

AVALIAÇÃO DE MÉTODOS PARA CONSTRUÇÃO DE UM ÍNDICE DE VULNERABILIDADE DE SAÚDE PARA LONDRINA-PR

EVALUATION METHODS FOR AN HEALTH VULNERABILITY INDEX ELABORATION IN LONDRINA CITY (PARANÁ STATE - BRAZIL)

Thamy Barbara Gioia

Universidade Estadual de Londrina

thamygioia@gmail.com

Adriana C. de Freitas Pereira

Universidade Estadual de Londrina

adricfp@gmail.com

Jaqueline Aparecida Raminelli

Docente do Departamento de Estatística

Universidade Estadual de Londrina

jaquest@gmail.com

RESUMO

Dentre os modelos de pesquisa interdisciplinares atuais, estão as proposições de indicadores em saúde como instrumento para o planejamento de ações. Alguns modelos propõem o estabelecimento de variáveis mais preditivas a partir da modelagem knowledge-driven enquanto outros buscam recursos em modelos data-driven. Neste sentido o objetivo desta pesquisa foi comparar dois modelos de Índice de Vulnerabilidade de Saúde – IVS para o Município de Londrina – PR, um em modelo knowledge-driven elaborado pela Secretaria de Saúde de Belo Horizonte e o segundo estabelecido por um modelo data-driven, através da estatística de análise fatorial. A metodologia de pesquisa utilizou dados do censo demográfico do IBGE (2010). No modelo knowledge-driven as variáveis mais preditivas foram às relacionadas ao saneamento básico enquanto que no modelo data-driven as variáveis mais preditivas foram raça, renda e analfabetismo. Tal condição influenciou na configuração espacial de classificação do índice. Setores distritais e rurais foram classificados em melhores condições no modelo data-driven quando comparados aos resultados do modelo knowledge-driven.

PALAVRAS-CHAVE: Índices de Saúde. Knowledge-driven. Data-driven. Análise Fatorial.

ABSTRACT

Various current interdisciplinary research models suggest using health indicators as an instrument for planning actions. Some models propose the establishment of more predictive variables based on knowledge-driven modeling, while others seek resources in data-driven models. Thus, the objective of this research was to compare two models of Health Vulnerability Index, for the city of Londrina, PR. One in a knowledge-driven model developed by the Belo Horizonte Health Department and the second established by a data-driven model, through factor analysis statistics. The research methodology used data from the IBGE demographic census (2010). In the knowledge-driven model, the most predictive variables were those related to basic sanitation. While in the data-driven model, the most predictive variables were race, income, and illiteracy. This condition influenced the spatial configuration of the index classification. District and rural sectors were classified with better conditions in the data-driven model than the results found in the knowledge-driven model.

KEYWORDS: Health Indexes. Knowledge-driven. Data-driven. Factor Analysis.

Recebido em: 26/03/2020

Aceito para publicação em: 09/11/2020.

INTRODUÇÃO

O entendimento de que variáveis ambientais e sociais influenciam nas condições de saúde de uma população não é algo recente. Em 1717 Lancisi elaborou um cartograma para identificação de áreas insalubres; em 1792 Finke estruturou seu “Ensaio de uma Geografia geral Médico-Prática”; em 1843 Budin escreveu uma obra intitulada “*Essai de Geographie Medicale*”; e em 1854 Snow apresentou um dos estudos mais conhecidos nesta área da ciência, a relação dos casos de cólera e o poço de abastecimento de água em Londres (BOUSQUAT; COHN, 2004; NOGUEIRA; REMOALDO, 2010; SANTANA, 2014; IÑIGUEZ; BARCELLOS, 2014).

Atualmente, a construção de indicadores que contemplem variáveis sociais e ambientais tem sido utilizada para planejamento de políticas públicas tendo em vista que as condições de renda, moradia, alimentação, emprego e educação são consideradas determinantes sociais em saúde e fazem parte de um grupo de fatores que influenciam na qualidade da saúde de uma população (BUSS; PELLEGRINI FILHO, 2007). Ainda, quando a síntese desses indicadores é apresentada através de mapas, utilizando para isso softwares de Sistemas de Informações Geográficas – SIG, é possível visualizar os resultados de distribuição no espaço, facilitando as associações de eventos em saúde (BRASIL, 2004).

