

UNIVERSIDADE FEDERAL DE ALFENAS – UNIFAL– MG  
INSTITUTO DE CIÊNCIAS SOCIAIS APLICADAS – ICSA

**ALEX DE SOUSA PEREIRA**

**OS PRINCIPAIS SETORES DE EMPREGO NA MESORREGIÃO DO  
SUL/SUDOESTE DE MINAS: UMA ANÁLISE MULTIVARIADA**

VARGINHA – MG

2021

**ALEX DE SOUSA PEREIRA**

**OS PRINCIPAIS SETORES DE EMPREGO NA MESORREGIÃO DO  
SUL/SUDOESTE DE MINAS: UMA ANÁLISE MULTIVARIADA**

Trabalho de conclusão de Piepex apresentado ao Instituto de Ciências Aplicadas da Universidade Federal de Alfenas como requisito parcial à obtenção do título de Bacharel em Ciência e Economia.

Orientador: Prof. Dr. Adriano Antonio Nuintin

**ALEX DE SOUSA PEREIRA**

**OS PRINCIPAIS SETORES DE EMPREGO NA MESORREGIÃO DO  
SUL/SUDOESTE DE MINAS: UMA ANÁLISE MULTIVARIADA**

A banca examinadora abaixo-assinada, aprova o trabalho de conclusão do PIEPEX (TCP), apresentado como parte dos requisitos para a obtenção do grau de Bacharel em Ciência e Economia pelo Instituto de Ciências Sociais Aplicadas da Unifal-MG.

Trabalho Aprovado em: 17/09/2021

Prof. Dr. Adriano Antonio Nuintin  
Universidade Federal de Alfenas

Assinatura: \_\_\_\_\_

Prof<sup>a</sup>. Dr<sup>a</sup>. Maria Aparecida Curi  
Universidade Federal de Alfenas

Assinatura: \_\_\_\_\_

Prof<sup>a</sup>. Dr<sup>a</sup>. Patrícia de Siqueira Ramos  
Universidade Federal de Alfenas

Assinatura: \_\_\_\_\_

## **AGRADECIMENTOS**

Primeiramente, gostaria de agradecer a Deus pela oportunidade que me foi dada de estar concluindo minha graduação e realizando meus sonhos.

Agradeço a toda a minha família, em especial a meus pais, por todo apoio e por me ajudarem sempre quando eu preciso.

Aos meus amigos André, Cláudio, Jessica e Júlio, que sempre estiveram ao meu lado, por nossa toda nossa amizade e por todas as vezes que me auxiliaram no decorrer do curso.

Ao meu orientador Adriano, por toda a ajuda durante a elaboração deste trabalho, sua disponibilidade e educação.

Ao Instituto de Ciências Sociais Aplicadas da UNIFAL-MG, por disponibilizar uma estrutura de ensino de qualidade e a oportunidade de cursar a graduação.

Por fim, a todos aqueles que contribuíram de alguma forma para que eu chegasse até este momento, me incentivando, não permitindo que eu desistisse e tornando mais leve a caminhada para esta conquista.

## RESUMO

O presente trabalho é uma análise empírica quantitativa que teve por objetivo verificar como se comportam os índices de empregabilidade (trabalhadores admitidos, demitidos e número de trabalhadores acumulados), para os municípios da Mesorregião do Sul/Sudoeste de Minas Gerais. Dessa forma, utilizou-se de técnicas de análise multivariada a fim de chegar ao objetivo proposto por este estudo. Os dados utilizados compõem um conjunto de 14 variáveis, retirados diretamente do CAGED no ano de 2020, sendo as informações referentes aos setores da economia: agropecuário, comércio, construção civil, indústria e serviços. As análises estabelecidas foram feitas através da linguagem de programação Python. Entre os métodos multivariados utilizados, empregou-se a análise de componentes principais e a análise de agrupamento através dos métodos hierárquico (Ward) e não hierárquico (k-médias). Os resultados obtidos demonstraram que os municípios de Varginha, Poços de Caldas, Pouso Alegre e Extrema foram os municípios que apresentaram maior índice de empregabilidade para a maioria dos setores da economia. Do mesmo modo, observou-se que os municípios como Monte Belo, Carmo da Cachoeira e Conceição do Rio Verde se demonstraram expressivos na geração de emprego no setor agropecuário. Por fim, Alfenas, Três Pontas e São Sebastião do Paraíso foram municípios que se apresentaram com altos índices de empregabilidade em todos os setores analisados.

**Palavras-chave:** Emprego; Desemprego; Minas Gerais; Análise de Agrupamento e Principais.

## SUMÁRIO

1. INTRODUÇÃO	6
2. FUNDAMENTAÇÃO TEÓRICA	8
2.1. Emprego e desemprego no sul de Minas	8
2.2. Análise multivariada	9
3. METODOLOGIA	14
4. ANÁLISE E DISCUSSÃO DOS RESULTADOS	17
4.1. Análise descritiva	17
4.2. Análise de componentes principais (ACP)	19
4.3. Análise de agrupamento (AA)	22
4.3.1. Método Ward	22
4.3.2. Método de K-médias	23
5. CONSIDERAÇÕES FINAIS	26
REFERÊNCIAS BIBLIOGRÁFICAS	28

## 1. INTRODUÇÃO

A economia mineira destaca-se por apresentar um dos maiores PIB (Produto Interno Bruto) do Brasil, e o Sul de Minas é uma das regiões mais importantes para geração da economia no estado (CENTRAL, 2018). A agricultura pode ser considerada como um dos setores mais bem explorados dentro desta região, com destaque principalmente no mercado de café como uns dos maiores produtores do mundo (ANDRADE, 1994). Também, destacam-se os setores de serviços, comércio e atividade industrial, beneficiados principalmente por sua proximidade com o estado de São Paulo (CENTRAL, 2018).

Hoje a Mesorregião do Sul/Sudoeste de Minas Gerais é composta por 146 municípios, sendo estes municípios distribuídos entre as seguintes microrregiões: Alfenas, Andrelândia, Itajubá, Passos, Pouso Alegre, Poços de Caldas, Santa Rita do Sapucaí, São Lourenço, São Sebastião do Paraíso e Varginha, conforme o IBGE (2021). Da mesma forma, a economia decorrente destes municípios é baseada nos setores da economia (Agropecuária, Comércio, Construção, Indústria e Serviços) conforme argumenta o CAGED (Cadastro Geral de Empregados e Desempregados).

Este trabalho tem como objetivo analisar como se comportam os índices de empregabilidade (admitidos, demitidos e número acumulado de trabalhadores) para os municípios da Mesorregião do Sul/Sudoeste de Minas Gerais. Para tanto, utilizou-se de técnicas multivariadas com a finalidade de analisar os dados computados pelo CAGED 2020. Foram empregadas duas técnicas para este trabalho: análise de componentes principais e análise de agrupamento (método Ward e k-médias).

A análise multivariada trata-se de um conjunto de técnicas estatísticas com objetivo de analisar várias variáveis ao mesmo tempo e auxiliar na tomada de decisão (HAIR et al., 2009). Uma dessas técnicas é a análise de componentes principais, que analisa aspectos de geração, seleção e interpretação a partir das variáveis investigadas (VICINI, 2005). Outra técnica bastante utilizada trata-se da análise de agrupamento ou análise de clusters, que tem por finalidade, agrupar objetos similares com base nas características que eles possuem (HAIR et al., 2009). Desse modo, a partir deste estudo, é possível ter uma visão mais clara de como os índices de empregabilidade se comportam nos municípios da Mesorregião.

