**Comparação de algoritmos de agrupamento espacial: um estudo de caso sobre a vulnerabilidade social da população da cidade de Belo Horizonte** (Deve conter no máximo 15 palavras e não deve conter expressões redundantes como: “Estudo de...”; “Influência de...”; “Elaboração de...” “Efeito de...“, “Análise de...”)

Renato Godoi da Cruz¹\*;Auberth Henrik Venson2

1 Nome da Empresa ou Instituição (opcional). Titulação ou função ou departamento. Rua Monte Alegre – São Lucas; 30240-282. Belo Horizonte, Minas Gerais, Brasil.

2 Nome da Empresa ou Instituição (opcional). Titulação ou função ou departamento. Endereço completo (pessoal ou profissional) – Bairro; 00000-000 Cidade, Estado, País

\*autor correspondente: renatogcruz@hotmail.com

**Comparação de algoritmos de agrupamento espacial: a vulnerabilidade social da população de Belo Horizonte** (Deve conter no máximo 15 palavras e não deve conter expressões redundantes como: “Estudo de...”; “Influência de...”; “Elaboração de...” “Efeito de...“, “Análise de...”)

**Resumo**

Tópico obrigatório para o depósito do TCC, porém opcional para a etapa dos Resultados preliminares.

**Palavras-chave:** principal component analytics; latent class growth modeling; k-means; multinomial logistic regression.

**Introdução**

O título da seção “Introdução” deve ser grafado em negrito, alinhado à esquerda e com a primeira letra maiúscula. Nesta seção o autor deve contextualizar o assunto, estabelecendo uma relação com outros trabalhos publicados sobre o tema e apresentar a justificativa para a realização do trabalho, situando a importância do problema estudado e solucionado, apoiado em citações bibliográficas de fontes confiáveis. O último parágrafo da introdução deve discriminar o objetivo do trabalho, que deve ser enunciado de forma coerente e direta.

A **Introdução** deve ser redigida em no **máximo duas páginas** e não deve conter subtópicos, figuras ou tabelas. Ressalta-se que o TCC não deverá exceder o total de 30 páginas, incluindo o(s) Apêndice(s) e/ou Anexo(s), sendo este o máximo de páginas permitido pelas “Normas de Elaboração do Trabalho de Conclusão de Curso” do MBA USP ESALQ.

Este trabalho consiste na análise da vulnerabilidade social da população da cidade de São Paulo. Para realizar o estudo, serão utilizadas as informações do último censo disponíveis para os 18.952 setores censitários do Município de São Paulo. Esses setores serão classificados de acordo com a maior ou menor presença de características sociais que contribuem para tornar uma família mais vulnerável socialmente. Para tanto, será necessário identificar, inicialmente, quais são as variáveis que melhor expressam o a vulnerabilidade social. Por vulnerabilidade social, entende-se, segundo Semzezem e Alves (2013), situações de empobrecimento da classe trabalhadora, relacionadas a dificuldades materiais para a manutenção da sobrevivência, assim como as dificuldades relacionais e culturais, que interferem na forma de viver dos trabalhadores e de suas famílias.

Dos Censos de 1990, 2000 e 2010, serão selecionadas variáveis utilizadas em estudos sobre a pobreza, que serão submetidas a uma técnica estatística chama analise fatorial. Essa técnica permite identificar, ao mesmo tempo, quais são as variáveis que apresentam coeficientes de correlação relativamente elevados entre si e estabelecer novas variáveis, chamado de fatores, que sejam capazes de captarem o comportamento conjunto das variáveis originais (Fávero e Belfiore, 2017).

Os grupos de vulnerabilidade serão classificados a partir do agrupamento da dimensão de privação socioeconômica e, também, da dimensão demográfica, já que o território, como afirmam Couto et al. (2010), é o terreno das políticas públicas, fundamental para a compreensão do conceito de vulnerabilidade social, posto que nele se determinam as condições de vida dos segmentos populacionais vulnerabilizados.

Tal agrupamento será realizado a partir da aplicação de outra técnica estatística denominada análise de agrupamentos, que permite agrupar os setores censitários que possuem características sociais semelhantes. A degradação em grupos de vulnerabilidade será feita considerando-se a intenção de captar a semelhança entre si, em relação às variáveis inseridas na análise, e consideravelmente diferença das observações de outros grupos (Centro de Estudos da Metrópole, 2004).