A modelagem espacial em SIG envolve a combinação de dados geográficos a fim de descrever e analisar possíveis integrações de variáveis para inferência geográfica e proposição de áreas vulneráveis, potenciais e adequadas, bem como na construção de índices que possam propiciar o suporte necessário à tomada de decisões. Tal procedimento pode considerar dois tipos de abordagem: o *knowledge-driven* (abordagem guiada pelo conhecimento) e o *data-driven* (abordagem guiada pelos dados) (MEIRELES; CAMARA; ALMEIDA, 2007).

Um método de formulação de índice no modelo *knowledge-driven* é o Índice de Vulnerabilidade de Saúde - IVS elaborado pela Secretaria de Saúde de Belo Horizonte - SS-BH com o objetivo de estabelecer áreas prioritárias de intervenção e alocação de recursos destinados ao setor de saúde do Município (PMBH, 2013). Neste modelo foram consideradas oito variáveis, sendo três de saneamento básico e cinco de caráter socioeconômico. A cada variável foi atribuído um peso determinado através da realização de entrevistas com técnicos de secretarias diversas do município e pesquisadores da academia.

Uma proposta alternativa para atribuição de pesos aos indicadores considera a utilização da análise fatorial, descrita por Manly (2008) como um conjunto básico de variáveis em termos de um número reduzido de índices ou fatores. Na prática, a análise fatorial permite chegar a cargas fatoriais específicas para cada variável, possibilitando inferir a ordem de participação (“peso”) de cada variável em um universo de dados, sendo, portanto, uma abordagem no modelo *data-driven*.

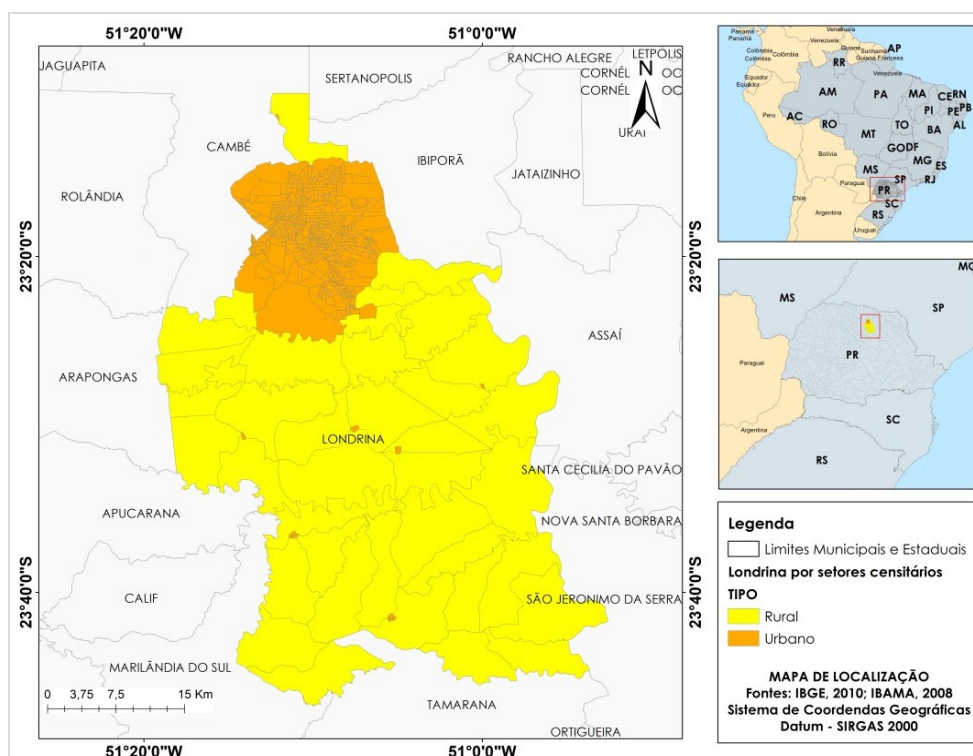
Com base no exposto, o objetivo desta pesquisa foi comparar dois métodos de construção de um IVS, o primeiro considerando a metodologia *knowledge-driven* proposta pela SS-BH e o segundo desenvolvido a partir das técnicas estatísticas de análise fatorial, na abordagem *data-driven*, partindo da hipótese de que a utilização de pesos atribuídos de forma qualitativa às variáveis constitutivas de um indicador podem subestimar resultados, negligenciando características de relevância em um município.