Este trabalho foi construído com base na seguinte estrutura: além desta introdução, na seção 2, “Fundamentação Teórica”, apresentam-se os caminhos teóricos que subsidiam a construção deste trabalho, sendo apresentado como se comporta a empregabilidade no sul de Minas Gerais, os setores da economia conforme CAGED e os métodos multivariados utilizados.

Na seção 3, “Metodologia”, é apresentada a metodologia utilizada, explicitando-se as técnicas, conceitos e dados utilizados. Na seção 4 “Análise e Discussão dos Resultados”, são expostos os resultados obtidos com a análise e sua a discussão. E por fim, na seção 5 apresentam-se as considerações finais sobre a pesquisa.



## 2. FUNDAMENTAÇÃO TEÓRICA

### 2.1. Emprego e desemprego no sul de Minas

A economia mineira destaca-se por ser uma das maiores participações no PIB (Produto Interno Bruto) brasileiro, sendo o Sul de Minas uma das regiões mais importantes para geração da economia no Estado. Conforme Central (2018), contribuem para este crescimento econômico principalmente os setores de serviços, comércio e atividade industrial, beneficiados principalmente por sua aproximação com o estado de São Paulo.

Já para Prates et al. (2010), em uma análise referente ao crescimento econômico na região Sul/Sudoeste de Minas entre os anos de 2000 a 2008, a Mesorregião não apresentou ganhos significativos na economia regional durante os anos. No entanto, a região ainda apresentou uma forte tendência ativa e receptora para o desenvolvimento da economia do estado.

Outro setor relevante ao crescimento econômico do Sul de Minas é o setor agropecuário, com ênfase no mercado de café. Andrade (1994) afirma que no Sul de Minas este mercado é composto em sua maioria por pequenos produtores. Todavia, segundo Vivaldi, Júnior e Alves (2020), apesar deste grande impacto para este setor da economia, o setor tem se modernizado cada vez mais, dificultando na geração de novos empregos.

A Tabela 1 apresenta o número de empregados referentes a cada setor da economia no ano de 2020.

Tabela 1 - Número de empregados acumulados nos setores da economia em 2020.

Microrregião	Agropecuária	Comércio	Construção	Indústria	Serviços
Alfenas	11,03%	8,94%	8,16%	5,59%	8,05%
Andrelândia	3,91%	1,61%	0,66%	1,74%	1,04%
Itajubá	2,93%	5,65%	3,71%	8,11%	5,92%
Passos	10,19%	8,74%	12,28%	5,34%	8,67%
Pouso Alegre	6,27%	18,96%	15,47%	23,15%	21,33%
Poços de Caldas	10,07%	15,11%	15,50%	15,89%	18,15%
Santa Rita do Sapucaí	6,56%	3,48%	12,06%	7,86%	3,26%
São Lourenço	10,32%	7,48%	5,29%	5,02%	6,75%
São Sebastião do Paraíso	13,75%	10,77%	8,54%	11,89%	8,55%
Varginha	24,97%	19,27%	18,33%	15,41%	18,28%
<b>Total</b>	<b>100,00%</b>	<b>100,00%</b>	<b>100,00%</b>	<b>100,00%</b>	<b>100,00%</b>

Fonte: Elaboração própria a partir de dados do CAGED (2021).

É possível observar que a microrregião de Varginha se destaca na geração de emprego nos setores de agropecuária, comércio e construção civil. Além disso, os municípios de Três Pontas, Andradas e Alfenas apresentaram o maior número de trabalhadores no setor agropecuário. No que se diz ao setor de comércio, as cidades de Pouso Alegre, Poços Caldas e Varginha foram as que tiveram maior número de trabalhadores. Por fim, em relação ao setor de construção civil, Poços de Caldas e Varginha foram as cidades que tiveram o maior número de trabalhadores.

Já a microrregião de Pouso Alegre destaca-se na geração de emprego nos setores de indústria e serviços. Todavia, no setor da indústria, os municípios de Pouso Alegre, Extrema e Poços Caldas tiveram a maior proporção de empregados. Já o setor de serviços, conseqüentemente, Poços Caldas, Pouso Alegre e Varginha foram as cidades que apresentaram o maior número de trabalhadores.

Desta forma, este trabalho propõe analisar o nível de emprego e desemprego nos municípios da Mesorregião do Sul/Sudoeste de Minas Gerais. O próximo tópico abordará de forma mais clara quais variáveis serão utilizadas neste estudo.

## **2.2. Análise multivariada**

Segundo Hair et al. (2009), a análise multivariada envolve um conjunto de técnicas estatísticas com o objetivo de analisar várias variáveis e auxiliar na tomada de decisão. Do mesmo modo, Grimm e Yarnold (2000) afirmam que a análise multivariada é aplicada em uma série de técnicas analíticas, fornecendo uma melhor compreensão de suas medidas de estudo.

Deste modo, o uso de técnicas multivariadas se vê cada vez mais presente nos trabalhos cujo caráter seja de análise quantitativa. É recomendável, conforme Ferreira (2008), o uso deste tipo de metodologia nas seguintes situações:

- Redução de dados ou simplificação estrutural: o caso a ser estudado é representado da maneira mais fácil possível, sem que se deixem informações importantes e tornando-as de fácil entendimento;
- Ordenação e agrupamento: agrupamento de observações ou variáveis similares, baseando-se em dados retirados de uma amostra ou em experimentos;
- Investigação da dependência entre variáveis: estuda-se as relações fundamentais entre as variáveis, que muitas vezes são de interesse do pesquisador;

- Predição: relacionam-se variáveis com objetivo de estimar o valor de uma ou mais, com base no número de observações de cada elemento do conjunto de dados;
- Construção e teste de hipóteses.

No que diz respeito ao aspecto dos dados multivariados, eles são representados por uma amostra aleatória com  $n$  observações, e para cada unidade amostral existem variáveis. Assim, têm-se uma matriz  $X$  com  $n$  (linhas) e  $p$  (colunas):

$$X_{n \times p} = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1p} \\ X_{21} & X_{22} & \dots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \dots & X_{np} \end{bmatrix}$$

em que,  $n$  representa o número de observações e  $p$  o número de variáveis do conjunto de dados.

Um exemplo prático de análise de multivariada pode ser dito com análise de componentes principais. Segundo Vicini (2005), a ACP (Análise de Componentes Principais), tem como objetivo analisar aspectos de geração, seleção e interpretação das variáveis investigadas, de modo que, pretende-se investigar os componentes com maior influência na formação de cada elemento. Deste modo, conforme Hair *et al.* (2009), esta técnica possui como finalidade condensar a informação de várias variáveis originais em um conjunto mínimo de variáveis, denominado fatores, em que se têm uma perda mínima de informação.

Dessa forma, a aplicação da ACP dentro dos setores da economia permite minimizar redundância de diversas variáveis quando analisamos diversos setores de emprego de uma típica região.

