



Atributos e Oportunidades do Sistema Mineiro de Inovação: um estudo a partir de análise multivariada

Autores:

Maria Isabel Batista D G Baptista - CEDEPLAR/UFMG - mibdgb@cedeplar.ufmg.br
João Pedro Figueira Amorim Parga - CEDEPLAR/UFMG - parga.joao@gmail.com

Resumo:

Esse artigo busca formular grupos (clusters) de microrregiões mineiras de forma a identificar capacidades inovativas locais. A relevância desse tema é eminente na literatura schumpeteriana e de geografia da inovação, que considera avanços tecnológicos o motor do desenvolvimento socioeconômico. A partir dos métodos de Análise de Componentes Principais (ACP) e Análise de Cluster, é proposta uma representação do arranjo das microrregiões de Minas Gerais de modo a esclarecer o estudo de seu emergente Sistema Regional de Inovação. Foram utilizados dados do ano 2010 para construir indicadores que captam as dimensões científica, tecnológica, urbana e social. Na conclusão do trabalho são identificadas regiões promissoras para alavancarem o desenvolvimento tecnológico no estado.



ATRIBUTOS E OPORTUNIDADES DO SISTEMA MINEIRO DE INOVAÇÃO

Um estudo a partir de análise multivariada

1 INTRODUÇÃO

Foi Schumpeter quem colocou o progresso tecnológico como motor do desenvolvimento econômico. Em sua visão, esse progresso aumenta a produtividade e desloca a economia para um patamar com maior renda e qualidade de vida.

O conjunto de tecnologias que caracterizam a atual Era do Conhecimento possibilitou a troca de informação rápida, simples entre quaisquer partes do planeta. Isso não diminui, entretanto, a importância de características locais para a criação de novas tecnologias. Pelo contrário, muito dos insumos necessários à atividade inovativa dependem de proximidade geográfica. O aprendizado através da interação direta entre agentes altamente qualificados e especializados, por exemplo, acontece apenas a nível local. É atual e emergente o estudo da relação entre o espaço local e o desenvolvimento de tecnologias no mundo globalizado.

O Sistema Regional de Inovação é o conjunto formado pelas estruturas científica, tecnológica e institucional que cria um ambiente favorável ao desenvolvimento tecnológico. Universidades, Empresas e Governo possuem seu papel nesse processo.

A difusão desse conceito coincide com ao surgimento do novo paradigma das tecnologias de informação. A instituição do paradigma da economia do conhecimento impôs uma nova dinâmica de concorrência na economia global após a década de 1980, caracterizada por uma crescente valorização das ideias em meio ao processo produtivo. Investimentos em qualificação da mão-de-obra e expansão de centros de pesquisa e desenvolvimento (P&D) passam a fazer parte da rotina das grandes empresas que buscam a inovação.

Minas Gerais é um estado com atividade produtiva lideradas por setores de baixa tecnologia (Indústrias Extrativas e Indústria de Transformação), na sua maioria privada e com capital estrangeiro. A rede de ensino superior é ampla, sendo contemplada por grandes universidades. As desigualdades regionais são marcantes, com a capital Belo Horizonte concentrando grande parte do dinamismo econômico.

Esse trabalho busca estudar o Sistema Regional de Inovação mineiro, procurando identificar possibilidades de investimento e políticas públicas para inovação. O Estado resume muitas das características típicas de outras regiões brasileiras, como a desigualdade regional. A superação dessas desigualdades é o caminho para o desenvolvimento econômico e para o bem-estar da população.



Esse artigo é composto por cinco partes, além da Introdução e da Conclusão. A primeira seção apresenta a revisão teórica sobre a relação entre inovação tecnológica e desenvolvimento econômico. A segunda narra a história da construção do Sistema Regional de Inovação em Minas Gerais. A terceira seção descreve as bases de dados utilizadas e as variáveis a partir delas construídas, bem como a justificativa de porque são indicadores de dimensões relacionadas à capacidade inovativa de uma microrregião. Em seguida, os métodos são revistos e descritos na seção de Metodologia. Por fim, os resultados do estudo são apresentados.

2 REVISÃO TEÓRICA: INOVAÇÃO E DESENVOLVIMENTO REGIONAL

A Teoria do Desenvolvimento Econômico (1911) de Schumpeter foi original ao defender o papel crucial das inovações no aumento do nível de renda de um país. Em *Capitalismo, Socialismo e Democracia* (1942), escreveu que a destruição criativa originada de um processo de inovação se configura como “o fato essencial do capitalismo”. A destruição criativa aumenta a taxa de crescimento econômico de longo prazo, renovando o sistema. Por isso, um país que busca o desenvolvimento econômico deve assimilar, transformar e criar novas tecnologias, garantindo sua competitividade nas trocas globais.

Para além do argumento do processo inovativo, Schumpeter itera fundamentos de forte corrente de pensamento econômico, ao defender que esse processo é possibilitado após acumulação de capital feita por agentes sociais em conflito. Aqui, importa o processo histórico que conduz a evolução institucional de uma sociedade. A capacidade inovativa de um país ou região é vista, assim, como resultado das relações entre agentes econômicos, políticos e sociais e reflete condições culturais e institucionais próprias.

Na década de 1970, trabalhos de Chris Freeman, Richard Nelson e Nathan Rosenberg refinaram o conceito de inovação, que não deveria ser vista como a descoberta de novos princípios técnicos e científicos. O processo inovativo é, explicam os autores, um aprendizado gerado por interações entre agentes nos níveis local, nacional e mundial. Agentes como indivíduos, firmas e outras organizações voltadas à busca de novos conhecimentos. Para o processo inovativo, foi reconhecida a importância de redes formais e informais de contato e cooperação, de comunicação com fontes externas de informação científica e tecnológica e da integração entre pesquisa e desenvolvimento (P&D), produção, design e marketing (SZAPIRO et al, 2017).

2.1 DETERMINANTES LOCAIS DO PROCESSO INOVATIVO

Externalidade é todo produto de atividades econômicas que afeta outros agentes não diretamente envolvidos na atividade. Esse é um conceito que surge naturalmente quando se trata de inovação, atividade que gera impactos muito difusos. é um caso de externalidade



positiva porque beneficia toda a sociedade, não apenas aqueles agente diretamente envolvidos com a sua elaboração.

Dois tipos de externalidades positivas locais são reconhecidos por exercerem papel importante no processo inovativo. Por um lado, há externalidades de especialização, que operam quando a estrutura produtiva de uma região é caracterizada pela forte especialização em determinado setor. De outro lado, as externalidades que emergem da diversificação, verificadas em estruturas produtivas complexas e diversificadas.

O principal autor que apontou a importância das economias de especialização foi Marshall (1920). A partir de seu trabalho, Arrow (1962) e Romer (1986) formalizaram o modelo que ficou conhecido com MAR (Marshall-Arrow-Romer). Nesse modelo, a especialização tem o importante efeito de estimular a circulação de informações, de ideias e o compartilhamento de conhecimento por meio de processos de imitação e interação entre os agentes. As externalidades de especialização ocorrem somente entre empresas de um mesmo setor.

Por outro lado, Jacobs (1969) apontou que a diversidade da estrutura produtiva de uma região ou de uma cidade é o principal motor da inovação. As estruturas produtivas diversificadas conduzem o intercâmbio de habilidades necessário para a emergência de novos setores e de novas atividades produtivas, uma vez que essas habilidades podem ser recombinadas para a produção de novas mercadorias e o atendimento a novos mercados. Além disso, a existência de uma adequada infraestrutura de transportes e comunicação, a proximidade aos mercados e o acesso privilegiado a serviços especializados são fontes adicionais de externalidades de diversificação, bastantes associadas à urbanização e muitas vezes chamadas de externalidades jacobianas.

Autores neoschumpeterianos iniciaram uma ampla agenda de pesquisa procurando avançar no entendimento sobre o processo de inovação e seus principais determinantes. Neste contexto, Nelson e Winter (1977) desenvolveram as bases daquela que, mais tarde, ficou conhecida como teoria evolucionista da inovação, e apresentaram seus elementos centrais de forma um pouco mais sistemática em seu livro clássico de 1982.

No final da década de 1980, foram associados elementos da teoria de economia regional à teoria de desenvolvimento de Schumpeter culminando na síntese de Sistemas de Inovação. Em 1988, foi publicada a clássica coletânea *Technical Change and Economic Theory* (DOSI et al, 1988) que introduziu a ideia de sistemas de inovação na literatura acadêmica.

A abordagem de sistema de inovação se torna um instrumento analítico extremamente importante porque reforça o caráter localizado da geração, assimilação e difusão da inovação, que não é anulado pelo aprofundamento da integração entre as economias nacionais. Não obstante o aumento da importância das relações globais de produção, a tecnologia não pode ser gerada em qualquer local e facilmente difundida para diferentes partes do mundo.

Na abordagem neoschumpeteriana, o processo de inovação é localizado, fundamentalmente social e enraizado culturalmente. Nesse aspecto, destaca-se a

importância da influência das instituições que oferecem incentivos ou restrições para a inovação tais como leis, regulações, normas culturais, regras sociais e padrões técnicos. Nessa percepção, os processos de inovação são marcados por mecanismos de feedback e relações de interação envolvendo ciência, tecnologia, produção, políticas públicas e demanda dos usuários (SZAPIRO et al, 2017).

A relação entre proximidade geográfica e as possibilidades de compartilhamento de conhecimentos entre agentes parte do pressuposto de que o conhecimento possui dois componentes distintos. O primeiro é o conhecimento decodificado, que pode ser transferido entre os agentes por meio de manuais. O segundo é o conhecimento tácito, cujo compartilhamento envolve, necessariamente, interação entre os agentes, o que dificulta a sua transferência a longas distâncias.