MATERIAIS E MÉTODOS

Área de estudo

Como área de estudo da pesquisa delimitou-se o município de Londrina, localizado na região norte do Estado do Paraná (Figura 1). De acordo com o censo demográfico do IBGE (2010), Londrina contava com 506.701 habitantes naquele ano, sendo que a maior parte residindo em área urbana, aproximadamente 97%. Para fins de coleta do censo demográfico, foram delimitados 717 setores censitários no município, sendo 678 em área urbana e 39 em área rural (IBGE, 2010).

Figura 1 – Mapa de localização da área de estudo



Fonte: Elaborado pelas autoras.

Dados utilizados e pré-processamento

O IVS que aqui será identificado como IVS original (elaborado pela Secretaria de Saúde de Belo Horizonte – MG e aplicado à área de estudo da presente pesquisa) parte da integração de oito variáveis socioeconômicas formadas a partir de 20 subvariáveis disponibilizadas no banco de dados do censo demográfico do IBGE (2010) (Quadro 1).

Quadro 1 - Variáveis e subvariáveis adotadas para composição do IVS original – SS-BH

Variáveis	Subvariáveis
V1 – Razão de moradores por domicílio	V ₀₂ – Planilha 1 – Total de domicílios V ₀₂ – Planilha 2 – Total de moradores
V2 - Percentual de domicílios particulares permanentes com abastecimento de água inadequado ou ausente	Arquivo domicílios V ₀₁₅ – Outra forma de abastecimento de água
V3 - Percentual de domicílios particulares permanentes com esgotamento sanitário inadequado ou ausente	Arquivo domicílios V ₀₁₉ – Esgotamento via fossa rudimentar V ₀₂₀ – Esgotamento via vala V ₀₂₁ – Esgotamento via rio, lago ou mar V ₀₂₂ – Outros tipos de esgotamento
V4 - Percentual de domicílios particulares permanentes com destino do lixo de forma inadequada ou ausente	Arquivo domicílios V ₀₃₈ – Lixo queimado na própria Propriedade; V ₀₃₉ – Lixo enterrado na propriedade V ₀₄₀ – Lixo jogado em terreno baldio ou logradouro V ₀₄₁ – Lixo jogado em rio, lago ou mar V ₀₄₂ – Outros destinos
V5 - Percentual de pessoas analfabetas	Arquivo alfabetização V ₀₁ - Pessoas alfabetizadas com 5 ou mais anos de idade

Continua

Conclusão

Variáveis	Subvariáveis
V6 - Percentual de domicílios particulares com rendimento per capita até ½ Salários mínimos	Arquivo renda V ₀₀₅ - Domicílios particulares permanentes com renda mensal de até 1/8 salário mínimo V ₀₀₆ - Domicílios particulares permanentes com renda mensal de 1/8 até 1/4 salário mínimo V ₀₀₇ - Domicílios particulares permanentes com renda mensal de 1/4 até 1/2 salário mínimo
V7 - Rendimento nominal mensal médio das pessoas responsáveis (invertido)	Arquivo pessoas responsáveis V ₀₂₂ – Total de rendimento médio mensal da pessoa responsável V ₀₂₀ – Responsável com ou sem rendimento
V8 - Percentual de pessoas de raça/cor preta, parda e indígena	Arquivo cor ou raça V ₀₀₃ – Total de negros V ₀₀₅ – Total de pardos V ₀₀₆ – Total de indígenas

Fonte: Elaborado pelas autoras.