Conforme Ferreira (2008), a ACP pode ser definida como  $i$ -ésimo componente principal ( $Y_i$ ), dado que a covariância possua autovalores  $(\lambda_i, e_i)$ ,  $i=1, 2, \dots, p$ , dado que  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ .

$$Y_i = e_i^t X = e_{i1} X_1 + e_{i2} X_2 + \dots + e_{ip} X_p \quad i=1, 2, \dots, p$$

sendo  $X$  o vetor que contém  $p$  variáveis após uma normalização e  $t$  representa que tal vetor (ou matriz) é transposto.

Após efetuado o cálculo dos componentes principais, tende-se a verificar o total da variação acumulada explicada por eles. De tal modo, a porcentagem da variação do  $k$ -ésimo componente principal, como exposto por Ferreira (2008), é dada por:

$$\%VarExp(Y_k) = \frac{\lambda_k}{\sum_{i=1}^p \lambda_i} \times 100$$

Por fim, o número máximo de componentes a serem considerados na análise poderá ser definido pelos seguintes critérios:

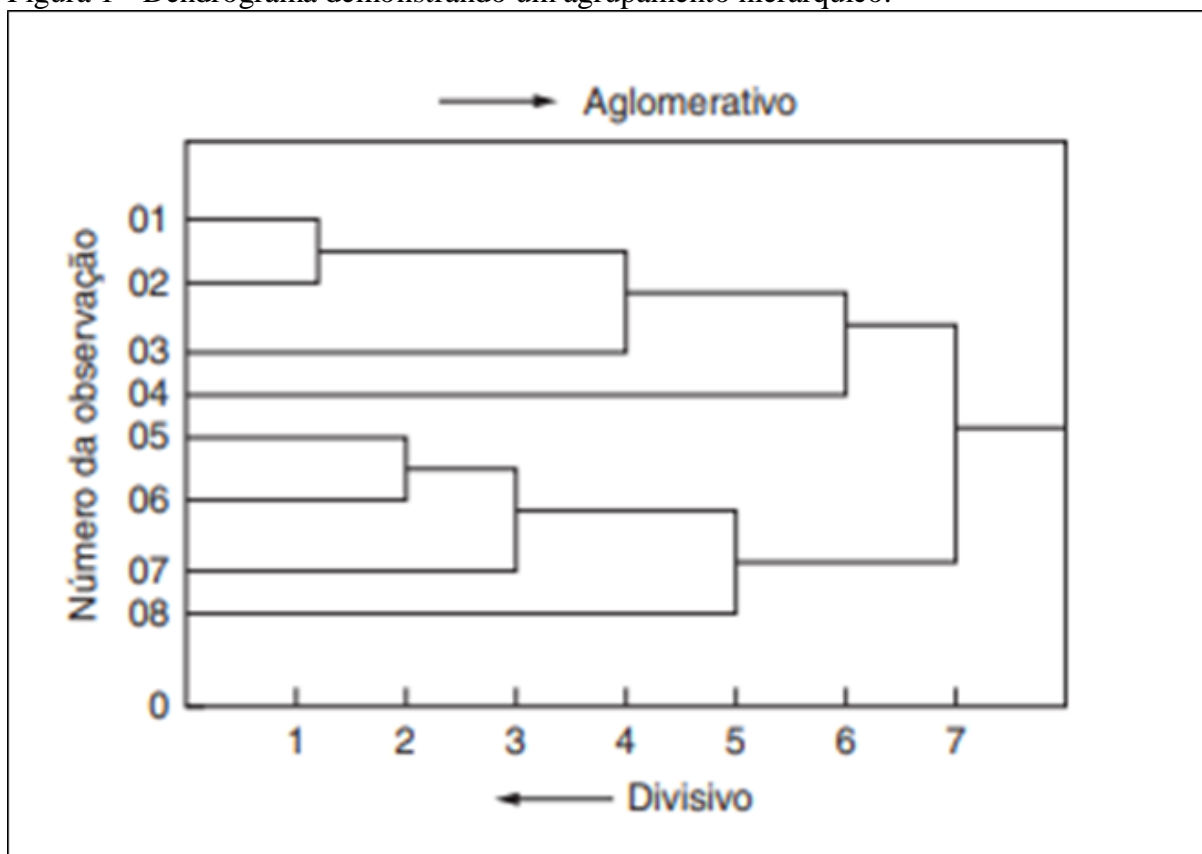
- I. Reter os  $k$  primeiros componentes que explicam grande parte (70% a 80%) da variação total;
- II. Caso a matriz  $R$  seja usada, somente com os componentes que possuam valores de  $\lambda > 1$  (KAISER, 1958). Outros autores optam por  $\lambda > 0,7$  (JOLLIFFE, 1972);
- III. Analisar um scree plot (gráfico de número de componentes). Deve-se procurar pelo “cotovelo”, ou seja, o ponto depois do qual a variação explicada diminui lentamente;
- IV. Verificar se o componente tem interpretação razoável e útil.

Outro método de análise multivariada é a AA (Análise de Agrupamento) ou AC (Análise de *Clusters*), pode ser definida segundo Hair *et al.* (2009), como um conjunto de técnicas multivariadas com o objetivo de agregar objetos com base nas características que eles possuem. De forma análoga, como argumenta Vicini (2005), o principal objetivo desta técnica é reunir todas as observações (indivíduos, elementos), verificando se existe homogeneidade dentro do grupo e heterogeneidade entre os grupos, propondo classificações entre eles.

Contudo, este tipo de análise pode ser feito de duas maneiras: por métodos hierárquicos ou não hierárquicos.

A utilização de métodos de agrupamento hierárquico tem como finalidade a criação de grupos baseados na árvore de decisão (dendrograma), que permite encontrar os grupos ou pontos de fusão baseados em cada estágio (Vicini, 2005). A Figura 1 demonstra um exemplo de dendrograma para oito observações.

Figura 1 - Dendrograma demonstrando um agrupamento hierárquico.



Fonte: Retirado de (HAIR *et al.*, 2009).

Os procedimentos hierárquicos podem ser divididos de duas formas: aglomerativos, em que cada observação começa com seu próprio agrupamento; e divisivos, em que todas as observações começam como um grande agrupamento e depois são divididas em 2 ou mais grupos até que suas observações se tornem um agrupamento cada uma (HAIR *et al.*, 2009).

Os métodos hierárquicos aglomerativos são considerados os mais usados na literatura por exigirem um menor processamento computacional (HAIR *et al.*, 2009). Entre os tipos de métodos hierárquicos, destacam-se os seguintes como principais, conforme apontam Hair *et al.* (2009):

- Vizinho Mais Próximo: sua finalidade é encontrar todas as menores distâncias entre as observações e unir as mais próximas;
- Vizinho Mais Distante: sua finalidade é encontrar a distância entre as distâncias entre as observações e unir as mais distantes;
- Distância Média: sua finalidade é encontrar a distância média entre as observações;
- Centroeide: sua finalidade é encontrar a distância medida entre os vetores médios dos grupos, denominados centroides;

- Ward: sua finalidade é minimizar a soma dos quadrados nos agrupamentos entre todas as variáveis.

Os métodos de agrupamento não hierárquicos são aqueles em que há a necessidade de se definir o número de grupos antes de se designar as observações. Este tipo de método é utilizado a fim de agregar todas as observações possíveis usando pontos de sementes iniciais com a finalidade de obter alocações mais precisas (HAIR et al., 2009).