O conhecimento tácito está enraizado nos indivíduos e nas rotinas das empresas. Nesse sentido, ele pode ser considerado o principal fator de concentração geográfica de atividades inovativas, porque seu papel central no processo de aprendizado interativo reforça os efeitos positivos da localização, da proximidade geográfica e da concentração espacial dos agentes econômicos (GARCIA, 2017).

2.2 SISTEMAS REGIONAIS DE INOVAÇÃO

Os avanços na tecnologia da informação criaram novos padrões de competitividade frente a um modelo globalizado de acumulação de capitais. As capacidades de geração, absorção e difusão de conhecimento passaram a figurar dentre os principais diferenciais para a competitividade no plano internacional. Isso fez com que as vantagens competitivas baseadas no conhecimento e na inovação tecnológica prevaleçam sobre as vantagens baseadas em preços e escala produtiva (SANTOS e CALIARI, 2012).

Os primeiros trabalhos acadêmicos e os relatórios de políticas utilizando a abordagem de Sistemas de Inovação apresentaram um recorte nacional. A decisão por essa dimensão de análise se explica pela trajetória histórica compartilhada por todos os agentes em uma nação, tais como: arcabouço institucional; sistemas educacionais e de ciência e tecnologia; bases socioculturais, estruturas de capital; quadro político e políticas industrial e de CT&I nacionais (SZAPIRO et al, 2017).

Não obstante, é no território compreendido como espaço cognitivo em que ocorrem os processos interativos e cooperativos que levam à inovação e onde se deve centrar a abordagem.

O sucesso de experiências regionais como a do Vale do Silício, nos Estados Unidos, e a dos distritos industriais, da chamada terceira Itália, reacenderam o debate sobre a importância da localidade e de suas peculiaridades para o desenvolvimento produtivo. Vieram à tona as vantagens da aglomeração industrial, com a presença de trabalho qualificado, a interação via cooperação e competição, a imersão local e externalidades, e seus reflexos sobre a produção de inovações (ASHEIM, 1995).

O conceito de sistemas regionais de inovação (SRIs) buscou superar uma limitação da teoria do SNIs, a de não levar em conta as peculiaridades regionais. A existência de desequilíbrios regionais é algo persistente principalmente em países subdesenvolvidos e de grandes dimensões como o Brasil. Dessa forma, a opção pela análise do SRI de Minas Gerais neste trabalho se mostra adequada.

O SRI é composto por relações de cooperação entre agentes da infraestrutura produtiva local que levam a geração e internalização de novos conhecimentos técnico-científicos. Os principais agentes são as empresas inovadoras, instituições criadoras de conhecimento como universidades e centros de pesquisa, instituições voltadas à transferência de tecnologia, instituições de suporte a inovação, e o sistema financeiro.

Lundvall, após analisar os determinantes da inovação, concluiu que o recurso mais estratégico à dinâmica da inovação é o aprendizado que surge da interação entre as pessoas imersas em um ambiente social particular.

O autor considera o aprendizado por P&D interna (*learning by searching*), pela interação com centros de pesquisas e universidades (*learning from advanced S&T*), o *learning by doing*, relacionado a melhorias no ambiente de produção, o *learning by using*, relacionado a ganhos pelo aprendizado com sistemas mais complexos, e o *learning by interacting*, relacionado a inovações de produtos surgidas da interação de produtos e usuários (BITTENCOURT e CÁRIO, 2017).

Deve-se destacar a existência de uma relação reflexiva entre conhecimento tácito e o contexto, uma vez que o conhecimento tácito, ao mesmo tempo, define e é definido pelo contexto social. A aprendizagem interativa e de longo prazo, capaz de transferir conhecimentos tácitos encontraria menos barreiras onde estão presentes afinidades culturais e linguísticas, por exemplo.

A partir da síntese neoschumpeteriana que sistematizou uma teoria sobre os determinantes regionais da inovação e sua relação com o desenvolvimento econômico, uma análise empírica é feita neste trabalho para se avaliar o sistema regional de inovação de Minas Gerais.

Foram utilizados dados que fornecem informações sobre os componentes da estrutura de fomento à inovação tais como qualificação da mão-de-obra, centros de P&D, informações sobre a estrutura do mercado e padrões de especialização e de diversificação.

3 O SISTEMA REGIONAL DE INOVAÇÃO EM MINAS GERAIS

O Estado de Minas Gerais é uma das unidades da federação brasileira mais marcadas por desigualdades regionais, com elevada concentração de atividades produtivas em poucas regiões, implicando também uma distribuição desigual tanto de renda como de receitas públicas.



O perfil da economia de Minas Gerais é de especialização produtiva com algumas aglomerações industriais. A indústria extrativa mineral é naturalmente concentrada em certas regiões porque não se pode alterar a localização das minas. Nesse caso, há notadamente concentração na região central do estado, onde está o Quadrilátero Ferrífero.

O complexo mineiro-metalúrgico se tornou a principal aglomeração industrial da economia mineira. No final da década de 1970, congregava um número significativo de grandes empresas nacionais e internacionais na área de mineração, siderurgia e refratária como Cia Vale do Rio Doce, Minerações Brasileiras - MBR, Usinas Siderúrgicas de Minas Gerais - Usiminas, Cia Siderúrgica Belgo-Mineira, Aços Especiais Itabira - Acesita, Aços de Minas Gerais S. A. - Açominas, Mannesman e Magnesita.

Entre 1970 e 1990, houve uma significativa mudança da estrutura industrial de Minas Gerais em direção a uma maior diversificação produtiva, que de uma especialização inicial em indústrias consideradas base, como mineração, metalurgia, minerais não-metálicos e de energia elétrica, passou por uma diversificação horizontal para as indústrias automobilística, mecânica, eletroeletrônica e serviços de telecomunicação (LEMOS E DINIZ, 1998).

Durante os programas de substituição de importações, cuja doutrina incluía a liderança do Estado na promoção do desenvolvimento, foram desenvolvidas instituições relacionadas ao planejamento regional e desenvolvimento tecnológico. Na década de setenta, Minas Gerais vivia uma experiência histórica de pesquisa, com o Centro Tecnológico de Minas Gerais (CETEC), a Empresa de Pesquisa Agropecuária (EPAMIG), a Fundação João Pinheiro (FJP), a Fundação Ezequiel Dias (FUNED) e o Centro de Pesquisa René Rachou.

Essas instituições foram duramente atingidas pelos cortes de despesas do Governo Estadual com a crise financeira brasileira nos anos 1980, que reorientou a estratégia econômica do governo para o pagamento da dívida. O contexto de crise fiscal crônica, aliado a mudanças no cenário internacional, resultou em uma forte reação contra o modelo de substituição de importações até então em vigor e induziu uma guinada neoliberal a partir dos anos 1990. Neste ano, o então presidente do Brasil, Fernando Collor, lançou o Programa Nacional de Desestatização (PND). O PND visava reduzir a máquina estatal, seguindo a receita neoliberal de ajuste macroeconômico para países subdesenvolvidos difundidas após o Consenso de Washington (1989).

O Brasil, como a maioria dos países da América Latina, passou por um processo de rápida redução do tamanho do Estado, entrada de capitais transnacionais, abertura comercial e privatização de empresas públicas. As privatizações foram parte importante da agenda econômica do governo Collor e foram mantidas pelos governos seguintes. Nesse contexto, foram privatizadas USIMINAS e Acesita, siderúrgicas mineiras localizadas nos municípios de Ipatinga e Timóteo, respectivamente. Com o impeachment de Fernando Collor, o seu vice-presidente Itamar Franco assume e dá continuidade às privatizações. Em seu governo, as privatizações mais emblemáticas foram as da Companhia Siderúrgica Nacional (CSN), da Açominas e da Embraer.

O sucessor de Itamar Franco, Fernando Henrique Cardoso, presidente do Brasil de 1995 a 2002, também adotou as recomendações neoliberais. Foi durante seu mandato que,

em maio de 1997, foi privatizada a então, e atual, maior exportadora de minério de ferro do mundo, a Companhia Vale do Rio Doce (CVRD), após ser incluída no PND em 1995.

Tem-se, então, um panorama do sistema regional de inovação de Minas Gerais. Em resumo, é caracterizado pelo protagonismo do setor extrativo mineral na economia, hoje gerenciado pela iniciativa privada. Na produção científica, instituições públicas de pesquisa são destaque.

O estado mineiro, pelo exposto, aparece como imagem da realidade brasileira de um país regionalmente desigual e com atividade econômica predominante de indústria de baixa tecnologia. Caracterizar o sistema regional de inovação mineiro esclarece, portanto, questões sobre o desenvolvimento econômico em economias periféricas de modernização tardia e incompleta.

4 FONTE DE DADOS

No presente trabalho, serão utilizadas as seguintes fontes de dados: o Censo Demográfico realizado no ano de 2010, Base de Dados de Comércio Exterior Brasileiro (Comex Stat), Relação Anual de Informações Sociais (RAIS) de 2010, Censo da Educação Superior (INEP) e Produto Interno Bruto (PIB) dos Municípios (IBGE). Algumas das informações são disponibilizadas por municípios, podendo ser agregados por microrregiões. A análise deste trabalho é feita para o espaço geográfico compreendido pelas 66 microrregiões do Estado de Minas Gerais.

O Censo Demográfico 2010, como os demais já realizados, apresenta informações sobre domicílios recenseados, segundo a espécie, e população residente, segundo as Unidades da Federação e municípios. O Comex Stat é um sistema para consultas e extração de dados utilizados para construção da balança comercial brasileira, detalhados por município do exportador/importador e divulgados mensalmente. A RAIS é um sistema estatístico responsável pelo censo anual do mercado de trabalho formal. O Censo da Educação Superior, realizado anualmente pelo Inep, é o instrumento de pesquisa do Brasil sobre as instituições de educação superior (IES), de seus alunos e docentes. O IBGE divulga as estimativas do PIB dos Municípios, referência 2010. São apresentados, a preços correntes, os valores adicionados brutos dos três grandes setores de atividade econômica, bem como os impostos, o PIB e o PIB per capita.