No decorrer da organização preliminar das planilhas observaram-se inconsistências no banco de dados: setores com zero morador, mas com domicílios recenseados; setores com zero domicílio, mas moradores recenseados; setores cujo número total de analfabetos era maior que o número total de moradores e o mesmo para o número total de população nas raças negras, pardas e indígenas; de forma que, durante o procedimento de cálculo, os resultados de agregação nestes setores foram negativos. Devido a esta condição, optou-se pela exclusão de todos os setores que apresentassem ao menos um dado inconsistente em qualquer uma das oito variáveis analisadas, o que resultou na exclusão de 55 setores, de forma que as demais etapas metodológicas foram aplicadas ao total de 662 setores censitários.

Os valores correspondentes as variáveis integrantes do IVS original da SS-BH foram então normalizadas para uma escala de 0 a 1 para “permitir a comparação e agregação dos indicadores em diferentes medidas [...]” (PMBH, 2013, p.9). Esta normalização foi executada para todas as variáveis dos 662 setores censitários através da Equação 1:

$$\text{Valor convertido} = \frac{\text{Valor bruto} - \text{valor mínimo}}{\text{valor máximo} - \text{valor mínimo}} \quad (1)$$

A planilha síntese constituída pelas variáveis descritas no Quadro 1 foi então concatenada à base cartográfica em formato vetorial *shapefile* dos setores censitários do município, também disponibilizada no banco de dados do censo demográfico do IBGE: http://geoftp.ibge.gov.br/organizacao_do_territorio/malhas_territoriais. Este procedimento foi realizado por meio de ferramentas disponíveis no *software* de SIG *ArcGIS* 9.3². Adicionalmente, todas as

² Licença disponível no laboratório IMAP&P - IMAGENS, PAISAGENS E PERSONAGENS. Departamento de Geociências do Centro de Ciências Exatas da Universidade Estadual de Londrina.

variáveis foram analisadas por meio de suas estatísticas descritivas a fim de identificar os valores médios, mínimos e máximos e o desvio padrão.

Modelagem: proposta original e proposta alternativa por meio da análise fatorial

Essa pesquisa, do tipo descritiva e de abordagem quantitativa, considerou duas metodologias de modelagem espacial: *knowledge-driven* (original) e *data-driven* (alternativa), onde foram testadas e comparadas duas propostas de IVS. A proposta original, que utilizou o modelo de IVS proposto pela SS-BH, descrita a seguir no subcapítulo 2.3.1. e a proposta alternativa, denominada de IVS alternativo, calculada por meio da estatística de análise fatorial, descrita no subcapítulo 2.3.2.