Um dos principais métodos não hierárquicos, o das k-médias, pode ser considerado como um dos mais utilizados. Este método é dividido nas seguintes etapas, conforme apresenta Ferreira (2008):

- I. Particionar os itens em k grupos iniciais arbitrariamente;
- II. Percorrer a lista de itens e calcular as distâncias de cada um deles para o centroide (médias) dos grupos;
- III. Fazer a realocação do item para o grupo em que ele apresentar mínima distância, obviamente se não for o grupo ao qual pertença;
- IV. Recalcular os centroides dos grupos que ganharam e perderam o item;
- V. Repetir os passos II, III e IV até que nenhuma alteração seja feita.

A seção seguinte demonstra a metodologia utilizada para a elaboração deste trabalho.

### 3. METODOLOGIA

O objetivo principal deste estudo, para construção do modelo multivariado, consistiu em analisar quais seriam os setores mais representativos no que tange a empregabilidade na Mesorregião do Sul/Sudoeste de Minas Gerais. Como primeira etapa, foi feita uma análise descritiva a fim de analisar como se comporta o nível de emprego e desemprego dentro da Mesorregião do Sul/Sudoeste de Minas Gerais.

O estudo possui um enfoque quantitativo, baseando-se em um projeto de levantamento, que segundo Creswell (2010), corresponde ao teste de um impacto, sobre um resultado, controlando todos os possíveis fatores que possam impactar neste resultado. Para a elaboração deste trabalho, utilizou-se dos seguintes procedimentos como ilustra a Figura 2.

Figura 2 - Etapas do procedimento metodológico



Fonte: Elaboração própria.

Primeiramente, usou-se da fundamentação teórica para explicar os princípios de técnica multivariada, com base nos trabalhos de Hair *et al.* (2009), Grimm e Yarnold (2000), Vicini (2005) e Ferreira (2008). Já para as análises exercidas, utilizou-se como fundamentação os trabalhos de Kubrusly e Sabóia (2006), Floriano (2019) e Cunha, Modotti e Setti (2008).

Já a coleta dos dados foi feita através do CAGED (Cadastro Geral de Empregados e Desempregados) no ano de 2020. De maneira que, foram coletados dados de 146 municípios que compõem a Mesorregião do Sul/Sudoeste de Minas Gerais. As variáveis propostas dizem respeito ao emprego, desemprego e número de trabalhadores acumulados de acordo com a classificação dos setores da economia conforme demonstra o CAGED (2020): agropecuário, comércio, construção civil, industrialização e serviços. O quadro 1 demonstra as variáveis utilizadas neste estudo.

Quadro 1 - Índices Empregabilidade nos Setores da Economia

Variável	Definição
Adm.Agro	Representa o número de trabalhadores admitidos no setor de agropecuária no ano de 2020.
Desl.Agro	Representa o número de trabalhadores desligados no setor de agropecuária no ano de 2020.
Trab.Agro	Representa o número de trabalhadores acumulado no setor de agropecuária no ano de 2020.
Adm.Comer	Representa o número de trabalhadores admitidos no setor de comércio no ano de 2020.
Desl.Comer	Representa o número de trabalhadores desligados no setor de comércio no ano de 2020.
Trab.Comer	Representa o número de trabalhadores acumulado no setor de comércio no ano de 2020.
Adm.Const	Representa o número de trabalhadores admitidos no setor de construção no ano de 2020.
Desl.Const	Representa o número de trabalhadores desligados no setor de construção no ano de 2020.
Trab.Const	Representa o número de trabalhadores acumulado no setor de construção no ano de 2020.
Adm.Indust	Representa o número de trabalhadores admitidos no setor de indústria no ano de 2020.
Desl.Indust	Representa o número de trabalhadores desligados no setor de indústria no ano de 2020.
Trab.Indust	Representa o número de trabalhadores acumulados no setor de indústria no ano de 2020.
Adm.Serv	Representa o número de trabalhadores admitidos no setor de serviços no ano de 2020.
Desl.Serv	Representa o número de trabalhadores demitidos no setor de serviços no ano de 2020.
Trab.Serv	Representa o número de trabalhadores acumulados no setor de serviços no ano de 2020.

Fonte: Elaboração própria, a partir dados do CAGED (2021).

Para tanto, a fim de utilizar-se a análise multivariada para ver como se comportam os municípios dentro da Mesorregião do Sul/Sudoeste de Minas Gerais, foi-se necessário padronizar as variáveis a fim de deixá-las na mesma escala. Deste modo, cada observação  $x_{ij}$  foi padronizada da seguinte forma, como especifica Hair et al. (2009):

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j},$$

sendo:

- $z_{ij}$ : valor normalizado da observação  $i$  na variável  $j$ ;
- $x_{ij}$ : valor da observação  $i$  na variável  $j$ ;
- $\bar{x}_j$ : valor médio da variável  $j$ ;
- $s_j$ : valor do desvio padrão da variável  $j$ ;

O processo de análise multivariada, se deu através de dois métodos: análise de componentes principais (ACP) através da matriz de correlação e pela análise de agrupamento (AA). Deste modo, utilizou-se do software de programação Python (2021) juntamente com os pacotes:

- pandas (para lidar com o conjunto de dados);
- numpy e scipy (para lidar com os cálculos vetoriais e matriciais);
- [altair, matplotlib e seaborn] (para gerar os gráficos);
- sklearn (para a análise de componentes principais);



- [scipy.cluster.hierarchy, scipy.spatial.distance, scipy.cluster.hierarchy e sklearn.cluster] (para a análise de agrupamento).

Os dados analisados tinham estrutura  $n \times p$ , sendo:  $n = 146$  observações (correspondente a cada município da Mesorregião Sul/Sudoeste de Minas Gerais) e  $p = 17$  variáveis (duas referente ao município/microrregião e, as demais, referentes às variáveis propostas pelo estudo).

No que concerne a Análise de Componentes Principais, objetivou-se de reduzir o número de variáveis propostas pelo estudo em componentes. Já análise de agrupamento, com finalidade de reduzir o número de observações em grupos. Contudo, utilizou-se tanto do método hierárquico quanto não hierárquico, todos utilizando da distância euclidiana para mensuração dos grupos.

Para a elaboração do procedimento hierárquico, foram utilizados 5 métodos hierárquicos: o vizinho mais próximo, vizinho mais distante, distância média, centroide e Ward. Observou-se que os grupos dos municípios ficaram melhor distribuídos pelo método Ward, sendo o ideal para este tipo de análise.

A escolha do número de grupos se deu através do corte do dendrograma, que tem como finalidade encontrar onde se tem o maior distanciamento entre um grupo e outro. Deste modo, observou-se que o número de grupos é 2 pelo método de Ward.

O método não hierárquico utilizado foi o de k-médias. No qual, após definido o número de grupos pelo método Ward, analisou-se as estatísticas descritivas (média e mediana) de cada grupo. Consequentemente, foi elaborado um gráfico que relaciona os agrupamentos com os componentes principais e os municípios utilizados neste estudo.