A partir das informações originais das variáveis das bases de dados em questão, foram elaboradas as variáveis a seguir, tendo como indivíduos de análise as microrregiões do Estado de Minas Gerais. As variáveis construídas podem ser divididas em quatro grupos.

Em primeiro lugar, o grupo dos indicadores de mercado de trabalho: rendimento médio mensal da população economicamente ativa (RMM), taxa de atividade (TXA), porcentagem de trabalhadores no mercado formal com ensino superior completo (ESC). Essa

dimensão busca captar a capacitação da mão-de-obra, responsável pela divulgação de conhecimento tácito.

Em segundo lugar, indicadores de urbanização e qualidade de vida da população: porcentagem de população urbana (PPU), porcentagem da população com saneamento inadequado (SAN) e número de empregados nos setores de intermediação financeira (FIN). A utilização desta variável como medida aproximativa de escala e densidade econômica urbana se inspira no conceito de Christaller de bem superior, e já foi utilizada na literatura afim (ALBUQUERQUE ET AL, 2002). Essa dimensão busca captar o potencial de geração de externalidades positivas de diversificação.

A terceira dimensão diz respeito às estruturas de mercado da indústria local: concentração de setores de alta tecnologia (QL_Tec), concentração de grandes empresas (QL_Gr) e concentração de estabelecimentos do setor de extração mineral (QL_Ext). Na teoria neoschumpeteriana, grandes empresas são principais responsáveis por inovação. Por isso, QL_Gr captura a concentração de empresas com mais de 250 empregados, critério reconhecido para a definição de uma empresa grande. Quando estágio de desenvolvimento econômico é tal que a articulação entre universidade e indústria ainda é baixo, os determinantes para a localização da indústria são mais relacionados com as lições da economia regional clássica, como a proximidade de matéria prima. Essa é a situação que a variável (QL_Ext) busca captar. Os setores de alta tecnologia foram selecionados de acordo com a classificação da OCDE.

A quarta e última dimensão trata das estruturas de produção científica, que foram contempladas através dos seguintes indicadores: investimento das instituições de ensino superior - IES - com investimento (INV), gasto com pesquisa e desenvolvimento das IES (P&D) e concentração de estabelecimentos de P&D (QL_P&D). Essas variáveis buscam contemplar atividades de ensino, pesquisa e formação de mão-de-obra qualificada.

A tabela contida no apêndice A traz o resumo das características das variáveis usadas e seu respectivo método de construção. A seção seguinte explica os métodos estatísticos empregados nesse artigo, que foram escolhidos por possibilitar a identificação de padrões de proximidade entre os indivíduos de análise em função de suas características. Por isso, será possível esclarecer as diversas formas como quatro dimensões interagem nas microrregiões do estado de Minas Gerais.

5 METODOLOGIA

Este trabalho faz uso métodos de técnicas de Análise Multivariada, compreendida como um conjunto de procedimentos estatísticos formulados para extrair informações e analisar simultaneamente múltiplos indivíduos ou objetos (HAIR JR. et al., 2013). As ferramentas aplicadas neste estudo visam identificar padrões de proximidade entre as observações, de acordo com as quatro dimensões mencionadas na seção anterior (relativas ao mercado de trabalho, à urbanização e qualidade de vida da população, às estruturas de mercado da indústria local e aos sistemas de inovação e produção científica).



Nas subseções seguintes serão revisados, brevemente, os métodos de análise multivariada empregados neste trabalho, notadamente a Análise de Componentes Principais (ACP) e a Análise de Clusters.

5.1 Quociente de Localização (QL)

O Quociente Locacional é um indicador que compara a participação percentual de uma região j qualquer em um setor específico i com a participação percentual da mesma região j na economia de referência. Em outras palavras, o QL “procura comparar duas estruturas setoriais-espaciais” (CROCCO et al., 2006, p. 218). Matematicamente:

$$QL_{ij} = \frac{E_{ij}/E_i}{E_{.j}/E_{..}}$$

onde E_{ij} é a produção do setor i na região j , E_i é a produção total do setor i , $E_{.j}$ é a produção total na região j e $E_{..}$ é a produção total.

Este índice consiste no parâmetro de localização e especialização mais comumente utilizado em estudos regionais, em especial em técnicas de análise exploratória. Como salientado por Simões (2005), especializações produtivas locais podem ser distinguidas a partir deste indicador, por exemplo.

Entretanto, algumas considerações devem ser feitas com relação ao uso do Quociente Locacional, “visto que a interpretação de seu resultado deve levar em conta as características da economia que está sendo considerada como referência” (CROCCO et al., 2006, p. 220). Em especial, a elevada disparidade regional existente no cenário brasileiro, em diferentes níveis de desagregação, faz com que o indicador possa apresentar distorções em regiões mais extremas das amostras. Por sua própria natureza, o QL tende a sobrevalorizar o peso de certos setores para regiões relativamente pequenas, cujas variáveis analisadas (relativas ao mercado de trabalho, estrutura produtiva, etc.) sejam demasiadamente deprimidas em contraste do resto do universo. Analogamente, há a subvalorização do indicador em regiões grandes. Desta maneira, é aconselhável que pesquisadores que empreguem este índice tomem suas conclusões com certa parcimônia e cientes das possíveis deturpações que ele pode gerar.

5.2 Análise de Componentes Principais

Dentro do arcabouço de análises estatísticas multivariadas, a Análise de Componentes Principais é provavelmente uma das técnicas mais comumente utilizadas, sendo empregada nas mais diversas áreas do conhecimento científico.



Esta ferramenta é aplicada para extrair as inter-relações mais relevantes de um conjunto composto por uma quantidade relativamente alta de variáveis heterogêneas que apresentam, não obstante, um alto grau de correlação entre elas. Por meio de combinações lineares do conjunto de variáveis originais, procura-se explicar as suas estruturas de variância e covariância a partir de índices que são não-correlacionados (ortogonais) e que podem ser ordenados de acordo com suas contribuições à variabilidade dos dados. A intuição desta técnica parte do pressuposto de que, apesar de a variabilidade total de um sistema ser dada por um número p de variáveis, admite-se que ela pode ser condensada e representada, sem perdas significativas à análise, por uma quantidade reduzida de k componentes, sendo $k < p$ (SIMÕES, 2005). Busca-se, neste sentido, sintetizar as informações que serão investigadas em um conjunto composto por um número menor de índices, denominados Componentes Principais (CPs), de modo que a quantidade de informação perdida seja minimizada (HAIR JR. et al., 2013).

Algebricamente, o procedimento para a obtenção dos componentes principais pode ser resumido da seguinte maneira¹. Seja um sistema composto por p variáveis aleatórias e dado pelo vetor $X' = X_1, X_2, \dots, X_p$. Os componentes principais são obtidos pelas combinações lineares não-correlacionadas Y_1, Y_2, \dots, Y_p das variáveis pertencentes ao vetor X' ,

$$Y_1 = a'_1 X' = a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p$$

$$Y_2 = a'_2 X' = a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p$$

⋮

$$Y_p = a'_p X' = a_{p1}X_1 + a_{p2}X_2 + \dots + a_{pp}X_p$$

dada a hipótese de maximização das variâncias $Var(Y_p)$ e sujeitas às restrições.

$$a'_1 = a_{11} + a_{12} \dots + a_{1p}$$

⋮

$$a'_p = a_{p1} + a_{p2} \dots + a_{pp}$$

Com isto em vista, o primeiro componente corresponde à combinação linear Y_1 que maximiza $Var(Y_1)$ sujeito à restrição a'_1 , o segundo a Y_2 que maximiza $Var(Y_2)$, sujeito à restrição a'_2 e cuja $Cov(a'_1 X a'_2 X) = 0$, e assim por diante. Em consequência, as variâncias dos CPs podem ser ordenadas: $Var(Y_1) \geq Var(Y_2) \geq \dots \geq Var(Y_p)$, de modo que o primeiro componente, Y_1 , explica a maior parcela da variação total das variáveis X' , o segundo, Y_2 , explica a segunda maior porção, porém inferior à primeira, e assim sucessivamente.

¹ Para explicações detalhadas dos procedimentos e provas algébricas do método de Componentes Principais, consultar Andrade (1989), Johnson e Wichern (2007) e Manly e Alberto (2017).



A natureza das equações relativas aos CPs, e suas derivações, remetem à matriz de covariância do sistema original. Como mencionado em Johnson e Wichern (2007), os componentes principais (Y_p) dependem exclusivamente da matriz de covariância C (ou de correlação ρ) do vetor X' . Esta matriz é simétrica e possui a forma a seguir

$$C = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{p1} & c_{p2} & \cdots & c_{pp} \end{bmatrix}$$

em que os elementos da diagonal principal, c_{ii} , correspondem à variância de X_i e o restante, $c_{ii} = c_{ji}$, equivale à covariância das variáveis de $X_i = X_j$. Os autovalores desta matriz C podem ser ordenados: $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$. Além disso, como demonstrado em Johnson e Wichern (2007) e Manly e Alberto (2017), tem-se que as variâncias dos componentes principais correspondem aos autovalores da matriz C , de modo que λ_p corresponde ao p-ésimo componente principal e

$$\sum_{i=1}^p Var(X_i) = \sum_{i=1}^p Var(Y_i) = \lambda_1 + \cdots + \lambda_p = c_{11} + \cdots + c_{pp}$$

ou seja, as somas das variâncias das variáveis originais e dos componentes principais são equivalentes. Em consequência, os componentes principais expressam toda a variância do sistema original.

Como ressaltado em Queiroz (2003), depreende-se que a utilização da ACP permite que sejam estabelecidas as relações entre o conjunto de variáveis que compõe a base de dados e o comportamento dos indivíduos². Adicionalmente, seus resultados podem ser empregados como insumos em outras ferramentas de análise multivariada ou até mesmo em análises de regressão. Além disso, em teoria, o número de componentes principais equivale à quantidade de variáveis existente no estudo. Entretanto, visto que a maior parte dos componentes traduz pequenos percentuais da variância total do sistema, uma grande parcela desta variância pode ser representada por um pequeno conjunto de componentes.

