**Projeto de Pesquisa e Planejamento de Atividades**

|  |  |  |
| --- | --- | --- |
| **Aluno**: Renato Godoi da Cruz | | **Data início curso**: 11/05/2021 |
| **Orientador**: Auberth Henrik Venson | | **Defesa em:**  12/2022 |
| **Curso**: MBA Data Science e Analytics | **Modalidade**: Distância | Turma: 211 |

1. **Título do projeto**

**Comparação de algoritmos de agrupamento espacial: um estudo de caso sobre a vulnerabilidade social da população da cidade de São Paulo**

1. **Introdução**

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).

1. **Objetivo**

O objetivo deste trabalho é aplicar técnicas estatísticas a fim de melhor caracterizar a vulnerabilidade social de setores censitários na cidade de São Paulo. Com o objetivo de descrever a vulnerabilidade social da população da cidade de São Paulo da melhor maneira, este estudo pretende comparar técnicas de agrupamento de dados do Censo para identificar grupos de setores censitários que seguiram trajetórias semelhantes de vulnerabilidade social em 2010.

Este trabalho se baseia em trabalhos anteriores definindo trajetórias de pobreza usando pontos ao longo tempo – *A comparison of two methods for classifying trajectories: a case study on neighbourhood poverty at the intrametropolitan level in Montreal* – e de mapeamento de vulnerabilidade social da população do Município de São Paulo – *O Mapa da Vulnerabilidade Social da População da Cidade de São Paulo*.

Sendo assim, o objetivo é encontrar a abordagem mais precisa para determinar grupos de setores censitários caracterizados por características semelhante da vulnerabilidade da população, com cada grupo sendo mais diferente entre si.

.

1. **Material e Métodos**

A área de estudo deste trabalho será o Município de São Paulo. Os dados publicados pelo IBGE estão organizados por 18.952 Distritos Administrativos [DA] ou Setores Censitários e serão coletados na 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.

1. **Resultados Esperados**

Com este trabalho, espera-se contribuir para o estudo de aplicações de técnicas estatísticas para o agrupamento de dados censitários a fim de verificar a vulnerabilidade social de determinada população.

Além do exercício estatístico, espera-se, com este trabalho, contribuir com a caracterização da vulnerabilidade social a partir de múltiplas dimensões; realizar o estudo a partir de uma unidade de análise bastante desagregada, o setor censitário, o que pode propiciar ao planejamento de intervenções locais que levem em consideração o contexto social em que residem as famílias; identificar possíveis locais prioritários para instalação de novos equipamentos a partir de critérios objetivos e produzir um mapeamento que possa servir de referência para a produção de estudos qualitativos. Espera-se, por fim, que este trabalho traga um olhar detalhado e qualificado sobre as distintas realidades locais ao longo do tempo, o que só é possível por meio de análises qualitativas, mais próximas de cada caso.

1. **Cronograma de Atividades**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Atividades planejadas** | **Mês** | | | | | | | | | |
| **mar** | **abr** | **mai** | **jun** | **jul** | **ago** | **set** | **out** | **nov** | **dez** |
| Pesquisa bibliográfica | **x** | **x** |  |  |  |  |  |  |  |  |
| Elaboração do projeto de pesquisa | **x** | **x** |  |  |  |  |  |  |  |  |
| Coleta dos bancos de dados do IBGE |  | **x** |  |  |  |  |  |  |  |  |
| Modelagem da técnica de PCA |  | **x** | **x** |  |  |  |  |  |  |  |
| Modelagem das técnicas de agrupamento |  |  | **x** | **x** |  |  |  |  |  |  |
| Modelagem da técnica de regressão logística multinomial |  |  |  | **x** | **x** |  |  |  |  |  |
| Análise dos resultados |  |  |  |  | **x** | **x** |  |  |  |  |
| Redação do TCC | **x** | **x** | **x** | **x** | **x** | **x** | **x** |  |  |  |
| Elaboração de artigo para publicação em revista especializada |  |  |  |  |  |  | **x** | **x** | **x** |  |
| Elaboração da apresentação da defesa |  |  |  |  |  |  |  |  | **x** |  |
| Defesa |  |  |  |  |  |  |  |  |  | **x** |
| Ajustes da monografia |  |  |  |  |  |  |  |  |  | **x** |
| Submissão do artigo em revista especializada |  |  |  |  |  |  |  |  |  | **x** |

Projeto de Pesquisa; Resultados Preliminares; Entrega do Trabalho de Conclusão de Curso; Entrega da Apresentação da Defesa.

Fonte: o autor (2022).

1. **Referências Bibliográficas**

APPARICIO, P.; RIVA, M.; SÉGUIN, A.-M. 2015. A comparison of two methods for classifying trajectories: a case study on neighbourhood poverty at the intrametropolitan level in Montreal. Cybergeo: European Journal of Geography.

CENTRO DE ESTUDOS DA METRÓPOLE. 2004. O Mapa da Vulnerabilidade Social da População da Cidade de São Paulo. Centro de Estudos da Metrópole. São Paulo, p. 01-115.

COUTO, B. R. et al. O sistema único de assistência social no Brasil: Uma realidade em movimento. 1ª edição. ed. São Paulo: Cortez Editora, 2010.

FÁVERO, L. P.; BELFIORE, P. 2017. Manual de Análise de Dados - Estatística e Modelagem Multivariada com Excel, SPSS e Stata. 1ª edição. Elsevier Editora Ltda, Rio de Janeiro, Rio de Janeiro, Brasil.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA [IBGE]. 2011. Base de informações do Censo Demográfico 2010: Resultados do Universo por setor censitário. Centro de Documentação e Disseminação de Informações do Instituto Brasileiro de Geográfia e Estatística. Rio de Janeiro, p. 125.

INSTITUTO BRASILEIRO DE GEOGRÁFIA E ESTATÍSTICA [IBGE]. 2022. Panorama Cidades. Cidades IBGE. Disponivel em: <https://cidades.ibge.gov.br/brasil/sp/sao-paulo/panorama>. Acesso em: 17 abril 2022.

SEMZEZEM, P.; ALVES, J. D. M. 2013. Vulnerabilidade social, abordagem territorial e proteção na política de assistência social. Serv. Soc. Rev. v. 16, p. p. 143-166