui uma correlação próxima de 0,9 com a de TEM de 60 anos ou mais.

Para descrever a mortalidade infantil, tanto a TMI quanto a TEM_0_4 não possuem correlação forte ou moderada com outras variáveis e, diferentemente da situação anterior, representam a mortalidade de faixas etárias bem próximas. Assim, para selecionar uma delas foi levada em que a TMI possui a vantagem de ser calculada com a amplitude de 1 ano (0 a 1 ano), o que minimiza o efeito da composição etária de cada município. Por exemplo, um determinado município possui mais recém nascidos (que possuem uma probabilidade maior de óbito) e outra possui mais crianças de 4 anos, a TEM para a primeira será ilusoriamente maior.

Sendo assim, as variáveis utilizadas para a realização do agrupamento foram: Taxa de mortalidade infantil, Taxa bruta de mortalidade padronizada, Razão de dependência total, Razão de dependência de jovens, Razão de dependência de idosos, Taxa de fecundidade total.

Para a realização do agrupamento dos municípios foi utilizado o método das k -médias, por meio da classe *KMeans* da biblioteca *scikit-learn* (PEDREGOSA et al., 2011). Como o método das k -médias requer que o número de grupos k seja definido a priori, existem alguns métodos para definir o valor de k . Neste trabalho foi empregado o método Ward para esse fim, através da biblioteca *SciPy* (VIRTANEN, P. et al., 2020). A Figura 2 apresenta o dendrograma obtido após a aplicação desse método.

Figura 2- Diagrama de árvore método Ward



Fonte: elaboração própria

A Figura 2 indica que é possível dividir os dados em dois a cinco grupos, na busca por uma maior distância no dendrograma. Com dois e três grupos, não haveria uma boa distinção entre os perfis de municípios formados. Com cinco grupos, a interpretação poderia ficar comprometida. Assim, a quantidade de grupos selecionada para este trabalho foi quatro, ou seja, $k = 4$ para aplicar o agrupamento das k -médias. Após a utilização do método, as medianas de cada variável foram obtidas de forma a traçar os perfis desses municípios similares. Além disso, um mapa com a localização dos municípios pertencentes a cada agrupamento foi obtido.

4. RESULTADOS E DISCUSSÃO

Após a aplicação do método das k -médias, com $k = 4$ grupos para as variáveis selecionadas, foi feito um resumo das características dos grupos. A quantidade de municípios (frequência) em cada grupo (grupos 0, 1, 2, 3), média e a mediana de cada variável por grupo estão expressas na Tabela 2. A linha ‘Brasil’ contém as médias e medianas considerando

todos os municípios do país.

Tabela 2 - Frequência e mediana de cada variável por grupo

	%	frequência		TFT	TBM_p	TMI	RDI	RDT	RDJ
Grupo 0	46,78	2.606	Mediana	1,63	7,14	11,23	0,21	0,54	0,33
			Média	1,65	7,20	11,13	0,20	0,53	0,33
Grupo 1	29,64	1.651	Mediana	1,54	6,27	0,00	0,30	0,60	0,28
			Média	1,55	6,19	6,12	0,31	0,60	0,29
Grupo 2	14,16	789	Mediana	1,45	6,85	34,87	0,27	0,58	0,30
			Média	1,47	6,90	39,61	0,27	0,58	0,31
Grupo 3	9,40	524	Mediana	1,90	6,75	14,49	0,17	0,62	0,44
			Média	2,00	6,71	15,09	0,17	0,63	0,46
Brasil	100,0	5.570	Mediana	1,62	6,79	11,23	0,23	0,57	0,32
			Média	1,65	6,80	12,88	0,23	0,57	0,33

Fonte: elaboração própria

A partir dos grupos formados (Tabela 2), é possível perceber que o Brasil possui municípios com perfis demográficos distintos. Observando os dois maiores grupos (0 e 1), tem-se o primeiro com RDI inferior à mediana de todos os municípios, enquanto o segundo possui uma RDI alta. De forma geral, o Brasil como um todo já está em um processo de envelhecimento avançado, visto que o grupo com a população com maior fecundidade total mediana é o 3, com 1,9 filhos por mulher. Em 1950, este número era de 6,2 para o Brasil inteiro. A taxa de fecundidade mediana de todos os grupos, são inferiores a 2,1. Dessa forma a maior parte dos municípios estão abaixo da taxa de reposição. (VASCONCELOS, A. M. N.; GOMES, M. M. F., 2012).

Mesmo comparando o grupo com a maior TFT mediana com o Brasil inteiro, a quantidade de filhos esperado para as mulheres ao longo de seu período reprodutivo em 2019 para esse grupo corresponde a 30,6%(100 x 1,9/6,2) da quantidade de filhos esperados para as mulheres de quaisquer municípios em 1950.