Neste sentido, segundo Johnson e Wichern (2007), dado um número elevado de variáveis originais (p), se a maior parte da variância total da população puder ser atribuída a poucos componentes, conclui-se que eles podem ser utilizados no lugar de p sem que haja uma perda de informação muito significativa. Em outras palavras, não há prejuízo à qualidade da análise ao serem descartados os componentes principais que possuem uma baixa capacidade de explicação da variabilidade total do conjunto de variáveis (SILVA et al., 2005).

Com base nesta fundamentação, destaca-se que neste trabalho é definido um limite mínimo, em percentagem, de variância que deve ser explicada por um número qualquer de componentes principais. A partir dele, a análise de clusters é conduzida utilizando apenas os

² Vale salientar que não é postulada qualquer relação de causalidade entre estes fatores.

primeiros componentes que exprimirem este mínimo previamente determinado, de modo a priorizar a maior parcela de variância do sistema e condensar o número de variáveis analisadas.

Em conclusão, a Análise de Componentes Principais permite a síntese das informações contidas em uma base de dados em um número reduzido de dimensões. Procura-se, neste sentido, verificar como cada região relaciona-se com cada eixo previamente definido (mercado de trabalho, urbanização e qualidade de vida, estruturas de mercado e inovação) e empregar os resultados obtidos nesta técnica no ferramental de agrupamento das microrregiões de MG, discutido na seção seguinte.

5.3 Análise de Clusters

A análise de Clusters consiste em um conjunto de técnicas de classificação de indivíduos de uma base de dados em agrupamentos, fundamentadas no relacionamento entre os objetos de estudo. Segundo Hair Jr. et al.(2013), com as técnicas de análise de cluster procura-se, primariamente, classificar as observações a partir das semelhanças e diferenças de suas características, de acordo com as suas inter-relações, em uma quantidade menor de grupos mutuamente exclusivos. Os agrupamentos gerados a partir destes instrumentos devem exibir elevados graus de homogeneidade interna (dentro de cada cluster) e de heterogeneidade externa (entre clusters distintos). Dito de outra maneira,

"métodos de aglomeração (*clustering*) podem ser caracterizados como qualquer procedimento estatístico que, utilizando um conjunto finito e multidimensional de informações, classifica seus elementos em grupos restritos homogêneos internamente, permitindo gerar estruturas agregadas significativas e desenvolver tipologias analíticas." (SIMÕES,2003, p. 109).

Adicionalmente, além deste objetivo primordial das técnicas de agrupamento, outras finalidades também podem ser destacadas (MANLY; ALBERTO,2017). Em especial, a análise de clusters resulta na diminuição da quantidade de dados a ser investigada. Com a decorrente síntese das informações em um número reduzido de observações, é possível que os grupos gerados pelo procedimento permitam uma visualização concisa e eficiente dos padrões de semelhanças e diferenças entre seus integrantes (e entre os diferentes conjuntos formados). Em particular, este aspecto mostra-se interessante para diagnósticos de populações ou amostras representadas por bases de dados que apresentam um número elevado de observações. Ademais, agrupamentos inesperados revelados pelo método podem sugerir relações ou levar a hipóteses antes inexploradas que, incidentalmente, merecem ser investigadas. Por fim, a identificação de *outliers* também é proporcionada pelo uso deste instrumental.

Para alcançar estes diferentes propósitos, três aspectos devem ser levados em consideração (HAIR JR. et al.,2013): a) à medida de semelhança entre os objetos do universo;



b) a maneira de agrupar objetos que apresentam características semelhantes; c) a forma de avaliar os grupos formados e suas heterogeneidades internas e externas. Abaixo, as diretrizes escolhidas para cada uma destas questões no presente trabalho serão brevemente revisadas³.

Em primeiro lugar, a similaridade de cada indivíduo com relação aos outros é de importância central para a formação dos agrupamentos das observações estudadas. O conceito de similaridade entre objetos corresponde justamente à medida quantitativa da correspondência, ou semelhança, entre as observações (HAIR JR. et al., 2013). Segundo Everitt et al. (2011), as medidas de proximidade podem ser divididas em dois grupos: de distância e de correlação⁴. Habitualmente, para variáveis contínuas, as medidas de distância são utilizadas com maior frequência em análises de clusters. Para este trabalho isto não é diferente. A similaridade é representada pela proximidade de cada uma das observações às outras ao longo de todas as variáveis que as compõem. Desta forma, estas medidas indicam, na realidade, a dissimilaridade entre as observações, de modo que valores elevados correspondem a um menor grau de similaridade entre os objetos.

Ou seja, "dois indivíduos são 'próximos' quando as suas dissimilaridades ou distâncias são pequenas ou suas similaridades amplas" (EVERITT et al., 2011, p.43). A medida de distância mais comumente empregada na literatura é a distância euclidiana, dada por

$$d_{ij} = \left[\sum_{k=1}^p (x_{ik} - x_{jk})^2 \right]^{1/2}$$

em que x_{ik} e x_{jk} correspondem ao k-ésimo valor das observações i e j , respectivamente.

Em seguida, pode ser feita a seguinte divisão com relação à maneira de agrupar os indivíduos: métodos hierárquicos ou de partição (não-hierárquicos) (KAUFMAN; ROUSSEEUW, 1990). Cada um deles será brevemente descrito a seguir.

Os métodos hierárquicos têm como ponto de partida o cálculo da distância entre cada objeto e todos os outros pertencentes ao universo e podem ser divididos em dois tipos: aglomerativos e divisivos. Em linhas gerais, o primeiro parte de uma situação inicial em que cada observação é considerada seu próprio cluster, de tal sorte que o número de objetos (n) é igual à quantidade de grupos existentes (K), isto é, $n = K$. Procede-se, então, à união dos grupos mais próximos, sucessivamente, até que exista apenas um único agrupamento. Já os hierárquicos divisivos incorrem em um processo análogo, porém em direção oposta. Todo o

³ Para referências mais aprofundadas sobre os métodos, tipos de algoritmos, passos necessários para a realização dos procedimentos em Análises de Cluster e outras dúvidas, consultar Kaufman e Rousseeuw (1990), Everitt et al. (2011) e Hair Jr. et al. (2013).

⁴ Segundo os autores, diferentes medidas de (dis)similaridade calculadas para a mesma base de dados podem resultar em soluções distintas dos agrupamentos gerados. Cabe ao pesquisador, portanto, avaliar o tipo de medida a partir de seu conhecimento sobre o objetivo da pesquisa, a natureza dos dados e as variáveis sob investigação (EVERITT et al., 2011).

universo é inicialmente concentrado em um único grupo e são feitas divisões consecutivas até que cada cluster seja formado por e contenha exclusivamente um único objeto.

Além disso, em métodos hierárquicos aglomerativos, a distância entre os grupos deve ser medida ao serem formadas os agrupamentos. Diversas maneiras podem ser citadas, como a de ligação simples (também conhecida como técnica de vizinhos mais próximos), em que a distância entre grupos é definida pelos indivíduos mais próximos e apenas um par de indivíduos de cada grupo é considerado; ligação completa (vizinhos mais distantes), em que os indivíduos que possuem a maior distância entre si determinam a distância entre os grupos; entre outras. Neste trabalho, optou-se por utilizar a medida de Ward como medida entre os grupos. Nela, a dissimilaridade entre dois clusters é baseada na minimização da variância total intra-clusters. Em cada estágio os agrupamentos são determinados pelos pares de clusters que possuem a menor distância entre seus centroides.

Os modelos não-hierárquicos, por sua vez, possibilitam o "trânsito" das observações em grupos distintos em diferentes etapas do processo de agrupamento (MANLY; ALBERTO, 2017). Estes métodos são aplicados para a classificação dos objetos em um número k de clusters, que pode ser especificado de antemão ou ser definido como parte do procedimento de clusterização. O objetivo é dividir os itens em k grupos de modo que as observações dentro dos clusters tenham um maior grau de proximidade (calculado usualmente por medidas euclidianas de distância) entre elas do que entre itens de grupos distintos. O ponto de partida para métodos de partição é a definição da quantidade de clusters ou os pontos que formam seus núcleos. Como explicado em Johnson e Wichern (2007), o método de *K-means* representa um dos procedimentos de partição mais populares na literatura. O processo pode ser resumido da seguinte maneira.

Em primeiro lugar, a base de dados é dividida em k agrupamentos iniciais. Em seguida, são computadas as distâncias euclidianas de cada item aos centroides (núcleos) dos grupos e, a partir disso, uma observação é designada ao cluster cujo centroide encontra-se mais próximo. Os núcleos dos grupos são recalculados, para que possam ser tomadas suas medidas de distância às observações restantes. Este processo é repetido iterativamente até não seja possível designar nenhum elemento aos grupos existentes (JOHNSON; WICHERN, 2007).

Dito isso, a definição do método empregado deve partir do tipo de análise empreendida, das características da base de dados e, fundamentalmente, do objetivo do estudo. Cada método possui seus conjuntos de vantagens e desvantagens (HAIR JR. et al., 2013; JOHNSON; WICHERN, 2007). Procedimentos hierárquicos foram os primeiros métodos de agrupamentos desenvolvidos, são mais simples de serem empregados e são versáteis, no sentido que podem ser aplicados para quase qualquer tipo de pesquisa. Em contrapartida, em função de sua natureza, podem acarretar em combinações indesejadas entre variáveis ou que não sejam condizentes com a realidade formadas no início que perduram até o fim da análise. Tal característica pode gerar resultados e, conseqüentemente, conclusões artificiais. Por outro lado, métodos de partição podem ser aplicados à base de dados extremamente maiores do que métodos hierárquicos. Não obstante, métodos não-hierárquicos mostram-se menos eficientes para situações que apresentam um elevado número possíveis soluções de agrupamento, além de dependerem

da ordem em que se encontram as observações nas bases de dados. Em especial, dentro das técnicas de partição, o método de *K-means* é mais suscetível a *outliers*.

