**Comparação de algoritmos de agrupamento espacial da 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...”).

Renato Godoi da Cruz¹\*;Auberth Henrik Venson2

1 Universidade de São Paulo. Instituto Pecege. Av. Pádua Dias, 11 – Agronomia; 13418-900. Piracicaba, São Paulo, 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 da vulnerabilidade social da população de Belo Horizonte**

**Resumo**

Opcional para a etapa dos Resultados preliminares.

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

**Introdução**

Este trabalho consiste na análise da vulnerabilidade social da população da cidade de Belo Horizonte. Para realizar o estudo, serão utilizadas as informações do último censo disponíveis para os setores censitários do Município de Belo Horizonte. 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.

Do Censo de 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 Belo Horizonte, no ano de 2010, apresentava uma população de cerca de 2.375.151 habitantes distribuídos por um território de 331,354 km². Apresentava 96,2% de domicílios com esgotamento sanitário adequado, 82,7% de domicílios urbanos em vias públicas com arborização e 44,2% 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 Belo Horizonte, 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**

A área de estudo deste trabalho será o Município de Belo Horizonte. Os dados publicados pelo IBGE estão organizados por Distritos Administrativos [DA] ou Setores Censitários e foram coletados do Censo Demográfico de 2010, resultados do universo, agregados por setores censitários.

Para identificar as variáveis que melhor expressam a vulnerabilidade social foi 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 Belo Horizonte, 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**

Inicialmente, foram escolhidas dezoito variáveis censitárias consideradas relevantes para a caracterização das múltiplas dimensões da privação buscando contemplar dimensões como escolaridade, condições de habitação, renda e estrutura etária. De modo a simplificar a análise, preservando a caracterização detalhada dos setores censitários, essas variáveis foram reduzidas, sempre visando à construção de indicadores que captassem a heterogeneidade de situações existentes no município. Para isso foi realizada uma análise fatorial das variáveis censitárias selecionadas, que consiste em uma técnica estatística de redução de variáveis – nesse caso, foi utilizado o método de componentes principais. Quando diferentes variáveis estão fortemente correlacionadas, como o caso das variáveis renda e educação, por exemplo, elas podem ser expressas por um único fator comum. Esse foi o procedimento utilizado para reduzir a quantidade de variáveis relevantes para a análise. Algumas das variáveis iniciais foram excluídas da análise por não apresentarem variações significativas entre os diferentes setores censitários e por não contribuírem significativamente para a explicação da variabilidade total. A exclusão de variáveis do modelo deveu-se à baixa comunalidade (menos de 0,40) e a pouca contribuição para o aumento da explicação da variância total do conjunto de dados.

A tabela 1 apresenta todas as variáveis consideradas e aquelas que foram selecionadas como as mais relevantes, que então compuseram duas dimensões explicativas principais ou, em termos estatísticos, dois fatores. Esses dois fatores explicam 76,70% da variabilidade total.

Tabela 1. Variáveis usadas na análise fatorial

É interessante notar, em primeiro lugar, que as variáveis relativas às condições gerais de habitação, com exceção do número de habitantes por domicílio, não foram selecionadas como variáveis relevantes. Isso significa que essas variáveis não apresentam variações significativas entre os diversos setores censitários do Município de Belo Horizonte. Esse conjunto de variáveis refere-se basicamente às condições de saneamento dos setores censitários, como cobertura de água, coleta de lixo, presença de sanitários e outras. Isso aponta para o fator da ampliação da cobertura estatal desses serviços, especialmente a partir da década de 1980, de modo que hoje a cobertura é quase universal em municípios como Belo Horizonte. Ao contrário do que ocorriam em décadas passadas, essas variáveis podem não ser mais tão relevantes na caracterização de populações segregadas, não podendo ser utilizadas como um indicador de geral de precariedade.

Além disso, nota-se a partir da observação da Tabela 1 que, em vez de se obter uma única dimensão explicativa da pobreza urbana, as variáveis selecionadas na análise fatorial foram divididas em dois fatores, o que reforça o argumento da heterogeneidade da pobreza.

**Conclusão(ões) ou Considerações Finais**

Opcional para a etapa dos Resultados preliminares.

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

Seção opcional nesta etapa.

**Referências**

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

**Apêndice**

Não se aplica nesta etapa.