IVS original

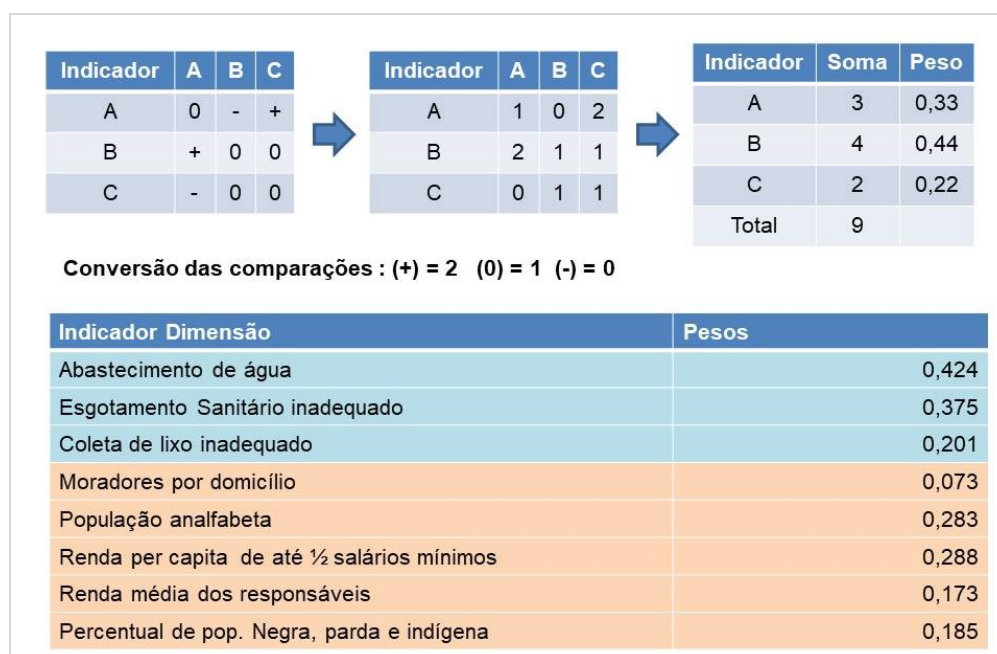
O primeiro modelo apresentado por esta pesquisa é o IVS construído pela SS-BH, elaborado em 1998 utilizando como base de dados o censo demográfico do IBGE de 1991, tendo sido recalculado posteriormente por aquela Secretaria com os censos de 2000 e 2010 devido ao incremento de variáveis e ao incremento de setores censitários. Em todas as atualizações as variáveis constituintes do índice foram delimitadas a partir do consenso entre técnicos da Secretaria de Saúde daquele município (PMBH, 2013).

A atualização a partir do censo de 2010 agregou oito variáveis (Quadro 1) ponderadas por pesos específicos que foram definidos a partir de um modelo *knowledge-driven* realizado pela SS-BH. Dezesesseis especialistas de áreas diversas foram convidados a opinarem com relação ao nível de importância de cada variável comparada a outra quando considerado a temática saúde.

Na matriz, o indicador disposto na linha era comparado a outro disposto na coluna. Quando marcado o sinal de positivo (+) o indicador linha era considerado mais importante do que o apresentado na coluna; se atribuído o sinal negativo (-) o indicador tinha menor importância, e se atribuído 0, igual importância.

Em seguida as comparações foram convertidas de (+) para 2, (-) para 0 e (0) para 1, os valores somados foram então distribuídos proporcionalmente constituindo-se os pesos às variáveis. O modelo de matriz utilizada para constituição dos pesos é apresentado na Figura 2.

Figura 2 – Modelo de atribuição de pesos do IVS original - SS-BH



Fonte: PMBH, 2013. Adaptado pelas autoras.

O IVS original da SS-BH foi então estabelecido e pode ser descrito pela Equação 2:

$$\text{IVS setor} = V1*P1 + V2*P2 + V3*P3 + Vn*pn \dots \quad (2)$$

sendo $V1, V2, Vn$ os valores de cada variável normalizada a partir dos dados do censo demográfico do IBGE e p o valor do peso atribuído a cada variável, conforme apresentando na Figura 2. Para obtenção dos resultados do IVS para o município de Londrina, por meio do método IVS original, optou-se por desenvolver o cálculo diretamente na planilha de dados concatenadas a base cartográfica dos setores censitários.

Proposta de IVS alternativo: análise fatorial

A proposta alternativa trabalha com análise fatorial na estimativa do IVS para o município de Londrina (PR), com base nas mesmas variáveis normalizadas do IVS original. Conforme descrito por Manly (2008), o modelo matemático de análise fatorial é dado pela Equação 3:

$$X_i = \alpha_{i1}F_1 + \alpha_{i2}F_2 + \alpha_{i3}F_3 + \dots + \alpha_{ij}F_j \quad (3)$$

X_i são as variáveis padronizadas; α_i são as cargas fatoriais que medem o grau de correlação entre as variáveis originais e os fatores, que por sua vez são estimados por uma combinação linear das variáveis originais, seguindo a Equação 4:

$$F_j = \omega_{j1}X_1 + \omega_{j2}X_2 + \omega_{j3}X_3 + \dots + \omega_{ji}X_i \quad (4)$$

sendo F_j os fatores comuns não relacionados (independentes) e ω_{ji} os coeficientes dos escores fatoriais, que resulta da multiplicação dos coeficientes (ω_{ji}) pelas variáveis originais; e X_i são as próprias variáveis originais.