## 4. ANÁLISE E DISCUSSÃO DOS RESULTADOS

### 4.1. Análise descritiva

Nesta seção são apresentadas as análises sobre as estatísticas descritivas e a análise de correlação entre as variáveis. A Tabela 2 apresenta o resumo estatístico, com os valores mínimos, máximos, médios, os quartis e desvios padrões das variáveis estudadas

Tabela 2 - Estatísticas descritivas do número empregados, demitidos e trabalhadores dos setores da economia.

	<b>Agropecuária</b>	<b>Comércio</b>	<b>Construção</b>	<b>Indústria</b>	<b>Serviços</b>
<b>Admitidos</b>					
<b>Count</b>	146,00	146,00	146,00	146,00	146,00
<b>Mean</b>	53,34	357,10	118,73	350,41	414,13
<b>Std</b>	105,85	851,91	341,96	690,54	1328,31
<b>Min</b>	0,00	0,00	0,00	0,00	0,00
<b>25%</b>	1,00	31,25	2,00	25,50	21,25
<b>50%</b>	11,00	95,00	11,00	107,00	64,00
<b>75%</b>	48,00	256,00	47,00	242,75	191,50
<b>Max</b>	715,00	5265,00	2687,00	4352,00	8838,00
<b>Demitidos</b>					
<b>Count</b>	146,00	146,00	146,0	146,00	146,00
<b>Mean</b>	50,69	340,80	106,4	338,65	419,92
<b>Std</b>	99,31	820,06	293,1	666,25	1348,99
<b>Min</b>	0,00	0,00	0,0	0,00	0,00
<b>25%</b>	2,00	28,00	3,0	23,25	17,75
<b>50%</b>	9,50	83,50	12,0	104,50	60,50
<b>75%</b>	52,75	243,00	42,5	282,50	196,50
<b>Max</b>	658,00	5254,00	1739,0	4071,00	9406,00
<b>Total de Trabalhadores</b>					
<b>Count</b>	146,00	146,00	146,00	146,00	146,00
<b>Mean</b>	383,78	868,31	136,29	950,49	1194,27
<b>Std</b>	460,77	1871,27	387,73	1969,10	3261,31
<b>Min</b>	5,00	3,00	0,00	0,00	4,00
<b>25%</b>	102,25	102,00	4,00	92,25	102,50
<b>50%</b>	203,50	256,50	18,50	235,00	271,50

Tabela 2 - Estatísticas descritivas do número empregados, demitidos e trabalhadores dos setores da economia.

	<b>Agropecuária</b>	<b>Comércio</b>	<b>Construção</b>	<b>Indústria</b>	<b>Serviços</b>
	<b>Total de Trabalhadores</b>				
<b>75%</b>	462,25	621,00	61,25	738,25	772,75
<b>Max</b>	2864,00	11373,00	2413,00	12994,00	23484,00

Fonte: Elaboração própria, a partir de dados do CAGED (2020).

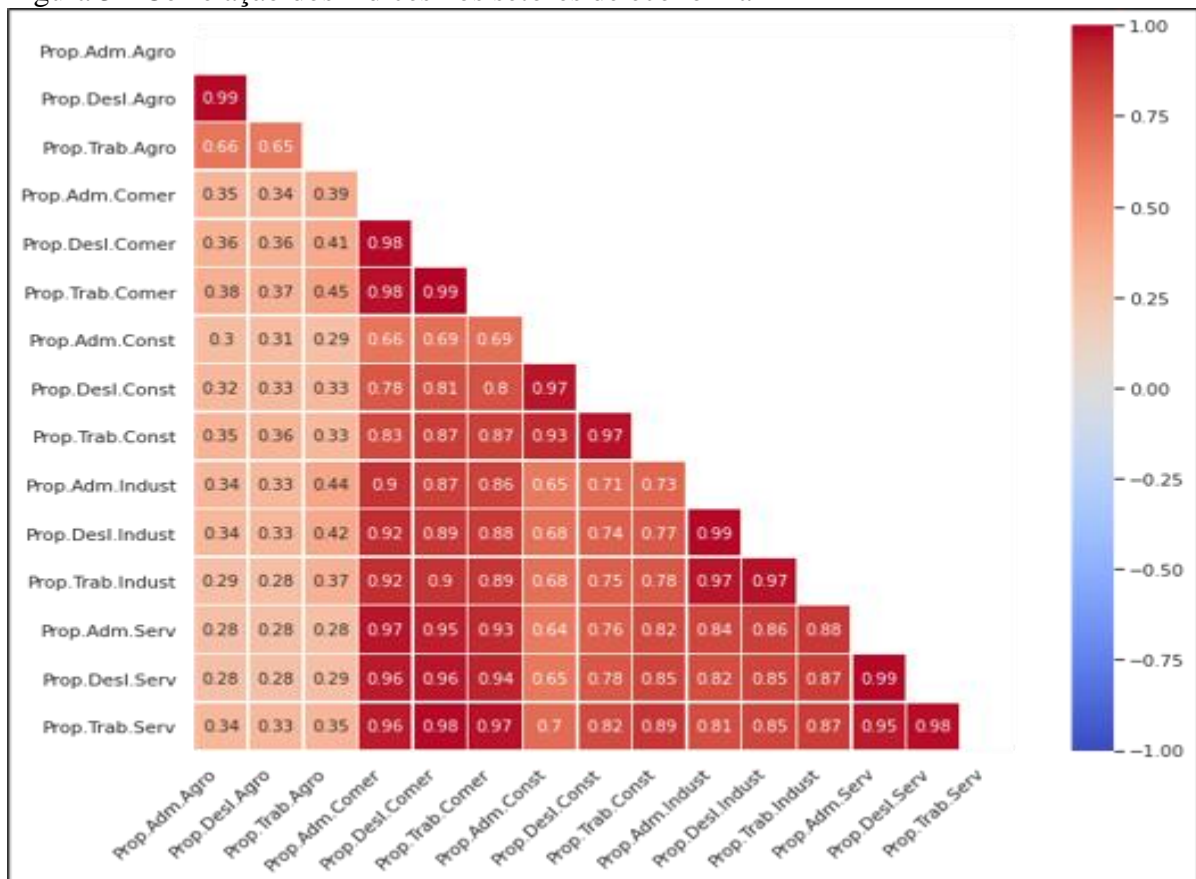
A partir das observações, pode-se notar que o setor de serviços é o setor da economia mais afetado pelos índices de emprego no Sul/Sudoeste de Minas Gerais. Isto pode ser visto em decorrência da elevada média e desvio padrão nos índices (admissão, desemprego, trabalho). Assim, a maioria dos trabalhadores da Mesorregião fazem parte deste setor.

Contudo, observa-se também que o setor de industrialização apresenta grande representatividade no Sul/Sudoeste de Minas Gerais. Entretanto, diferentemente do setor de serviços, a indústria se destaca por ter um grande número de empregados admitidos do que demitidos, apesar de sua média e desvio padrão serem inferiores, o setor apresenta um aumento maior no número de empregos gerados no ano de 2020.

O setor de agropecuária foi o que apresentou menor média, demonstrando que o número de empregos neste setor é o mais baixo entre os demais. Isto pode ser explicado pelo alto avanço tecnológico nos últimos anos, decorrente da escassez de mão de obra neste setor conforme argumenta Vivaldi, Júnior e Alves (2020).