Para avaliar o desempenho relativo dos modelos de agrupamento na identificação de vulnerabilidade social, uma análise regressão logística multinomial (MLR) será conduzida para examinar empiricamente como um conjunto de preditores teoricamente associados à vulnerabilidade explica cada grupo.

De acordo com o site Panorama Cidades (2022), do Instituto Brasileiro de Geografia e Estatística, o Município de São Paulo, no ano de 2010, apresentava uma população de cerca de 12.396.372 habitantes distribuídos por um território de 1.521,202 km². Apresentava 92.6% de domicílios com esgotamento sanitário adequado, 74.8% de domicílios urbanos em vias públicas com arborização e 50.3% de domicílios urbanos em vias públicas com urbanização adequada (presença de calçada, pavimentação, meio-fio e bueiro).

Quando comparado com as outras cidades do estado de São Paulo, ocupava a 225º, 517º e 83º posições num universo de 645 cidades, respectivamente. Já quando comparado aos municípios do Brasil, suas posições eram 317º, 2779º e 419º de 5570, respectivamente.

A taxa de mortalidade infantil média no município era de 11.21 para 1.000 nascidos vivos. As internações devido a diarreias são de 0.3 para cada 1.000 habitantes. Comparado com todas as cidades do estado, ficava nas posições 277º e 332º, respectivamente, do total de 645. Quando comparado a municípios do Brasil, essas posições eram de 2796º e 3907º de 5570, respectivamente.

O salário médio mensal, em 2019, era de 4,1 salários mínimos. A proporção de pessoas ocupadas em relação à população total era de 47,1%. Na comparação com as outras cidades do estado, ocupava as posições 4º e 23º de 645, respectivamente. Já na comparação com municípios do país, ficava na posição, respectivamente, 17º e 79º de 5570. Tinha 31.6% da população em domicílios com rendimentos mensais de até meio salário mínimo por pessoa, o que o colocava na posição 305º de 645 dentre os municípios do estado e na 4372º posição de 5570 dentre os municípios do Brasil.

O setor censitário corresponde a menor unidade territorial, formada por área contínua contida, integralmente, em área urbana ou rural tem dimensão adequada à operação de pesquisas (IBGE, 2011). Por essa razão, arquivos com dados agregados por setor censitário serão usados como amostras para o trabalho. Além das variáveis de identificação geográfica – Grandes Regiões, Unidades da Federação, Mesorregião, Microrregião, Região Metropolitana ou RIDE, Município, Distrito, Subdistrito, Bairro, Setor, Situação do Setor e Tipo do Setor – das quais os setores interessam para este trabalho, há informações distribuídas em planilhas, que abrangem as seguintes características da população residente: sexo, idade, cor ou raça, condição no domicílio; pessoas responsáveis pelo domicílio; alfabetização; registro de nascimento das crianças de até 10 anos de idade; e características dos domicílios particulares (IBGE, 2011).

**Material e Métodos**

Agregados por Setores Censitários (Instituto Brasileiro de Geográfia e Estatística, 2010).

Todos os trabalhos deverão ser executados de forma que os(as) alunos(as) utilizem as ferramentas e os conhecimentos adquiridos ao longo do curso.

Neste tópico deve ser apresentada a descrição detalhada de todo o material utilizado e todos os métodos utilizados na obtenção de dados e análise dos resultados da pesquisa, de forma a permitir fácil compreensão e interpretação dos resultados, bem como a reprodução do estudo ou a utilização do método por outros profissionais. No entanto, apenas novos procedimentos devem ser descritos detalhadamente; procedimentos previamente publicados (descritos na literatura) devem ser citados e referenciados, mencionando de modo breve quaisquer modificações significativas efetuadas.

Na descrição do local da pesquisa não deve ser mencionado o nome da propriedade ou instituição ou empresa onde foi realizado o estudo, ou onde foram coletados os dados, bem como nome do(s) proprietário(s). O local deve ser identificado apenas pela localização geográfica (coordenadas) ou nome da cidade e estado. A população estudada, técnicas e métodos de amostragem, espécie animal ou vegetal utilizada, com a pertinente classificação zoológica ou botânica, técnicas agronômicas ou zootécnicas e outras, além da descrição dos procedimentos analíticos, devem ser detalhados na sequência lógica em que o trabalho foi conduzido.

A área de estudo deste trabalho será o Município de Belo Horizonte. Os dados publicados pelo IBGE estão organizados por xxxxxx Distritos Administrativos [DA] ou Setores Censitários e foram coletados da ~~Secretaria de Desenvolvimento Urbano do Município de São Paulo~~.