Todos os procedimentos de análise fatorial foram realizados no *software R* versão 3.5.1 por meio do pacote *FactoMineR* (HUSSON; JOSSE; LÉ, 2018) e consistiram na execução das seguintes etapas:

1. Concepção e análise da matriz de correlação linear sob as variáveis originais;
2. Concepção de fatores por meio do método de estimação de máxima verossimilhança;
3. Fixação do número de fatores respondentes considerando autovalores maiores ou iguais a 1 (um);
4. Estimação de cargas fatoriais através do método de componentes principais;
5. Rotação de fatores para melhora da interpretação através do método oblíquo *promax*;
6. Estabelecimento dos modelos de cálculos a partir da equação de Manly (2008).

A exclusão dos 55 setores censitários não inviabilizou a modelagem em análise fatorial. Conforme orientações de HAIR JR *et al* (2009) recomenda-se a observação de pelo menos 50 amostras para aplicação de uma análise fatorial robusta, sendo mais indicada observações amostrais iguais ou superiores a 100.

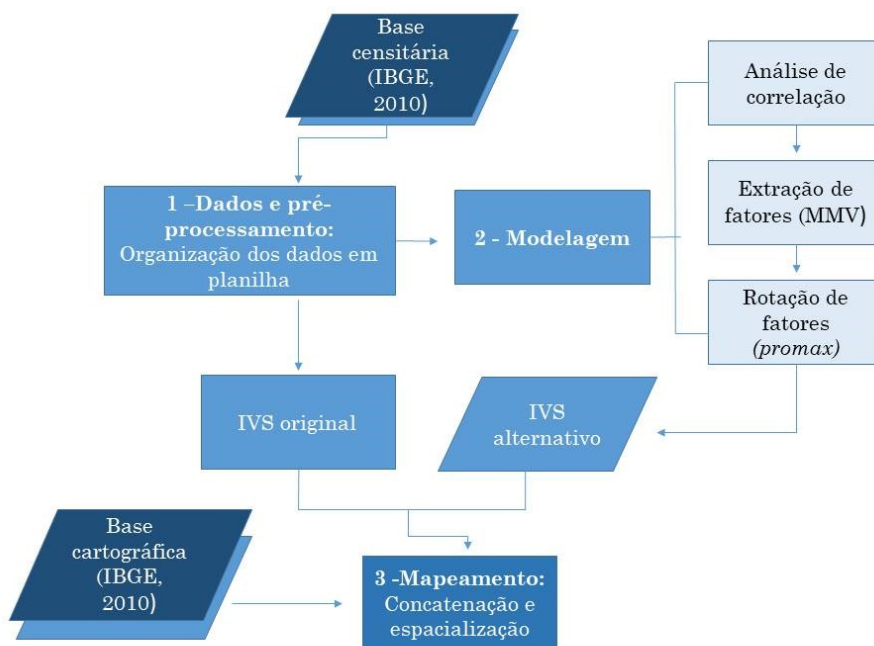
Metodologia para mapeamento dos resultados

Para mapeamento dos resultados, a planilha do IVS alternativo foi concatenada a base cartográfica vetorial de setores censitários com a agregação resultante do IVS original. Este procedimento foi realizado por meio da ferramenta *join*, disponível no *software ArcGIS* versão 9.3. Os resultados foram analisados conforme categorização proposta na própria metodologia da SS-BH (PMBH, 2013):

- **Médio risco:** setores censitários com valores do IVS com $\frac{1}{2}$ desvio padrão (DP) em torno da média (média +/- 0,5DP);
- **Baixo Risco:** setores com valores IVS inferiores ao IVS Médio;
- **Elevado Risco:** setores com valores acima do IVS médio até o limite de 1,5 desvio padrão acima da média (limite superior do IVS médio + 1 DP);
- **Muito elevado risco:** setores com valores acima do IVS elevado;

A síntese dos procedimentos metodológicos utilizados na pesquisa pode ser observada no fluxograma da Figura 3.