O grupo 0 corresponde a 46,7% dos municípios brasileiros, ou seja, quase metade deles. Os municípios desse grupo possuem consideravelmente mais jovens do que idosos, visto que sua RDJ é de 33% enquanto a de idosos é de 20%, e sua taxa de mortalidade padronizada é maior que a mediana para todos os municípios. As medianas dos indicadores para esse grupo são bem próximas às medianas para todos os municípios.

Já os municípios do grupo 1 estão em por uma transição demográfica mais adiantada, pois eles apresentam uma grande quantidade de idosos (sua RDI mediana, 30%, é maior que a RDJ mediana, que é 28%) e possuem baixas taxas de mortalidade e fecundidade, sendo 6,27 e 1,65, respectivamente.

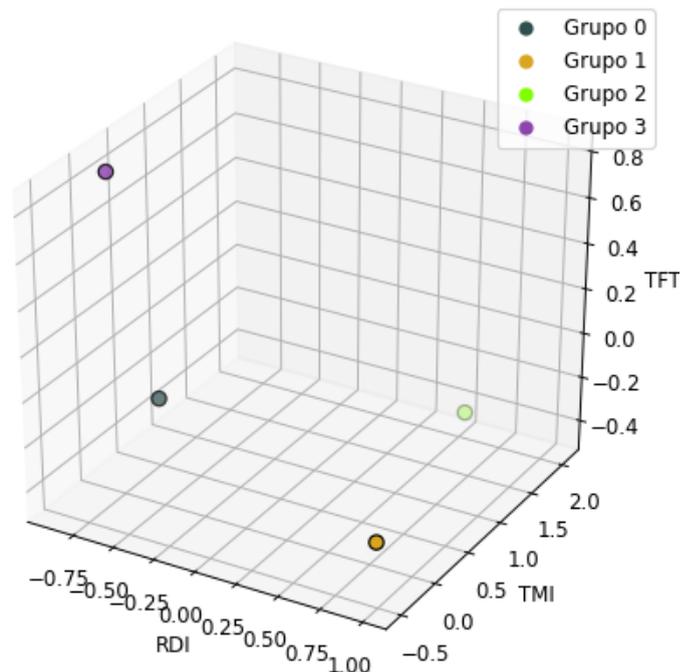
O grupo 2 possui uma grande quantidade de jovens e idosos, sua taxa de fecundidade é baixa (1,45) e sua mortalidade infantil é alta (34,87) se comparada à dos demais grupos. Assim, a transição demográfica nesses municípios é parcialmente adiantada. Por um lado, as

peças têm poucos filhos e há uma quantidade considerável de idosos na população, por outro, a mortalidade infantil é muito alta.

O grupo 3 é o grupo com a população mais jovem dentre todos os grupos, sua RDJ é de 44,5% , enquanto sua RDI é de apenas 17%. Sua TMI é superior à mediana brasileira, e sua TBM_p é inferior à mediana para todos os municípios, e sua TFT mediana (1,9) é a mais alta entre todos os grupos. Esse é o grupo onde a transição demográfica é menos adiantada, visto que possui TFT, TMI e RDJ elevados, e sua TBM_p (6,75) é bem próxima à mediana do Brasil, que é 6,79.

A Figura 3 se trata de um diagrama de dispersão dos centros de cada grupo obtido através do método das *k*-médias. Para possibilitar esse tipo de visualização foi necessário escolher apenas três variáveis, entre as seis variáveis usadas para o agrupamento, a escolha foi feita de forma que tivesse uma variável para representar a mortalidade (No caso foi escolhido a TMI, pois ela apresenta valores mais distantes da mediana para todos os municípios), outra para representar a natalidade (TFT), e outra que descreva a quantidade de idosos (RDI). Isso resulta em uma certa perda de informação. Mesmo assim, optou-se por apresentar o gráfico por ele trazer uma ilustração das diferenças entre os quatro grupos.