A correlação entre as variáveis se mostrou alta entre os principais índices referentes aos setores de economia. A Figura 3 demonstra essa correlação, com os valores do coeficiente de correlação de Pearson.

Figura 3 - Correlação dos índices nos setores de economia



Fonte: Elaboração própria, a partir dados do CAGED (2021).

Analisando o gráfico da Figura 4, pode-se observar que as medidas referentes ao setor agropecuário não apresentam alta correlação com os demais setores, indicando que essas medidas não parecem estar associadas com as dos outros. Contudo, ao se analisar as demais variáveis, pode-se observar que existe uma alta correlação entre os outros setores. Isso indica uma associação positiva entre eles, ou seja, à medida que o número de empregados, desempregados e trabalhadores de um setor cresce, o número dos outros (exceto o agropecuário) tende a crescer.

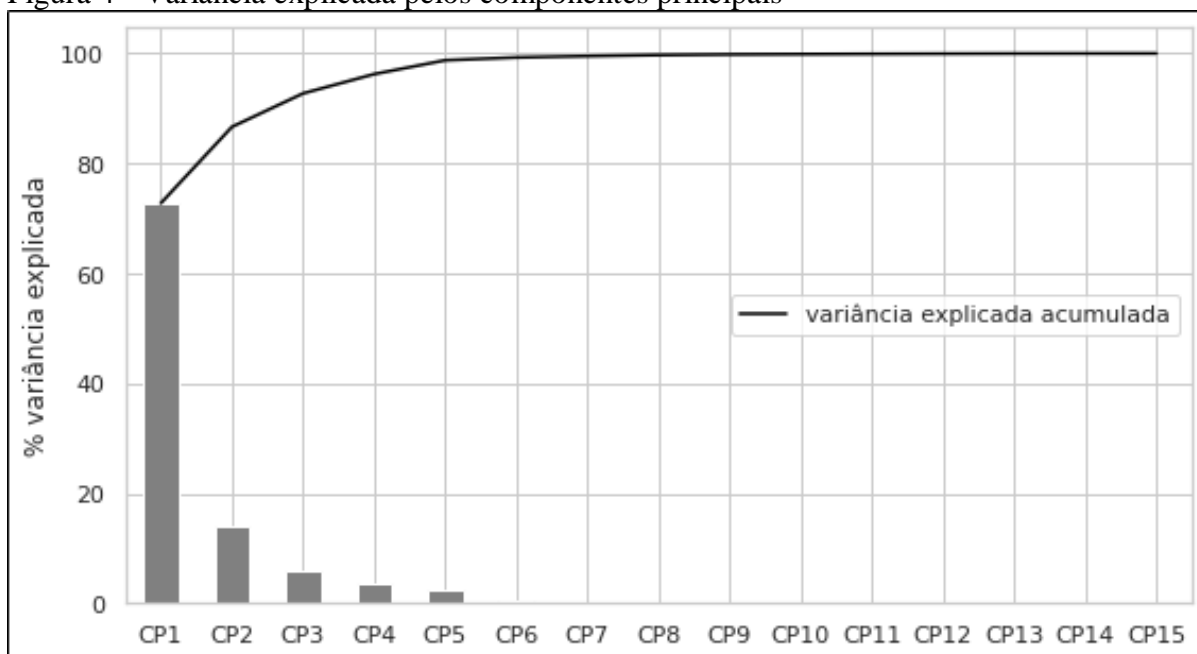
O setor da construção apresentou uma correlação um pouco mais baixa que os demais setores, principalmente no que diz respeito ao índice de geração de emprego, que apresenta uma correlação menor, mas ainda alta (maior do que 0,6) para provocar um aumento nos outros setores a cada unidade inserida no setor de construção.

#### 4.2. Análise de componentes principais (ACP)

O modelo de variância de um componente principal explica o quanto cada componente explica os dados utilizados. Desta maneira, a Figura 4 apresenta o gráfico scree plot, que tem

como finalidade demonstrar a representação de cada componente. O CP1 explica 72,75 % da variação total, o que já seria suficiente para representação dos dados. Contudo, optou-se por analisar dois componentes a fim de trazer uma análise mais detalhada dos dados. Então, foram utilizados os dois primeiros CPs que, juntos, explicam 86,63 % da variação dos dados originais.

Figura 4 - Variância explicada pelos componentes principais



Fonte: Elaboração própria, a partir dados CAGED (2020).

A Tabela 3 apresenta os coeficientes dos dois primeiros componentes principais (CPs), que explicam de forma resumida o conjunto de dados, informando a contribuição de cada variável para cada CP.

Tabela 3 - Escores dos componentes principais

	<b>Agropecuária</b>	<b>Comércio</b>	<b>Construção</b>	<b>Indústria</b>	<b>Serviços</b>
<b>Admitidos</b>					
CP1	0,13	0,29	0,24	0,28	0,28
CP2	0,6	-0,07	-0,03	-0,04	-0,13
<b>Demitidos</b>					
CP1	0,13	0,3	0,26	0,28	0,29
CP2	0,6	-0,06	-0,048	-0,05	-0,13
<b>Total de Trabalhadores</b>					
CP1	0,14	0,29	0,28	0,28	0,29
CP2	0,47	-0,03	-0,05	-0,09	-0,08

Fonte: Elaboração própria, a partir dados CAGED (2021).

Analisando o CP1, pode-se observar que existem coeficientes positivos para todas as variáveis, indicando que este componente é afetado positivamente por todos os setores analisados. Porém, pode-se notar que este componente possui um menor valor no coeficiente nos índices do setor agropecuário, indicando que os municípios que têm mais relação com este componente tendem a ter menores valores nos índices do setor agropecuário do que os demais setores. Deste modo, é possível definir o CP1 como um componente urbano, pois os maiores coeficientes são aqueles relacionados aos setores de comércio, construção, indústria e serviços.

Já o CP2 apresenta altos coeficientes nos índices do setor agropecuário. Os outros setores apresentam índices negativos e menores coeficientes, revelando que os municípios que têm uma relação maior deste componente tendem a ter menores índices dos outros setores, quando comparado ao setor agropecuário. Então o CP2 pode ser considerado como um componente rural, pois a maior contribuição é das variáveis ligadas ao setor agropecuário.

Conseqüentemente, a Figura 6 apresenta o diagrama de dispersão dos escores do CP1 e do CP2 dos municípios do Sul/Sudoeste de Minas Gerais. Desta forma ao analisar o gráfico de dispersão dos municípios nota-se que os municípios como Pouso Alegre, Poços de Caldas e Varginha que se apresentam mais à direita do gráfico, são cidades que podem ser denominadas como mais urbanas, logo que apresentam maior relação com CP1.

Já os municípios como Monte Belo, Carmo da Cachoeira e Conceição Rio Verde que se apresentaram mais à esquerda do gráfico, são cidades que ser denominadas como mais rurais, logo que apresentaram maior relação com CP2. Enfim, municípios como de Alfenas, Três Pontas e São Sebastião do Paraíso que se apresentaram centralizadas no gráfico, são aquelas cidades que podem ser chamadas tanto mais rurais quanto mais urbanas, logo que apresentaram alta relação com os dois componentes.