Figura 3 – Fluxograma metodológico



Fonte: Elaborado pelas autoras.

Considerações sobre os métodos avaliados

É possível observar que a metodologia proposta pela SS-BH (2013) demanda menos etapas de cálculos e avaliações estatísticas do que a proposta da análise fatorial, o que favorece o processo de execução. No entanto, ao partir de avaliações de especialistas para definir os pesos atribuídos às variáveis, incorre-se ao risco de análises subjetivas, motivadas pelas experiências particulares dos técnicos envolvidos no processo. Nestas condições, variáveis socioeconômicas de relevância poderiam ser negligenciadas.

O método estabelecido pela análise fatorial, no entanto, fornece o que se poderia chamar de “cenário de análise preliminar”, já que os fatores reúnem grupos de variáveis que melhor explicam a variância dos dados do universo analisado. Ainda, para cada variável, é gerado o escore de classificação, ou seja, valor de representatividade de determinada variável em determinado fator, de forma que, os pesos mais significativos podem ser atribuídos às variáveis de deficiências mais significativas. No entanto, trata-se de um procedimento que demanda análises estatísticas robustas e o uso de softwares para modelagem, devido às inúmeras etapas de cálculos pertinentes ao método, o que poderia limitar o uso da metodologia a depender dos recursos técnicos e pessoais disponíveis.

Mesmo se tratando de realidades geográficas distintas, tanto do ponto de vista da configuração urbana e das condições socioeconômicas, considera-se pertinente a aplicação do modelo de IVS original de Belo Horizonte para o município de Londrina devido à potencialidade de aplicação do índice como instrumento de políticas públicas já consolidadas no município de referência.

Ressalta-se também que neste estudo não se pretende apresentar uma discussão profunda acerca dos conceitos de vulnerabilidade ou sobre a pertinência das variáveis propostas para construção do

IVS original da SS-BH. O foco principal da pesquisa está em comparar as duas abordagens: *knowledge-driven* e *data-driven* para construção do índice.

RESULTADOS E DISCUSSÕES

Na Tabela 1 apresenta-se uma síntese da estatística descritiva sobre os valores brutos das oito variáveis constituintes do IVS para Londrina. Observa-se que enquanto alguns dos setores apresentaram 0% de deficiência em saneamento básico, outros indicaram 100% de deficiência. Ainda, a variável V7 (Rendimento nominal mensal médio das pessoas responsáveis) indicou setores com renda média de R\$ 258,10 enquanto outros, uma renda média de R\$ 13.563,48 por responsável.

Tais resultados permitem inferir preliminarmente, a significativa diferenciação socioespacial quanto às condições socioeconômicas da população residente no município e consequentemente, as variações a serem encontradas quanto aos resultados do índice de vulnerabilidade.

Tabela 1 – Estatística descritiva para as oito variáveis brutas componentes do IVS para Londrina

Variável (unid.)	Mínimo	Máximo	Média	Desvio Padrão
V1 (Razão)	1,34	4	3,07	0,37
V2 (%)	0,00	100	1,15	6,32
V3 (%)	0,00	100	17,47	33,02
V4 (%)	0,00	100	5,59	20,58
V5 (%)	0,00	41,17	11,23	5,46
V6 (%)	0,00	71,17	13,12	11,96
V7 (Média em R\$)	258,10	13.563,48	1.718,28	1.523,92
V8 (%)	0	67,5	25,45	14,53

Fonte: Elaborado pelas autoras.