Como solução para estes problemas, a combinação entre os dois métodos é recomendada, dada por dois passos (HAIR JR. et al., 2013). Primeiramente o método hierárquico é utilizado para que os conjuntos completos de solução de cluster e o número apropriado de partições, k , sejam definidos. Em seguida, emprega-se esta quantidade de grupos k como ponto de partida em métodos não-hierárquicos. Este trabalho utiliza justamente este tipo de procedimento. O agrupamento é feito por meio do *critério de Ward* (nos componentes principais selecionados). Com isso, o número de clusters é definido. A técnica *K-means* é empregada a partir desta quantidade de agrupamentos obtida.

Por fim, deve-se avaliar a quantidade desejável de grupos formados. Para tanto, deve ser considerado o trade-off entre homogeneidade interna e o número de grupos formado. Um número reduzido de agrupamentos acarreta em menor homogeneidade intra-grupos, enquanto um maior número de grupos determina uma maior homogeneidade. Hair Jr. et al. (2013) argumentam que é desejável que haja parcimônia a este respeito, de modo que a estrutura básica seja atingida (ou seja, um número mínimo de clusters seja formado) juntamente com um nível satisfatório de heterogeneidade entre grupos.

Em resumo, primeiramente utilizou-se o método hierárquico aglomerativo, com as distâncias entre observações definidas por medidas euclidianas e entre os grupos pelo *critério de Ward*. O número de clusters foi escolhido a partir desta abordagem e, em seguida, empregado no método de partição *K-means*. Os resultados oriundos das análises multivariadas conduzidas neste estudo são apresentados na seção seguinte.

6 Resultados

Esta seção contém os resultados obtidos pelas técnicas de análise multivariada empregados para as microrregiões de Minas Gerais, tendo em vista aspectos relativos ao mercado de trabalho, à urbanização e qualidade de vida, a estruturas de mercado e à inovação. A Análise de Componentes Principais foi empregada para sintetizar a quantidade de observações que devem ser analisadas, além de indicar características da base de dados. A partir dos resultados alcançados por esta técnica, foram obtidos agrupamentos por meio da Análise de Clusters.

6.1 Tipologia a partir da Técnica de ACP

As variáveis que mais contribuem individualmente para o primeiro componente são, em ordem crescente, Rendimento Médio Mensal (RMM), Proporção da População Urbana (PPU), Taxa de Atividade (TXA), contribuindo positivamente, e Proporção de Domicílios com Saneamento Inadequado (SAN), negativamente. Esse componente sintetiza características das dimensões urbana e mercado de trabalho, fato que a variável com contribuição negativa



reforça. Colocam-se, de um lado, microrregiões mais urbanizadas, com alta taxa de atividade e remuneração dos trabalhadores formais, nomeadamente: Belo Horizonte, Uberaba, Uberlândia e Juiz de Fora. Do outro lado, há microrregiões como Grão Mogol, Araçuaí, Peçanha e Pedra Azul, distinguidas por baixos índices de urbanização e qualidade de vida e desenvolvimento socioeconômico de modo geral.

O segundo componente é construído basicamente a partir das variáveis Trabalhadores com Superior Completo (ESC) e Despesa com Investimento por IES (INV_IES), ambas contribuindo positivamente. Assim, reúne informações da qualificação da mão-de-obra e expansão da estrutura da formação dessa qualificação. As microrregiões que mais contribuem positivamente aqui são Diamantina, Viçosa, Belo Horizonte e Ouro Preto. Ou seja, possuem trabalhadores qualificados e seus centros de formação estão se expandindo. Alternativamente, Divinópolis, Campo Belo, Oliveira e Ubá se encontram em situação oposta.

Sobre o terceiro componente, Empregados dos Setores de Intermediação Financeira (FIN), QL de grandes empresas (QL_Gr) contribuem positivamente e Gasto com P&D (PeD_IES) e QL de Estabelecimentos de P&D (QL_PeD) contribuem negativamente. Esse componente resume características sobre grau de urbanização, industrialização e estrutura científica. Assim, centros urbanos, com concentração de grandes empresas e de estabelecimentos de P&D se destacariam positivamente. O fato de ter dois Quociente de Localização modelando a maior parte de sua variância pode explicar algumas distorções, como a proeminência de microrregiões como Grão Mogol e Janaúria.

Finalmente, o quarto componente é formado, principalmente, pelas contribuições positivas das variáveis QL Setores Extrativos Minerais (QL_Ext) e QL Setores de Alta Tecnologia (QL_Tec) e contribuição negativa de QL de Estabelecimentos de P&D (QL_P&D). Esse componente apresenta informações sobre as estruturas científica e tecnológica, bem como proximidade com matéria-prima para a indústria. Pedra Azul, Santa Rita do Sapucaí, Aracuaí e Ouro Preto despontam contribuindo positivamente, e a interpretação é que estas são microrregiões ricas em recursos minerais que concentram atividades produtivas de alta tecnologia. Do lado oposto, há Janaúba, Viçosa, Capelinha e Ituiutaba contribuem negativamente nesse componente. Mais uma vez, a explicação para distorções é dada pela forte presença de variáveis na forma de Quociente de Localização.

Os resultados obtidos a partir desta técnica de Análise Estatística Multivariada podem ser resumidos pelas tabelas 1 e 2, apresentadas nesta seção. Como pode ser visto, os quatro primeiros CPs explicam quase 70% da variância total da amostra. Para Johnson e Wichern (2007), os componentes que explicarem uma maioria expressiva da variância podem substituir a amostra original nas subseqüentes análises feitas, sem perdas significativas de informação. Este é o procedimento feito neste trabalho, de modo que os agrupamentos são criados tendo em vista os quatro primeiros componentes gerados.

Tabela 1 - Análise de Componentes Principais - Total da variância explicada

Componentes	Variância (%)	
	Individual	Acumulada
1	34,27	34,27
2	15,25	49,52
3	10,21	59,72
4	8,70	68,42
5	7,94	76,37
6	7,66	84,03
7	5,67	89,69
8	4,63	94,33
9	2,49	96,82
10	2,03	98,84
11	0,81	99,65
12	0,35	100,00

Fonte: Elaboração própria a partir dos dados do Censo (2010), Base de Dados de Comércio Exterior Brasileiro (ComexStat), Relação Anual de Informações Sociais (RAIS) de 2010, Censo da Educação Superior (INEP) e Produto Interno Bruto (PIB) dos Municípios (IBGE).

Tabela 2 – Matriz dos Coeficientes dos Componentes Principais - Contribuições das variáveis

Variáveis	Componentes			
	1	2	3	4
RMM	0,931	0,033	0,137	0,041
TXA	0,827	-0,278	0,035	0,018
ESC	0,354	0,814	0,127	-0,096
PPU	0,879	-0,252	-0,022	0,120
SAN	-0,917	0,176	0,092	-0,157
FIN	0,425	0,383	0,652	0,004
QL_Tec	0,311	-0,018	0,188	0,447
QL_Gr	-0,445	0,301	0,458	0,096
QL_Ext	-0,375	0,136	-0,143	0,744
INV_IES	0,217	0,738	-0,278	0,069
PeD_IES	0,186	0,340	-0,539	0,234
QL_PeD	0,352	0,276	-0,348	-0,414

Fonte: Elaboração própria a partir dos dados do Censo (2010), Base de Dados de Comércio Exterior Brasileiro (ComexStat), Relação Anual de Informações Sociais (RAIS) de 2010, Censo da Educação Superior (INEP) e Produto Interno Bruto (PIB) dos Municípios (IBGE).

6.2 Tipologia a partir da Análise de Clusters

Os resultados da técnica de Componentes Principais apresentados na subseção anterior foram usados como insumos para o agrupamento das microrregiões de Minas Gerais de acordo com as similaridades entre as observações. Procurou-se, portanto, reduzir as dimensões analisadas para que fosse feito a junção das variáveis. A metodologia de análise de clusters empregada pode ser resumida em duas etapas: *i)* método hierárquico aglomerativo, utilizando medidas euclidianas e o *Critério de Ward* como medidas de



distância entre observações e grupos (para a obtenção do número de clusters e da visualização do processo para identificar possíveis agrupamentos que fizessem mais sentido para o estudo em questão); *ii*) método não-hierárquico, mais especificamente conhecido como *K-means*, com uso do número de clusters obtido na primeira etapa. Dito isso, os produtos das técnicas aplicadas são apresentados abaixo.

Foram obtidos seis agrupamentos para as microrregiões de MG. A Tabela 3 contém uma breve caracterização de cada um deles, informando o número de microrregiões que os compõem e as variáveis mais associadas a cada grupo, de maneira que são indicados as médias obtidas para estas variáveis em cada conjunto formado, as suas médias totais e o *p*-valor a elas associado⁵.

O primeiro e segundo clusters formados localizam-se, em sua grande maioria, na porção oeste do estado de Minas Gerais. Ambos apresentam uma associação negativa com o primeiro Componente Principal 1, traduzido, especialmente, em agrupamentos compostos por microrregiões cujos indicadores relativos à urbanização e qualidade de vida e ao mercado de trabalho são mais deprimidos. O primeiro agrupamento é formado por apenas três microrregiões, Grão Mogol, Araçuaí e Pedra Azul, localizadas em áreas contíguas no Norte de Minas e no Jequitinhonha. O cluster é marcado por áreas debilitadas em quesitos socioeconômicos com indicativos de uma urbanização precária, com as mais baixas médias de rendimento médio mensal da população, maior proporção de residentes em áreas não-urbanas e maiores taxas de população com acesso a saneamento inadequado dentre todos os grupos formados. Continuando na porção oeste do estado, o segundo cluster, carrega ao todo 19 territórios, pertencentes principalmente ao Norte de Minas, Jequitinhonha, Vale do Mucuri, Vale do Rio Doce e Zona da Mata. Similarmente ao primeiro este conjunto também apresenta um índice de acesso a saneamento apropriado e população residentes em áreas urbanas relativamente baixos. Além disso, estas microrregiões possuem, em média, taxas de atividade e um nível de renda menores do que a média do estado.