**Anexo**

|  |  |
| --- | --- |
| **VARIABLE** | **MEANING** |
| **“Education” Dimension** | |
| **P\_GRAD** | Percentage of people for whose kind of the highest completed degree was higher education |
| **P\_MEST** | Percentage of people for whose kind of the highest completed degree was master |
| **P\_DOUTOR** | Percentage of people whose kind of the highest completed degree was doctorate |
| **P\_SEM\_INST** | Percentage of people whose level of education is unschooled or incomplete Primary school |
| **P\_FUND** | Percentage of people whose level of education is complete primary school and incomplete high school |
| **P\_ENSMED** | Percentage of people whose level of education is complete high school and incomplete higher education |
| **P\_ENSSUP** | Percentage of people whose level of education is complete higher education |
| **“Mobility” Dimension** | |
| **P\_OUTROMUNC** | Percentage of people working in another municipality |
| **P\_CASADIA** | Percentage of people returning home from work daily |
| **P\_ATE5** | Percentage of people whose usual time spent commuting from home to work is up to 5 minutes |
| **P\_6A30** | percentage of people whose usual time spent commuting from home to work is up to 6 to 30 minutes |
| **P\_1A2** | percentage of people whose usual time spent commuting from home to work is 1-2 hours |
| **P\_MAISDE2** | percentage of people whose usual time spent commuting from home to work is more than 2 hours |
| **“Poverty” Dimension** | |
| **MEDIA\_DENSMORA** | Resident density per room |
| **P\_POBREZA** | % of people in poverty line: whose per capita household income per month is less than or equal to R$ 255.00 or US$144.89 (half minimum wage in 2010) |
| **P\_PPI\_POBREZA** | % of people in the poverty line and race, black, brown or indigenous |
| **P\_BOLSA\_FAM** | percentage of people who in July 2010 had a regular monthly income from the *Bolsa Família* Social Program or the Child Labor Eradication Program (PETI) |
| **P\_OUTROSPROG** | percentage of people who in July 2010 had regular monthly income from other social programs or transfers |
| **“Material deprivation” Dimension** | |
| **P\_ALVSREV** | Percentage of homes with uncoated masonry |
| **P\_REDE\_ESG** | Percentage of households with general sewerage |
| **P\_REDE\_AGUA** | Percentage of households with general water distribution network |
| **P\_LIXO** | Percentage of households with garbage collected directly by cleaning service |
| **P\_ENERGIA** | Percentage of households with electricity from electricity distribution company |
| **P\_TV** | Percentage of households with TV |
| **P\_MAQLAV** | Percentage of households with washing machine |
| **P\_GELADEIRA** | Percentage of households with refrigerator |
| **P\_MAQTVGEL** | Percentage of households with washing machine, TV and refrigerator |
| **P\_CELULAR** | Percentage of households with cell phones |
| **P\_COMP\_INT** | Percentage of households with computer with internet access |
| **P\_CELCOMPINT** | Percentage of households with mobile phone and internet computer |
| **P\_MOTO** | Percentage of households with motorcycle for private use |
| **P\_CARRO** | Percentage of households with private car |
| **P\_ADEQ** | Percentage of households with adequate housing |
| **P\_TUDOADEQ** | Percentage of households with access to sewerage, water supply, garbage collection, electricity and adequate housing |
| **P\_NEM\_MOTO\_CARRO** | Percentage of households without motorcycles or cars ownerships for private use |
| **P\_SO\_MOTO** | Percentage of households with only motorcycles ownership for private use |
| **P\_SO\_CARRO** | Percentage of households with only private car ownership |
| **“Income” Dimension** | |
| **MED\_RENDDOM** | monthly household income in July 2010, in Brazilian *Reais* |
| **“Wealth” Dimension** | |
| **P\_ALUG1000** | percentage of rented households with rental value of R$1,000.00 (US$ 568.20) or more |
| **P\_BANH4OUMAIS** | Percentage of households with 4 or more bathrooms |
| **P\_IDOSO10SM** | % of people aged 65 years and over with a monthly income equal to or above R$ 5,100.00 (US$ 2,897.72 or 10 Brazilian minimum wages) |
| **“Segregation” Dimension** | |
| **ICE\_RENDA** | (number of people with income above R$ 5,400.00 - number of people with income below R$ 1,000.00) / number of respondents [figures were calculated based on the 20 and 80 percentiles of income V6529 in the PERSON spreadsheet 2010 Census microdata] |
| **ICE\_EDU** | (Number of persons with completed higher education - Number of persons without education and incomplete elementary school)/Total respondents [V6400] |
| **ICE\_RENDA\_PRETO** | (number of whites with income over R$ 5,400.00 - number of blacks with income equal to or less than R$ 1,000.00) / total number of people who answered both questions [V6529 and V0606] |
| **ICE\_RENDA\_PPI** | (number of whites with income over R$ 5,400.00 - number of black + brown + indigenous with income equal to or less than R$ 1,000.00) / total number of people who answered both questions [V6529 and V0606] |
| **ICE\_BRANCO\_RENDA** | (number of whites with income over R$ 5,400.00 - number of whites with income equal to or less than R$ 1,000.00) / total number of people who answered both questions [V6529 and V0606] |