Figura 5 - Diagrama de dispersão dos municípios com base nos componentes principais



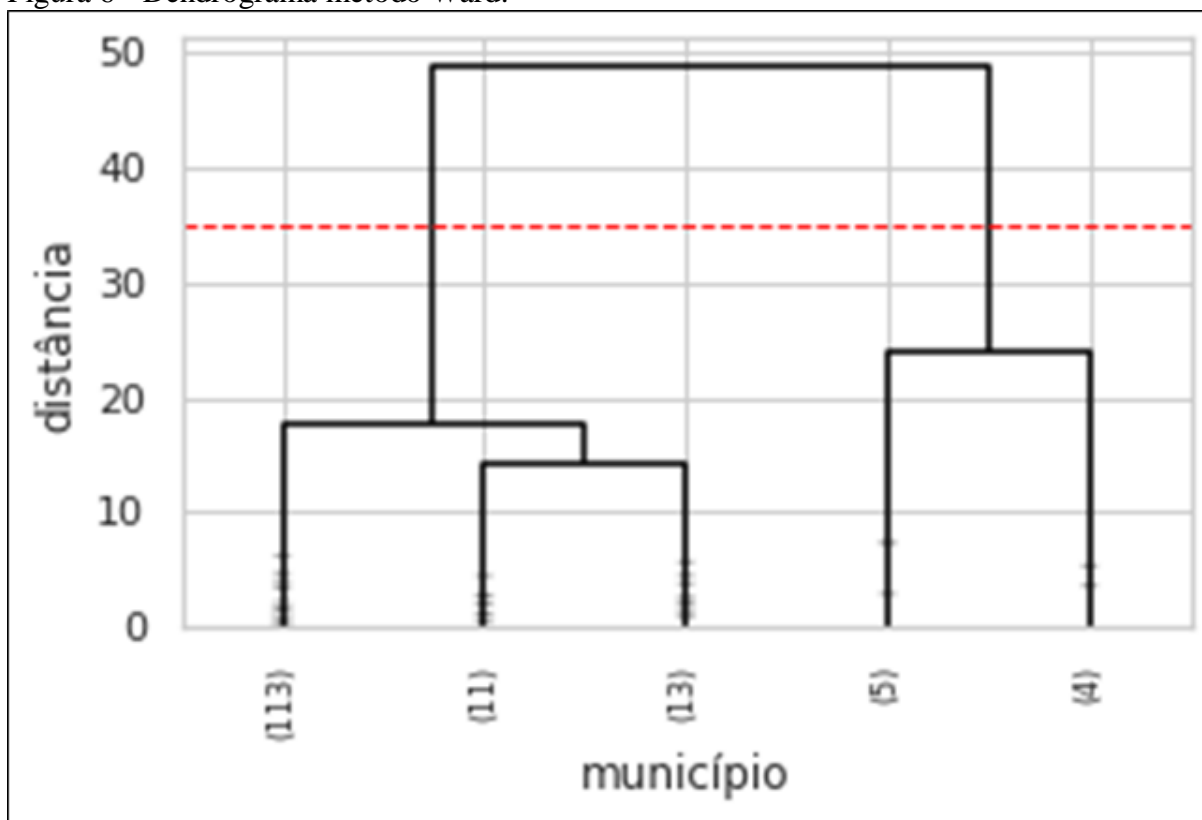
Fonte: Elaboração própria, a partir de dados do CAGED (2020).

### 4.3. Análise de agrupamento (AA)

#### 4.3.1. Método Ward

Para a análise de agrupamento hierárquico, utilizou-se do método Ward por ter sido o que mais distribui os grupos dos municípios sem que houvesse a presença de outliers. De tal modo, a figura 6 demonstra os resultados referente especificamente para o dendrograma do método de Ward, em que se teve a distribuição dos municípios em 2 grupos. O *primeiro grupo* representa cerca de 93,84 % dos municípios e o *segundo grupo* cerca 6,16 % dos municípios.

Figura 6 - Dendrograma método Ward.



Fonte: Elaboração própria, a partir dados do CAGED (2020).

Assim, a partir do número de grupos encontrados, utilizou-se do método não hierárquico com a finalidade de observar melhor, como se encontram particionados os municípios em cada grupo que será demonstrado na próxima subseção.

#### 4.3.2. Método de K-médias

Para análise de agrupamento não hierárquico, utilizou-se o método k-médias. Deste modo, optou-se por utilizar 2 grupos para a sua elaboração com base nos resultados encontrados pelo método Ward. O primeiro grupo representa 95,21% dos municípios e o segundo grupo 4,79% dos municípios.

No que diz respeito às estatísticas descritivas dos grupos obtidos, a Tabela 4 apresenta a média e a mediana para estes grupos. Analisando a Tabela 4 pode-se observar que os municípios presentes no grupo 0 apresentam média e mediana menores do que o grupo 1. Assim, aproximadamente 95% dos municípios da Mesorregião do Sul/Sudoeste de Minas Gerais apresentaram um baixo nível de empregabilidade, desemprego e número de trabalhadores acumulados.



Já no grupo 1 observa-se que estão localizados os municípios em que os índices de emprego (admitidos, demitidos e trabalhadores) são elevados para todos os setores da economia. De tal forma que este grupo apresenta aproximadamente 5 % dos municípios que se tem uma alta média e mediana, podendo ser considerados como os municípios que fornecem mais mão de obra na Mesorregião.

Tabela 4 - Média e mediana dos grupos pelo método k-médias.

		Agropecuária	Comércio	Construção	Indústria	Serviços
<b>Admitidos</b>						
<b>Grupo 0</b>	Média	44	196.17	57.95	234.4	167.53
	Mediana	8	80	10	89	60
<b>Grupo 1</b>	Média	235.43	3552.71	1325.57	2654.0	5310.86
	Mediana	189	4393	1176	2281	7071
<b>Demitidos</b>						
<b>Grupo 0</b>	Média	42.10	190.41	52.81	223.06	172.46
	Mediana	8	81	11	96	55
<b>Grupo 1</b>	Média	221.29	3327.14	1170.71	2634.00	5333.71
	Mediana	180	2885	1214	2371	5841
<b>Total de Trabalhadores</b>						
<b>Grupo 0</b>	Média	356.34	532.91	60.87	615.79	600.09
	Mediana	191	226	17	207	243
<b>Grupo 1</b>	Média	928.71	7528.43	1633.86	7596.71	12993.00
	Mediana	895	7196	1952	7041	8928

Fonte: Elaboração própria, a partir de dados do CAGED (2020).

A Figura 7 demonstra como se comportam os municípios dentro de cada grupo no gráfico com os eixos representados pelos componentes principais. Pode-se observar que os municípios tendem a ficar dispersos em relação aos grupos. Observa-se que as cidades localizadas no grupo 0 tiveram maior relação CP2, conseqüentemente os municípios como de Monte Belo, Carmo da Cachoeira e Conceição do Rio Verde apresentaram alta empregabilidade no setor agropecuário. Já Três Pontas e São Sebastião foram municípios que apresentaram alta relação tanto CP1 e CP2, indicando que estas cidades tem alta geração de emprego em todos os setores.