O terceiro cluster é composto pelo mesmo número de regiões do primeiro (3), porém com características fundamentalmente opostas. Este agrupamento pode ser ligado ao segundo e terceiro Componentes Principais. Como consequência, possuem níveis de indicadores relacionados à estrutura científica e qualificação de seus trabalhadores com valores extremamente acima da média estadual. As microrregiões de Diamantina, Ouro Preto e Viçosa diferenciam-se no cenário mineiro por suas estruturas de produção científica, com a presença de universidades federais (Universidade Federal de Ouro Preto e Universidade Federal de Viçosa, por exemplo). Os gastos elevados com Investimento e Pesquisa & Desenvolvimento por parte das Instituições de Ensino Superiores relacionam-se a estas peculiaridades, demonstrando à força relativa com relação à inovação e à produção de conhecimento científico destas microrregiões no contexto mineiro.

Passando para a parte central-ocidental do estado, o quarto cluster corresponde ao maior agrupamento obtido pelos métodos empregados, contendo 30 microrregiões. Este cluster é formado por microrregiões que englobam todo o Noroeste de Minas, passando pelo Triângulo Mineiro e Alto do Parnaíba, Central Mineira, Região Metropolitana de Belo

⁵ Para uma visualização espacial dos clusters formados, consultar o mapa contido no Apêndice B.

horizonte, Vale do Rio Doce, Oeste, Sul e Sudeste de Minas e Campos das Vertentes, até a Zona da Mata. Em sua maioria, seus integrantes possuem uma situação mais favorável em relação ao mercado de trabalho do que os clusters 1 e 2, por exemplo, sendo identificados tanto por um nível de rendimento quanto por uma proporção da população economicamente ativa com relação à população acima da média. Em contrapartida, não parecem serem ligados à ambientes com estruturas de produção científica e de inovação consolidadas, tendo em vista suas relações negativas ao segundo componente, que se traduz em médias bastante inferiores de despesas de ensinos superiores e de qualificação dos trabalhadores em comparação à média de MG.

O cluster de número 5 formado pelo procedimento previamente descrito abarca regiões espacialmente diversas do estado de Minas Gerais. Seus membros encontram-se no Norte, Oeste e Sul do estado, com integrantes do Norte de Minas (Montes Claros), Triângulo Mineiro (Ituiutaba, Uberlândia e Uberaba), Sul e Sudoeste de Minas (Alfenas, Pouso Alegre e Itajubá), Campos das Vertentes (Lavras e São João Del Rei) e Zona da Mata (Juiz de Fora). Salienta as afinidades demonstradas por este cluster com relação aos dois primeiros Componentes Principais da metodologia aplicada, relacionados principalmente a fatores positivos do mercado de trabalho, qualidade de vida da população e da estrutura de produção científica. Com relação à dimensão do mercado de trabalho, é notório que a média do indicador de renda deste agrupamento seja inferior apenas a Belo Horizonte, capital e um polo de concentração de renda e estrutura produtiva do estado (CERQUEIRA; SIMÕES, 1997). Além disso, todos os índices desta dimensão do quinto cluster encontram-se acima da média estadual, de maneira significativa. Adicionalmente, o cluster comporta microrregiões que exibem estruturas espaciais-setoriais especializadas em Pesquisa e Desenvolvimento em relação ao restante de Minas, expresso pelo Quociente Locacional deste fator e do Investimento gasto por IES.

O sexto e último cluster é composto somente pela microrregião de Belo Horizonte. A concentração espacial de renda, urbanização e de estrutura produtiva são fatos notórios com relação a capital do estado e amplamente discutidos na literatura. Vale salientar apenas que os indicadores mais característicos deste agrupamento, pela metodologia empregada, apontam para um diferencial de serviços financeiros, qualificação dos trabalhadores e rendimento médio da população.



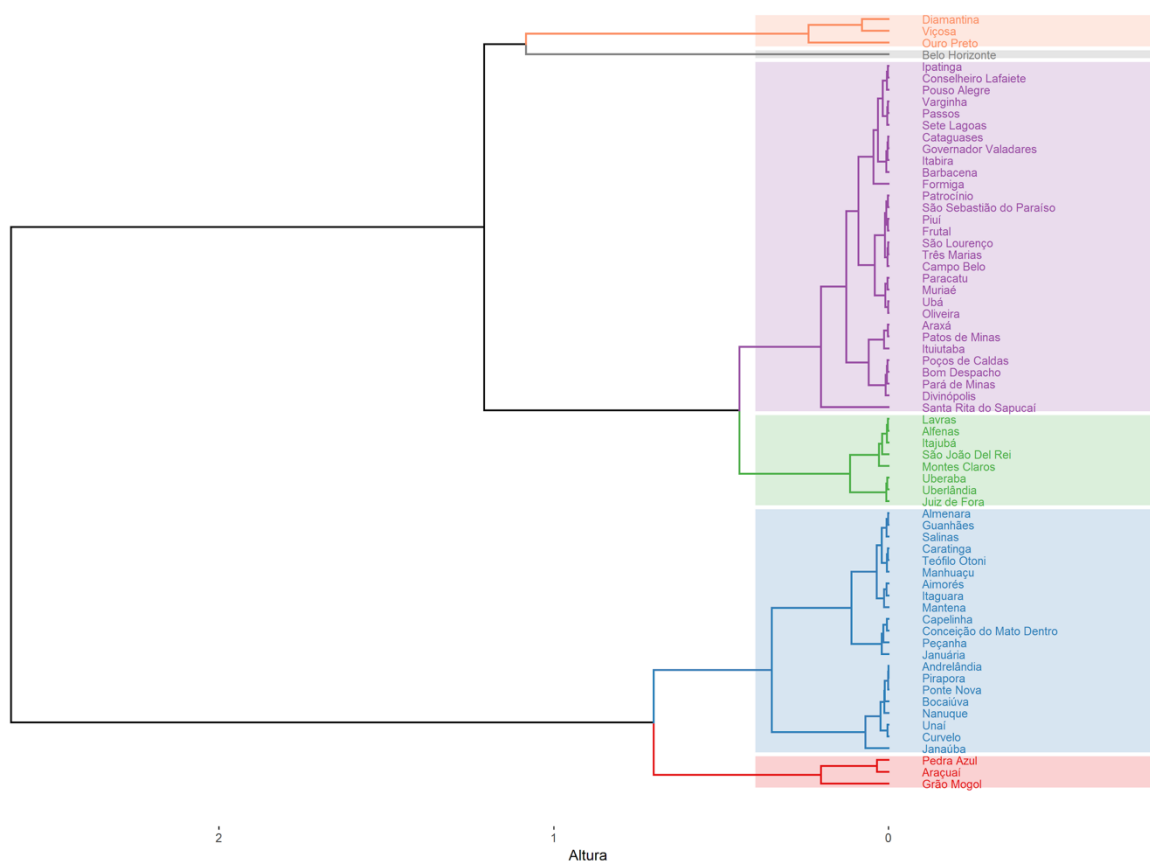
Tabela 3 - Clusters das Microrregiões de MG - Variáveis mais associadas a cada Cluster

Cluster	Cor	Número de Microrregiões	Variáveis	Média na Categoria	Média Total	P-valor
1	Vermelho	3	QL_Ext	6,354	1,432	1,6E-08
			QL_Gr	2,224	0,918	2,4E-06
			SAN	0,340	0,141	0,001
			RMM	659,49	1.057,53	0,004
			TXA	0,423	0,490	0,002
			PPU	0,550	0,778	0,002
2	Azul	19	SAN	0,243	0,141	3,7E-07
			INV_IES	104.270,60	1.618.553,53	0,033
			TXA	0,456	0,490	4,8E-06
			PPU	0,661	0,778	3,9E-06
			RMM	810,65	1.057,53	2,2E-07
3	Laranja	3	INV_IES	13.755.735,86	1.618.553,53	4,7E-09
			PeD_IES	2.695.904,36	198.293,18	1,3E-06
			ESC	0,181	0,123	0,000
4	Roxo	30	PPU	0,848	0,778	8,1E-05
			TXA	0,511	0,490	1,0E-04
			RMM	1.153,98	1.057,53	0,004
			INV_IES	494.604,18	1.618.553,53	0,023
			ESC	0,113	0,123	0,006
			QL_Gr	0,726	0,918	0,004
			SAN	0,086	0,141	7,5E-05
5	Verde	10	ESC	0,149	0,123	0,001
			RMM	1.285,35	1.057,53	0,001
			QL_PeD	1,619	0,771	0,002
			INV_IES	4.517.831,11	1.618.553,53	0,007
			TXA	0,516	0,490	0,025
			PPU	0,862	0,778	0,028
			SAN	0,064	0,141	0,011
6	Cinza	1	FIN	31892	1021	1,8E-15
			ESC	0,217	0,123	0,001
			RMM	1.814,09	1.057,53	0,002

Fonte: Elaboração própria a partir dos dados do Censo (2010), Base de Dados de Comércio Exterior Brasileiro (ComexStat), Relação Anual de Informações Sociais (RAIS) de 2010, Censo da Educação Superior (INEP) e Produto Interno Bruto (PIB) dos Municípios (IBGE).