Figura 3- Diagrama de dispersão centros dos grupos



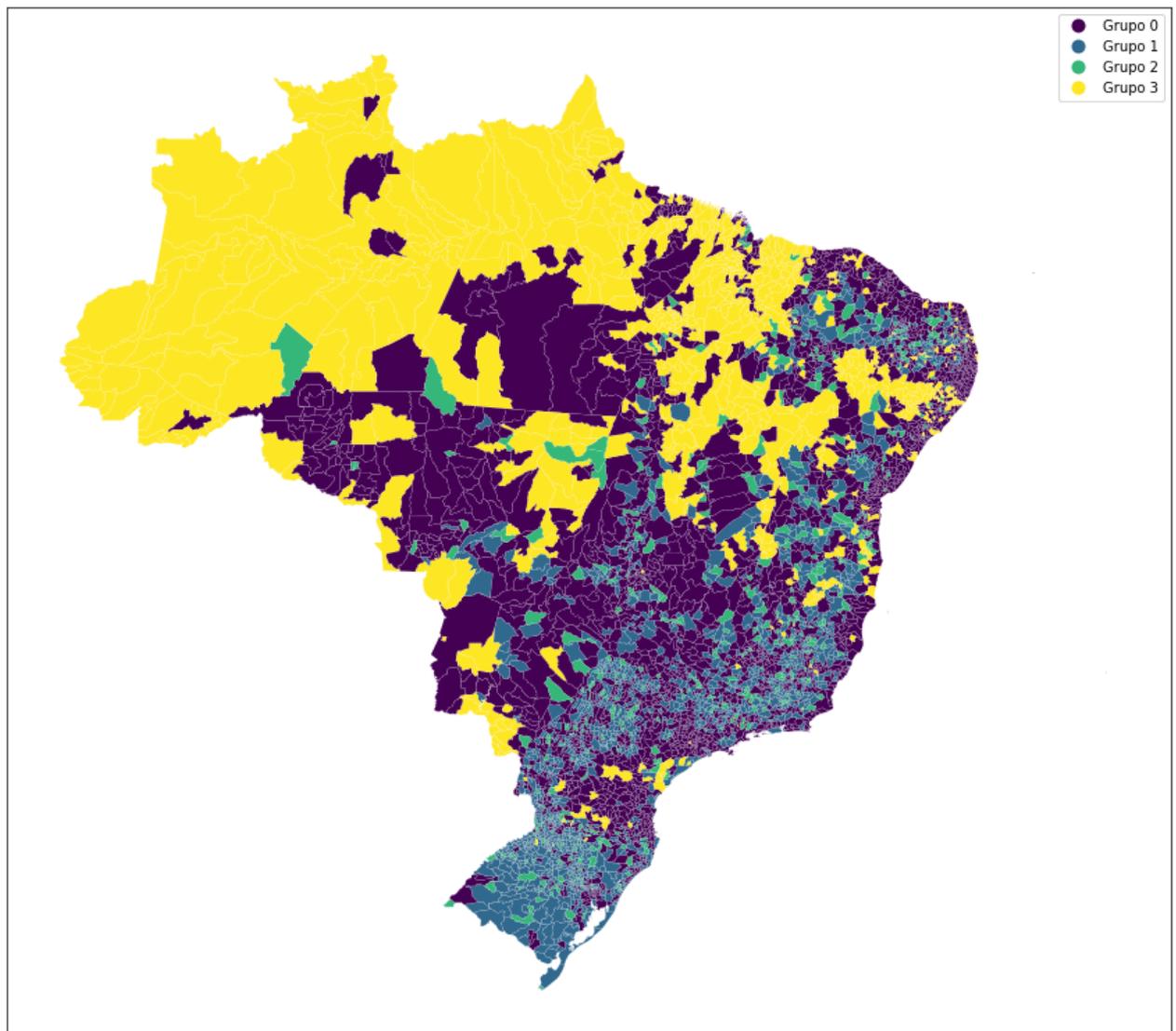
Fonte: Elaboração própria

A Figura 3 mostra que os grupos 0 e 3 estão em um processo de transição demográfica menos adiantados que os outros dois grupos, visto que eles possuem uma razão de dependência de idosos mais baixa. Entre esses, o grupo 3 enfrenta esse processo de forma

mais branda que o 0, pois sua taxa de fecundidade é mais alta. Já os grupos 1 e 2 apresentam um processo de envelhecimento mais rápido, e a principal diferença entre esses grupos é a taxa de mortalidade infantil, que no grupo 2 é maior, o que mostra que a transição demográfica nesse grupo é menos adiantada que no grupo 1.

Na Figura 4 é apresentado um mapa do Brasil com a localização dos municípios pertencentes a cada um dos grupos obtidos.

Figura 4 - Mapa com as localizações dos municípios pertencentes aos quatro grupos



Fonte: Elaboração própria

A partir do mapa da Figura 4 é possível perceber que os municípios do grupo 1, em azul, que são aqueles em que a transição demográfica está ocorrendo de forma mais adiantada, estão concentrados na parte leste e sul do Brasil. Já a parte norte contém mais municípios do grupo 3, em amarelo, onde este processo ocorre de forma menos adiantada. O grupo 0, em

roxo, está mais concentrado na parte central do país, e o grupo 2, em verde, está disperso por todo o território brasileiro porém é mais concentrado próximo a municípios do grupo 1.