Para identificar as variáveis que melhor expressam a vulnerabilidade social ~~será~~ utilizada uma técnica estatística chamada de Análise Fatorial por Componentes Principais. A intenção em usar essa técnica é procurar identificar uma quantidade relativamente pequena de fatores que representam o comportamento conjunto de variáveis originais interdependentes. Segundo Fávero e Belfiore (2017), as técnicas de analise fatorial são uteis quando há a intenção de se trabalhar com variáveis que apresentam, entre si, coeficientes de correlação relativamente elevados e, concomitantemente, se deseja estabelecer novas variáveis que captem o comportamento conjunto das variáveis originais.

Os grupos de vulnerabilidade serão classificados a partir da combinação da dimensão de privação socioeconômica e da dimensão demográfica anteriormente descrita. Tal combinação será feita a partir da aplicação de outra técnica estatística denominada análise de agrupamentos, que permite agrupar os setores censitários que possuem características sociais comuns.

Para o agrupamento de grupos de vulnerabilidade será utilizada uma segunda técnica estatística chamada de análise de agrupamentos. Esta análise representa um conjunto de técnicas exploratórias muito úteis e que podem ser aplicadas quando há a intenção de se verificar a existência de comportamentos semelhantes entre observações que, neste caso, serão os setores censitários do Município de São Paulo, em relação a determinadas variáveis e o objetivo de se criarem grupos em que prevaleça a homogeneidade interna. Nesse sentido, esse conjunto de técnicas tem por objetivo principal a alocação de observações em uma quantidade relativamente pequena de agrupamento das observações (Fávero e Belfiore, 2017). A ideia é que os setores censitários de determinado grupo sejam relativamente semelhantes entre si e consideravelmente diferentes dos setores censitários de outros grupos. Os métodos de agrupamento a serem comparados serão o Latent Class Growth Modelling [LCGM] e k-means.

Para avaliar o desempenho relativo dos modelos de agrupamento na identificação de vulnerabilidade social será conduzida uma regressão logística multinomial [MLR] para examinar empiricamente como um conjunto de preditores teoricamente associados à vulnerabilidade social explica cada grupo. Diferentemente da tradicional técnica de regressão estimada por meio de métodos de mínimos quadrados, as técnicas de regressão logística são utilizadas quando o fenômeno a ser estudado apresenta-se de forma qualitativa (Fávero e Belfiore, 2017). A intenção é estimar a probabilidade de ocorrência de fenômenos qualitativos e usar o R² e as estatísticas de ajuste do modelo desta análise para informar qual das soluções de cluster resume melhor a variação na concentração de vulnerabilidade social

**Resultados Preliminares**

O título da seção Resultados Preliminares deve ser alinhado à esquerda, grafado em negrito com as primeiras letras das palavras em letras maiúsculas. É permitido que a seção seja dividida em subtópicos com formatação de acordo com a descrição no item 1.1. Nesta seção devem ser apresentados os resultados parciais da pesquisa, ou seja, os resultados obtidos até o momento.

**Conclusão(ões) ou Considerações Finais**

Tópico obrigatório para o depósito do TCC, porém opcional para a etapa dos Resultados preliminares.

**Agradecimentos** (opcional, 1 parágrafo sucinto)

Trata-se de seção opcional, de no máximo três linhas, na qual o autor agradece aqueles que contribuíram de maneira relevante para o desenvolvimento do trabalho e elaboração do TCC, mas que não tiveram o envolvimento intelectual necessário à atribuição de coautoria do mesmo, abstendo-se totalmente da menção ou citação de nomes de empresas, instituições ou pessoas que permitiram ou contribuíram com o desenvolvimento do trabalho, a menos que esteja documentalmente autorizado a fazê-lo.

**Referências**

Neste tópico deverá conter todas as referências dos trabalhos citados no texto e formatadas seguindo rigorosamente as normas do MBA USP ESALQ. Para mais informações, vide o manual de “Normas para Elaboração do Trabalho de Conclusão de Curso” disponível no sistema TCC.

**Apêndice ou Anexo** (opcional)

Apêndices são textos e/ou documentos que foram elaborados pelo autor e que são importantes para complementar a argumentação do trabalho. Anexos são textos ou documentos que ilustram, mas que não foram elaborados pelos autores. Apêndices deverão seguir as mesmas normas de formatação do restante do texto, inclusive para figuras e tabelas.

O TCC deverá conter no máximo 30 páginas, incluindo o(s) Apêndice(s) e/ou Anexo(s).