Os municípios contidos no grupo 1, conseqüentemente, tiveram uma maior relação com o CP1. Deste modo, as cidades de Pouso Alegre, Poços de Caldas e Varginha possuem alta relação com o CP1, ou seja, estes municípios têm tendem a ter uma alta empregabilidade nos setores de comércio, construção, industrialização e serviços. Conseqüentemente, a cidade de



## 5. CONSIDERAÇÕES FINAIS

O presente trabalho teve como objetivo analisar o como se comportam os índices de empregabilidade (admitidos, demitidos e número de trabalhadores) para os municípios da Mesorregião do Sul/Sudoeste de Minas Gerais. Para tanto, utilizou-se de técnicas multivariadas com a finalidade de analisar os dados computados do CAGED 2020. De tal modo, utilizaram-se duas técnicas: Análise de Componentes Principais e Análise de Agrupamento (método Ward e de K-Médias). O Quadro 2 ilustra de forma resumida os procedimentos adotados para utilização dos métodos multivariados.

Quadro 2 – Resumo dos procedimentos multivariados.

<b>Método</b>	<b>Objetivo</b>	<b>Nº Componentes/Grupos</b>	<b>Resultados Gerais</b>
Componentes Principais	Encontrar número de componentes que resume as variáveis estudadas.	2	Observou-se que 2 componentes são ideais para resumir os índices de empregabilidades nos setores da economia.
Ward	Encontrar número de grupos que resume as observações estudadas.	2	Observou-se que 2 grupos são ideais para resumir os municípios da Mesorregião do Sul/Sudoeste de Minas Gerais.
K médias	A partir do número grupos escolhidos, particionar número de observações em cada grupo no qual se tenha a menor distância.	2	Observou-se que após particionado os municípios em dois grupos: o primeiro grupo ficou com 139 municípios e o grupo 1 com 7 municípios.

Fonte: Elaboração própria.

A partir das análises empregadas pode-se notar que existe uma distinção entre os índices de empregabilidade empregados. De tal modo, que se pode observar que os índices do setor agropecuário tendem a se diferenciar dos demais setores da economia. Isto, pode ser observado pela questão da modernização maquinaria que vem captando cada vez mais empregos deste setor.

Através da análise de componentes principais pode-se observar que as variáveis estudadas ficaram melhor particionadas em dois componentes principais. O CP1 denominado um componente urbano, teve maior relação com os setores de comércio, construção, indústria e serviços. Já o CP2 denominado um componente rural, teve maior relação com o setor agropecuário.

A análise de agrupamento demonstrou que as cidades de Varginha, Poços de Caldas e Pouso Alegre e Extrema que foram os municípios que geraram mais empregabilidade para

maioria dos setores da economia. Da mesma forma, que em ambos casos, municípios como Monte Belo, Carmo da Cachoeira e Conceição do Rio Verde se demonstraram pertinentes na geração de emprego no setor agropecuário.

Do mesmo modo, a cidade de Alfenas, que mostrou alta relação com os dois componentes principais, assim como Três Pontas e São Sebastião do Paraíso, esses municípios possuem alta geração emprego em todos os setores da economia.

O estudo dos índices de empregabilidade nos setores da economia a partir de métodos multivariados possibilita que se tenha uma visão de como se comportam a geração de emprego perante os municípios, já que através destas técnicas é possível verificar como cada observação se relaciona com as variáveis utilizadas. Desta forma, para análises futuras sugere-se de utilizar outras variáveis que geram economia para cidades a fim de explicar a geração de emprego e como ela distribuída.

## REFERÊNCIAS BIBLIOGRÁFICAS

ANDRADE, R. G. R. **A expansão da cafeicultura em Minas Gerais: da intervenção do Estado à liberação do mercado**. 1994. 173 f. Dissertação (Mestrado em Economia) - Universidade Federal de Minas Gerais, Belo Horizonte, 1994.

CAGED. **PDET**, 2020. Disponível em: <http://pdet.mte.gov.br/novo-caged>. Acesso em: 2 de setembro de 2021.

CADASTRO GERAL DE EMPREGADOS E DESEMPREGADOS (CAGED). Painel de Informações do Novo Caged. **CAGED**, 2020. Disponível em: <https://app.powerbi.com/view?r=eyJrIjoiNWl5NWl0ODEtYmZiYy00Mjg3LTkzNWUtY2UyYjIwMDE1YWI2IiwidCI6IjNlYzkyOTY5LTVhNTEtNGYxOC04YWM5LWVmOThmYmFmYTtk3OCJ9&pageName=ReportSectionb52b07ec3b5f3ac6c749>. Acesso em: 28 de agosto de 2021.

CRESWELL, John W.; CLARK, Vicki L. Plano. **Pesquisa de Métodos Mistos-: Série Métodos de Pesquisa**. Penso Editora, 2015.

CUNHA, A.; MODOTTI, M.; SETTI, J. Classificação de caminhões através de agrupamento por análise de cluster. In: **ANPET. Anais do XXII Congresso de Pesquisa e Ensino em Transportes da ANPET. Fortaleza/CE**. 2008. p. 1447-1459.

FERREIRA, Daniel Furtado. **Estatística multivariada**. Lavras: Editora Ufla, 2008.

FLORIANO, I. **As principais culturas agrícolas de Minas Gerais: uma análise multivariada**. Trabalho de Conclusão de Piepex (Graduação no Bacharelado Interdisciplinar em Ciência e Economia), Universidade Federal de Alfenas, Varginha, 2019.

GRIMM, Laurence G.; YARNOLD, Paul R. **Reading and understanding MORE multivariate statistics**. American psychological association, 2000.

HAIR, Joseph F. et al. **Análise multivariada de dados**. Bookman editora, 2009.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). Divisão Regional do Brasil. **IBGE**, 2021. Disponível em: <https://www.ibge.gov.br/geociencias/organizacao-do-territorio/divisao-regional/15778-divisoes-regionais-do-brasil.html?=&t=o-que-e>. Acesso em: 2 de setembro de 2021.

KAISER, Henry F. The varimax criterion for analytic rotation in factor analysis. **Psychometrika**, 1958, 23.3: 187-200.

KUBRUSLY, Lucia Silva; SABOIA, João Luiz Maurity. Uma análise multivariada da população ocupada nas regiões metropolitanas brasileiras. **Ensaio FEE**, v. 27, n. 2, 2006.

MOURA, F.; VALE, A. Mercado de trabalho nos pequenos municípios e polarização regional: uma análise sobre Divisa Nova e Alfenas, no Sul/Sudoeste de Minas. **Espaço em Revista**, v. 20, n. 2, p. 1-18, 2018.

JOLLIFFE, Ian T. Discarding variables in a principal component analysis. I: Artificial data. **Journal of the Royal Statistical Society: Series C (Applied Statistics)**, v. 21, n. 2, p. 160-173, 1972.

PRATES, Ângelo Marcos Queiróz et al. A Dinâmica Da Estrutura Produtiva Da Região Sul/Sudoeste De Minas Gerais Nos Anos Recentes: Apontamentos De Sua Inserção Na Economia Mineira. In: **Anais do XIV Seminário sobre a Economia Mineira [Proceedings of the 14th Seminar on the Economy of Minas Gerais]**. Cedeplar, Universidade Federal de Minas Gerais, 2010.

PYTHON. **The Python programming language**. Disponível em: <https://docs.python.org/3/reference>. Acesso em 28 de agosto de 2021.

VICINI, Lorena. **Análise multivariada: da teoria à prática**. 2005.