Adicionalmente, o processo de obtenção dos agrupamentos pode ser visualizado na Figura 1. Este diagrama, conhecido como Dendrograma, corresponde a representação gráfica da geração dos agrupamentos pelos métodos hierárquicos empregados neste trabalho. Nele, os nós terminais correspondem às microrregiões, as ligações entre cada haste determinam os clusters que foram formados durante o processo e seus comprimentos representam as distâncias em que eles são agrupados (EVERITT et al., 2011). Os conjuntos resultantes do final do processo foram coloridos para uma melhor visualização. Com este objeto, a partir desta representação visual da distância entre as observações, percebe-se, por exemplo, como a microrregião de BH pode ser entendida como um *outlier* na amostra, em função de indicadores completamente díspares com relação ao resto da amostra. Além disso, outro ponto interessante no processo de agrupamento refere-se à microrregião de Santa Rita do Sapucaí, reconhecida como um polo de inovação no estado mineiro, expresso indicadores relativos a estruturas locais da indústria e de produção científica acima da média. Como pode ser visto no Dendrograma, há uma distância considerável entre suas variáveis e o restante do quarto cluster, que ela pertence. Isto indica a peculiaridade desta região com relação às dimensões estudadas neste trabalho.

Figura 1 – Dendrograma da Análise de Cluster

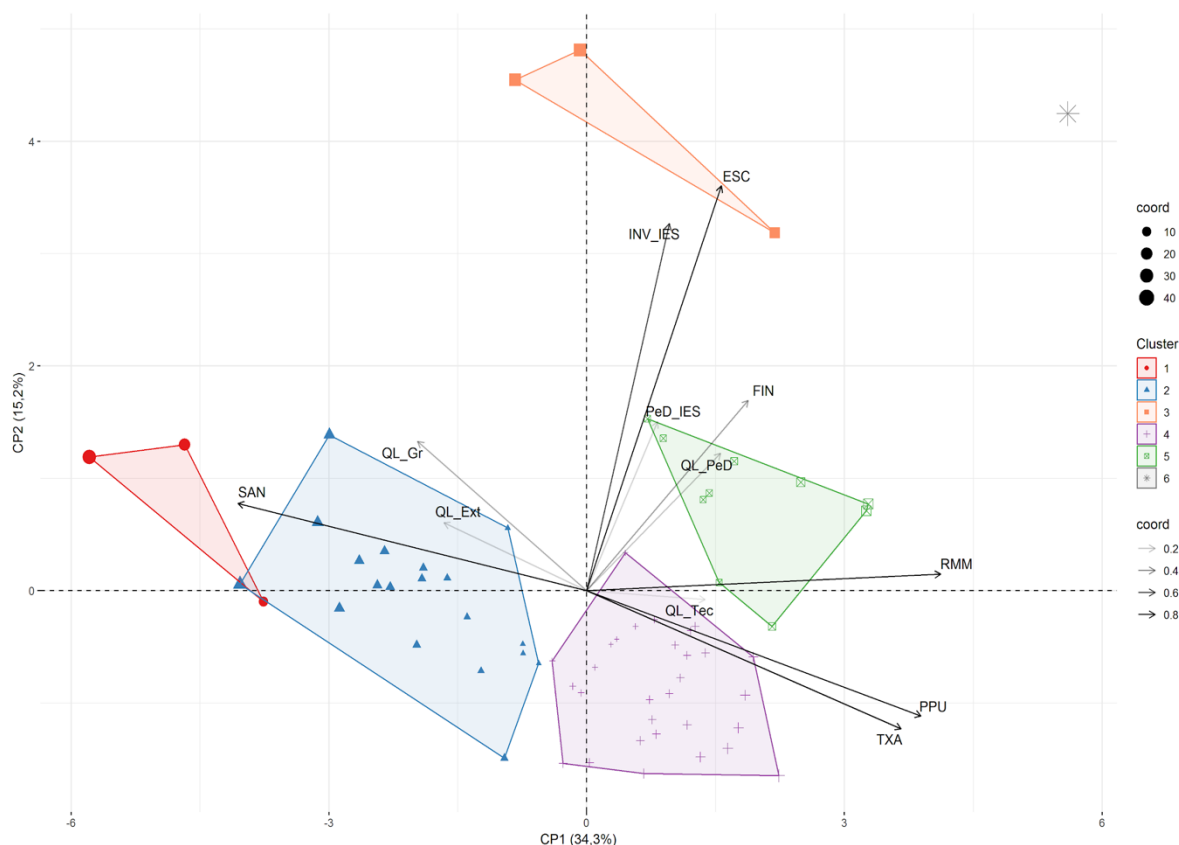


Fonte: Elaboração própria a partir dos dados do Censo (2010), Base de Dados de Comércio Exterior Brasileiro (ComexStat), Relação Anual de Informações Sociais (RAIS) de 2010, Censo da Educação Superior (INEP) e Produto Interno Bruto (PIB) dos Municípios (IBGE).

Finalmente, a Figura 2 pode ser vista como uma síntese dos resultados obtidos neste trabalho. Todas as variáveis e observações encontram-se dispostas em dois planos, compostos pelos dois primeiros CPs, 1 e 2, obtidos na técnica de ACP, eixos horizontal e vertical, respectivamente. As relações entre todas as variáveis (com relação a estes componentes) estão contidas na imagem, de modo que variáveis que possuam correlações positivas encontram-se agrupadas e direcionadas em um mesmo sentido. A distância entre a variável e a origem e a nitidez da seta de cada uma indicam a sua qualidade no mapa. As observações, agrupadas pelos clusters formados, foram plotadas nos planos de acordo com suas contribuições para variância destes dois primeiros componentes. Desta forma, as relações descritas anteriormente entre as variáveis, microrregiões, os clusters formados e os dois primeiros componentes principais são expressas na Figura.

Como mencionado, os clusters 1 e 2 possuem relações negativas com indicadores de urbanização e qualidade de vida, dada o direcionamento das setas relativas às variáveis SAN, PPU e TXA. O terceiro cluster, formado por Diamantina, Ouro Preto e Viçosa, caracteriza-se por uma alta correlação com a dimensão estrutura da produção científica, sendo um possível polo científico. O quarto cluster, apesar de apresentar indicadores de urbanização e mercado de trabalho ligeiramente mais elevados do que a média estadual, possui uma baixa relação com estruturas relacionadas à inovação. Como esperado, o sexto cluster, constituído somente pela capital, é claramente um *outlier* da base estudada. E, como destaque final, pode ser salientado as médias mais elevadas relativamente à média estadual do cluster 5 com relação a alguns indicadores de mercado de trabalho, estrutura da indústria e de produção científica.

Figura 2 – Resultados das técnicas de ACP e Clusters



Fonte: Elaboração própria a partir dos dados do Censo (2010), Base de Dados de Comércio Exterior Brasileiro (ComexStat), Relação Anual de Informações Sociais (RAIS) de 2010, Censo da Educação Superior (INEP) e Produto Interno Bruto (PIB) dos Municípios (IBGE).

7 Conclusão

Esse artigo procurou apresentar um estudo das 66 microrregiões de Minas Gerais quanto à suas contribuições ao Sistema Regional de Inovação do estado e identificar padrões de similaridade. A atividade inovativa é complexa e compreende atividades científicas, tecnológicas e informacionais. Pessoas, empresas e Estado se relacionam no contexto local de forma a compartilhar conhecimento.

Um estudo que busca classificar um Sistema de Inovação deve levar em consideração todas estas dimensões. Esse trabalho o fez através da construção de 12 indicadores que buscaram captar aspectos do mercado de trabalho, da estrutura de mercado e intensidade tecnológica do setor produtivo, das instituições de ensino e pesquisa e da infraestrutura urbana.

Após a aplicação de técnicas de análise estatística multivariada, nominalmente as análises de Componentes Principais e de Cluster, foi possível organizar as microrregiões em grupos homogêneos internamente com características que informam seu papel no sistema regional. Foram identificados 6 clusters, sendo que a microrregião contendo a capital do estado formou um cluster unitário.

Os clusters 1 e 2 (vermelho e azul) localizados, em sua maioria, na porção nordeste do estado, são formados por microrregiões com pouca indústria de alta tecnologia, proporção de população urbana menor do que a média do estado, menores níveis de renda e gastos irrisórios com pesquisa e desenvolvimento. Por isso, pode-se dizer que não são regiões promissoras em termos de inovação, contribuindo mais com fornecimento de matéria prima e alimentos.

O cluster 3 (laranja) é formado por apenas 3 microrregiões: Ouro Preto, Diamantina e Viçosa. É o cluster especializado em atividade científicas, pois apresenta altos níveis de gasto em estrutura física e em pesquisa, além de elevada proporção de trabalhadores altamente qualificados.

O cluster 4 (roxo) possui indicadores de mercado de trabalho melhor do que a média do estado, porém não se destaca nos outros determinantes da inovação.

O cluster 5 (verde), por outro lado, não só apresenta mão-de-obra qualificada e especializada, como concentra indústrias de alta tecnologia e instituições de pesquisa e desenvolvimento. É, portanto, o retrato das regiões promissoras em termos inovativos.

Finalmente, o cluster unitário formado pela microrregião de Belo Horizonte possui índices que podem ser explicados pelo fato de conter a capital do Estado. É dinâmico, diversificado, rico, inovador e concentrador de infraestruturas urbanas, científicas, tecnológicas, produtivas e humanas.