5. CONSIDERAÇÕES FINAIS

Os métodos usados neste trabalho com os dados utilizados para o agrupamento, sugeriram que os municípios brasileiros fossem divididos em quatro grupos. O agrupamento foi feito via k médias, e obteve-se quatro perfis, com diferenças bem traçadas. Cerca de 46,78% dos municípios estão contidos no grupo 0, nesse grupo a população dependente é predominantemente jovem (33%), e a taxa bruta de mortalidade padronizada é um pouco elevada.

O segundo maior grupo (29,64% dos municípios) obtido por essa técnica é o 2, e diferentemente do grupo anterior, este possui uma população dependente mediana, com mais idosos(30%) do que jovens (28%). Juntos, esses dois grupos somam mais de 75 por cento dos municípios brasileiros.

Ambos já sofrem com os efeitos da transição demográfica, entretanto no grupo 1 esses efeitos estão ocorrendo de forma mais avançada. Os menores grupos (2,3) embora não representem uma parte grande dos municípios, foram importantes para análise pois identificaram um grupo de municípios, que possui uma transição demográfica significativamente menos adiantada que nos outros grupos (3), e outro grupo (2), que embora possua municípios com taxa de fecundidade total baixa e razão de dependência de idosos elevada. possuem também uma taxa de mortalidade infantil elevada.

Assim o objetivo deste trabalho foi alcançado. A ausência de indicadores municipais recentes impossibilita um melhor aproveitamento do algoritmo, além disso os dados têm amplitude alta o que impacta o resultado dos indicadores calculados para o agrupamento, de forma que inviabilize o cálculo de outros indicadores que poderiam ser benéficos ao agrupamento.

6. REFERÊNCIAS BIBLIOGRÁFICAS

ALVES, J. E. D. A transição demográfica e a janela de oportunidade. 2008. Disponível em: https://www.ufjf.br/ladem/files/2009/05/transicao_demografica.pdf

CARVALHO, J. A. M.; SAWYER, D. O.; RODRIGUES, R. N. Introdução a alguns conceitos básicos e medidas em demografia. **Associação Brasileira de Estudos Populacionais**, 2ª edição, 1998.

CASTIGLIONI, A. H. Transição urbana e demográfica no Brasil: características, percursos e tendências. **Ateliê Geográfico**, Goiânia, v. 14, n. 1, p. 6–26, 2020.

DATASUS – Departamento de Informática do Sistema Único de Saúde. 2019 Disponível em: http://tabnet.datasus.gov.br/cgi/menu_tabnet_php.htm

GUMINIAK, S. C. Crise previdenciária no Brasil: o impacto da transição demográfica sobre o fluxo de caixa atual e potencial do RGPS até 2060. 2017. Disponível em: <https://repositorio.uniceub.br/jspui/bitstream/235/11407/1/21554960.pdf>

MATUDA, N. S. Introdução a demografia notas de aula, 2009. Disponível em: http://wiki.dpi.inpe.br/lib/exe/fetch.php?media=ser457-cst310:aulas-2014:leituras:matuda_2009.pdf

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

PAIVA, P. T. A.; WAJNMAN, S. Das causas às consequências econômicas da transição demográfica no Brasil. 2005. Disponível em: <https://doi.org/10.1590/S0102-30982005000200008>

PALMA, L. F. Agrupamento de dados: k-médias. 2018, Disponível em: https://www2.ufrb.edu.br/bcet/components/com_chronoforms5/chronoforms/uploads/tcc/20190604200511_2018.2_TCC_Luann_Farias_Palma-_Agrupamento_de_dados_-_K_medias.pdf

REIS, C.; BARBOSA, L. M. L. H.; PIMENTEL, V. P. O desafio do envelhecimento populacional na perspectiva sistêmica da saúde. **BNDES Setorial**, Rio de Janeiro, n. 44, p. [87]-124, set. 2016. Disponível em: <http://web.bndes.gov.br/bib/jspui/handle/1408/9955>

PEDREGOSA et al., Scikit-learn: Machine Learning in Python, JMLR 12, pp. 2825-2830, 2011.

SOUZA, L. G.; RAMOS, P. S.; FRIAS, L. O envelhecimento populacional nos municípios do Sul/Sudoeste de Minas Gerais: uma análise de agrupamento. **Revista Debate Econômico**, v. 5, n. 2, 2017.

VASCONCELOS, A. M. N.; GOMES, M. M. F. Transição demográfica: a experiência brasileira. 2012. Disponível em: http://scielo.iec.gov.br/scielo.php?script=sci_arttext&pid=S1679-49742012000400003

VICINI, L., Análise multivariada da teoria à prática, 2005 Disponível em: <http://w3.ufsm.br/adriano/livro/Caderno%20dedatico%20multivariada%20-%20LIVRO%20FINAL%201.pdf>

VIRTANEN, P. et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. **Nature Methods**, 17, pp.261–272, 2020.