REFERÊNCIAS

- ALBUQUERQUE, E.; SIMÕES, R.; BAESSA, A.; CAMPOLINA; B.; SILVA, L. A Distribuição Espacial da Produção Científica e Tecnológica Brasileira: uma Descrição de Estatísticas de Produção Local de Patentes e Artigos Científicos. *Revista Brasileira de Inovação*, v.1, n.2, p. 225-251, jul-dez. 2002
- ANDRADE, T. A. Métodos estatísticos e econométricos aplicados à análise regional. In: HADDAD, P. R. (Ed.). *Economia Regional: teorias e métodos de análise*. Fortaleza: BNB/ETENE, 1989. cap. 9, p. 427–507.
- ARROW, K. J. The economic implications of learning by doing. *Review of Economic Studies*, v. 29, p. 155–172, 1962.
- ASHEIM, B. Industrial Districts as “learning regions”: condition for prosperity? Em: *IGU Commission on Interdependent and Uneven Development: global-local perspectives*. Seoul, p.7-11, ago. 1995, Seoul: Step Group.
- CERQUEIRA, H. E. da G.; SIMÕES, R. F. Modernização e diferenciação econômica em Belo Horizonte. *Varia História*, v. 2, n. 18, p. 443–463, set. 1997.
- CROCCO, M. A. et al. Metodologia de identificação de aglomerações produtivas locais. *Nova Economia*, v. 2, n. 16, p. 211–241, maio-ago. 2006.
- EVERITT, B. S. et al. *Cluster Analysis*. 5. ed. Chichester: John Wiley Sons, Inc., 2011.
- HAIR JR., J. F. et al. *Multivariate Data Analysis*. Pearson new international edition. Harlow: Pearson Education Limited, 2013.
- DOSI, G. The nature of the innovative process. Em: *DOSI, G. (Ed.). Technical Change and Economic Theory*. Londres: Pinter, 1988. cap.10, p.221- 238.
- JACOBS, J. *The Economy of Cities*. New York: Random House Inc, 1969.
- JOHNSON, R. A.; WICHERN, D. W. *Applied multivariate Statistical Analysis*. 6. ed. New Jersey: Pearson Prentice Hall, 2007.
- KAUFMAN, L.; ROUSSEEUW, P. J. *Finding Groups in Data: An introduction to cluster analysis*. New York: John Wiley Sons, Inc., 1990.
- LEMOS, Mauro B.; DINIZ, Clélio C. *Sistemas Regionais de Inovação: o caso de Minas Gerais*. Rio de Janeiro: IE/UFRJ, mar. 1998. (Nota Técnica 06/98).
- LUNDEVALL, B. (Ed.) *National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning*. Londres: Pinter, 1992.

MANLY, B.; ALBERTO, J. A. N. *Multivariate statistical methods: a primer*. 4. ed. Boca Raton: CRC Press, 2017.

NELSON, R. e ROSENBERG, N. *Technical Innovation and National Systems*. Em: NELSON, R. (Ed.). *National Innovation Systems*. New York: Oxford University Press, 1993. cap.1, p.3–21

NELSON, R.; WINTER, S. G. In Search of Useful Theory of Innovation. *Research Policy*, V.6, p. 36-76. 1977.

QUEIROZ, E. A. de. *Pluriatividade e inserção das famílias rurais no processo de urbanização do rural: uma aplicação de técnicas estatísticas de análise multivariada para minas gerais*. Mestrado em Economia — Centro de Desenvolvimento e Planejamento Regional, Universidade Federal de Minas Gerais, Belo Horizonte, 2003.

SANTOS, U. P.; CALIARI, T. Distribuição Espacial das Estruturas de Apoio às Atividades Tecnológicas no Brasil: uma análise multivariada para as cinquenta maiores microrregiões do país. *Economia*. v.13, set. 2012.

SCHUMPETER, J. *A Teoria do Desenvolvimento Econômico*. São Paulo: Nova Cultural, 1985 [1911].

SILVA, H. et al. Agropecuária e urbanização: Uma análise multivariada para minas gerais, 1995-2000. *Revista Econômica do Nordeste*, Fortaleza, v. 39, n. 2, abr. - jun. 2005.

SIMÕES, R. F. *Localização industrial e relações intersetoriais: uma análise de fuzzy cluster para Minas Gerais*. Doutorado em Ciências Econômicas — Universidade Estadual de Campinas, Campinas, 2003.

SIMÕES, R. F. Métodos de análise regional e urbana: diagnósticos aplicados ao planejamento. *Texto para discussão*, Belo Horizonte, n. 259, 2005.

SZAPIRO, M.; MATTOS, M.; CASSIOLATO, J. *Sistemas de Inovação e Desenvolvimento*. Em: RAPINI, M.; SILVA, L.; ALBUQUERQUE, E. (Org.). *Economia da Ciência, Tecnologia e Inovação: fundamentos teóricos e economia global*. 1.ed. Curitiba: Prismas, 2017. cap.10, p371-412.



APÊNDICE A: VARIÁVEIS E INDICADORES UTILIZADOS

Tabela 4 – Variáveis e Indicadores utilizados

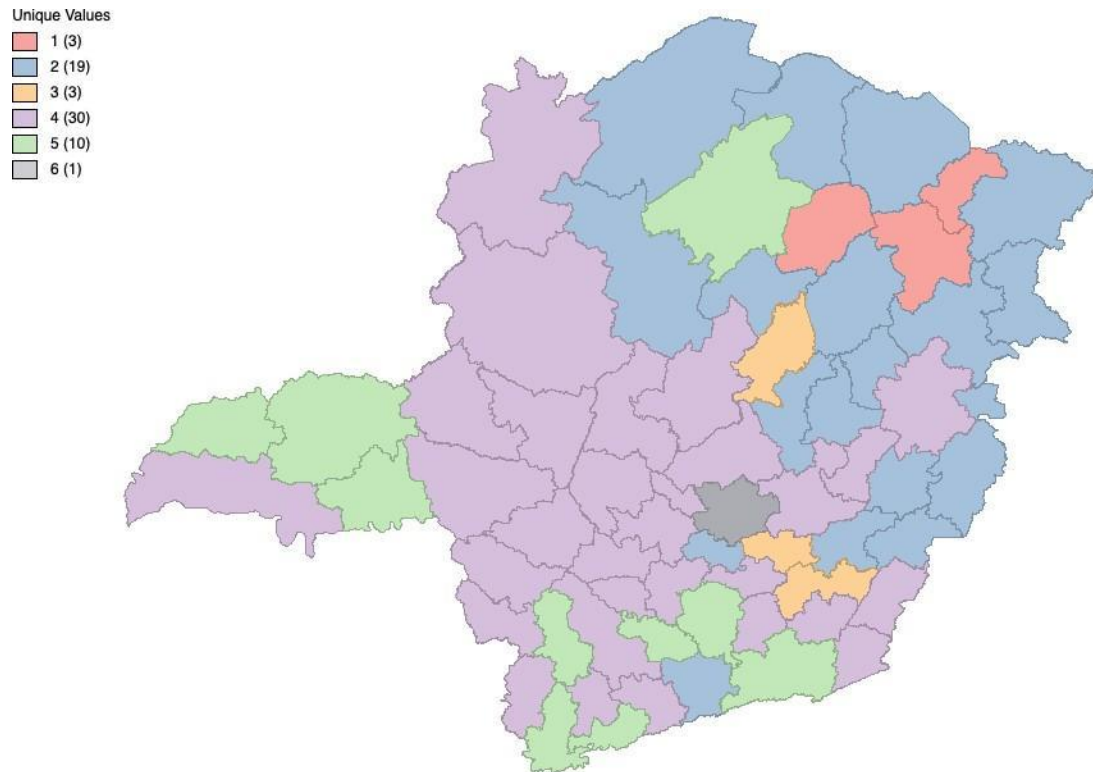
Dimensão	Sigla	Variável	Descrição	Base de dados
Mercado de Trabalho	ESC	Trabalhadores Superior Completo	Porcentagem de Trabalhadores com ensino Superior Completo no mercado formal	RAIS 2010
Mercado de Trabalho	RMM	Rendimento Médio Mensal	Valor do rendimento nominal médio mensal das pessoas de 10 anos ou mais de idade, com rendimento (R\$) economicamente ativas na semana de referência	Censo Demográfico 2010
Mercado de Trabalho	TXA	Taxa de Atividade	População Economicamente Ativa sobre População total com 10 anos de idade ou mais	Censo Demográfico 2010
Urbanização e Qualidade de Vida	PPU	Proporção População Urbana	População residente urbana sobre população total	Censo Demográfico 2010
Urbanização e Qualidade de Vida	SAN	Proporção população com saneamento inadequado	População residente em domicílios particulares com saneamento inadequado por população residente em domicílios particulares permanentes	Censo Demográfico 2010
Urbanização e Qualidade de Vida	FIN	Total de empregados do setor de intermediação financeira	Soma empregados Div-CNAE 64, 65 e 66	RAIS 2010
Estruturas da Indústria Local	QL_Tec	QL de setores de Alta Tecnologia	Quociente de Localização setores de alta tecnologia (Div-CNAE 29,30,31,33, e 35)	RAIS 2010
Estruturas da Indústria Local	QL_Gr	QL de Grandes Empresas	QL de empresas que apresentam 250 ou mais empregados	RAIS 2010
Estruturas da Indústria Local	QL_Ext	QL de setores Extrativos Minerais	Quociente de Localização calculado a partir de dados de número de estabelecimentos	RAIS 2010
Estruturas da Produção Científica	QL_PeD	QL de estabelecimentos de P&D	Quociente de Localização calculado a partir de dados de número de estabelecimentos	RAIS 2010
Estruturas da Produção Científica	INV_IES	Investimento por IES	Despesa média com Investimento por Instituições de Ensino Superior (IES) de cada microrregião	Censo da Educação Superior 2010
Estruturas da Produção Científica	PeD_IES	P&D por IES	Despesa média com P&D por IES de cada microrregião	Censo da Educação Superior 2010

Fonte: Elaboração própria a partir dos dados do Censo (2010), Base de Dados de Comércio Exterior Brasileiro (ComexStat), Relação Anual de Informações Sociais (RAIS) de 2010, Censo da Educação Superior (INEP) e Produto Interno Bruto (PIB) dos Municípios (IBGE).



APÊNDICE B: MAPA DOS CLUSTERS

Figura 3 – Clusters das Microrregiões de Minas Gerais



Fonte: Elaboração própria a partir dos dados do Censo (2010), Base de Dados de Comércio Exterior Brasileiro (ComexStat), Relação Anual de Informações Sociais (RAIS) de 2010, Censo da Educação Superior (INEP) e Produto Interno Bruto (PIB) dos Municípios (IBGE).