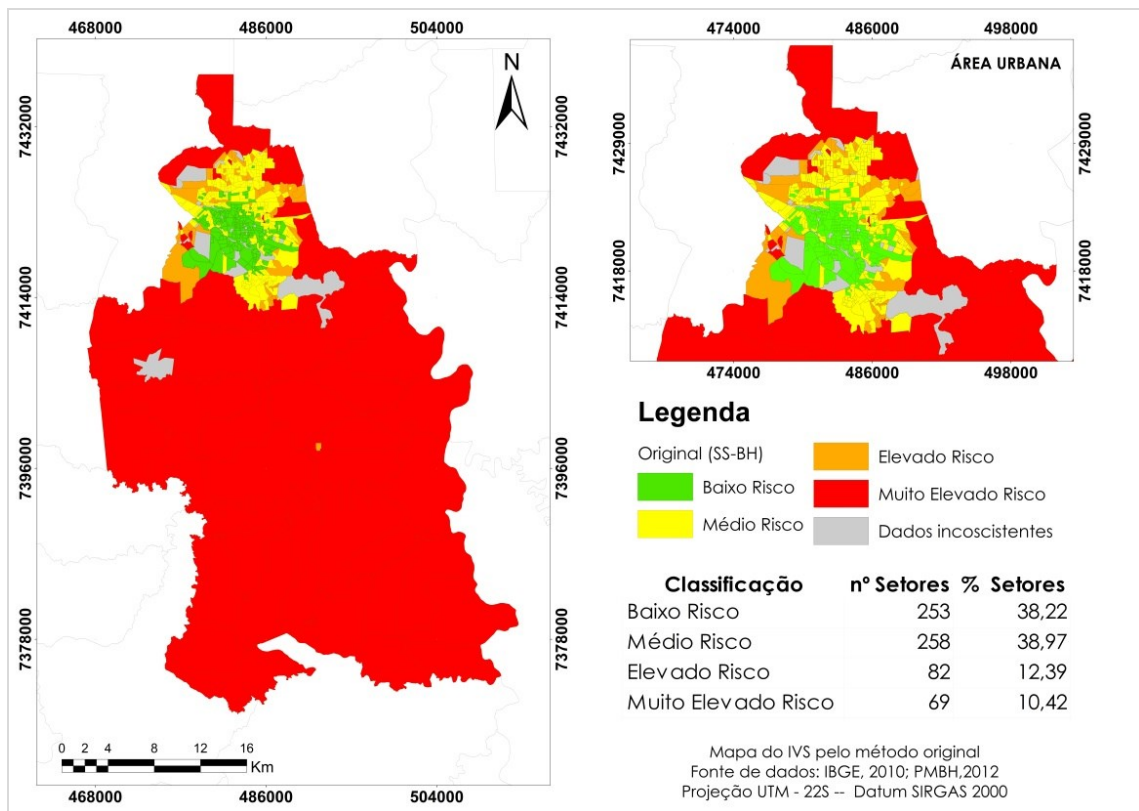
O estabelecimento do IVS original a partir do modelo de atribuição de pesos no método *knowledge-driven* (Figura 2) resultou na espacialização apresentada na Figura 4. Os setores classificados como baixo risco estão representados pelos setores em verde e totalizaram 253 setores, todos localizados em área urbana, concentrados na região centro-sudoeste e abrangendo, portanto, 38% dos setores do município.

Em amarelo apresentam-se os setores com médio risco, totalizando 258 setores nesta categoria, aproximadamente 39% dos setores. Em laranja categorizam-se os setores de elevado risco, localizados nas áreas mais periféricas do município, totalizando 82 setores e aproximadamente 12% do total; por último, em vermelho, os setores com muito elevado risco em saúde concentrados nas áreas mais periféricas da área urbana e praticamente em todos os setores da área rural do município, totalizaram 69 setores, cerca de 10% do total.

A proposição de uma avaliação alternativa a partir da análise fatorial exigiu a análise de algumas etapas de procedimento estatístico, a primeira delas referente à avaliação das correlações entre as variáveis. O número substancial de referência para correlações entre variáveis não é unanimidade entre os referenciais bibliográficos, no entanto, neste artigo, optou-se por considerar significantes, os valores acima de 0,30 (HAIR JR *et al*, 2009).

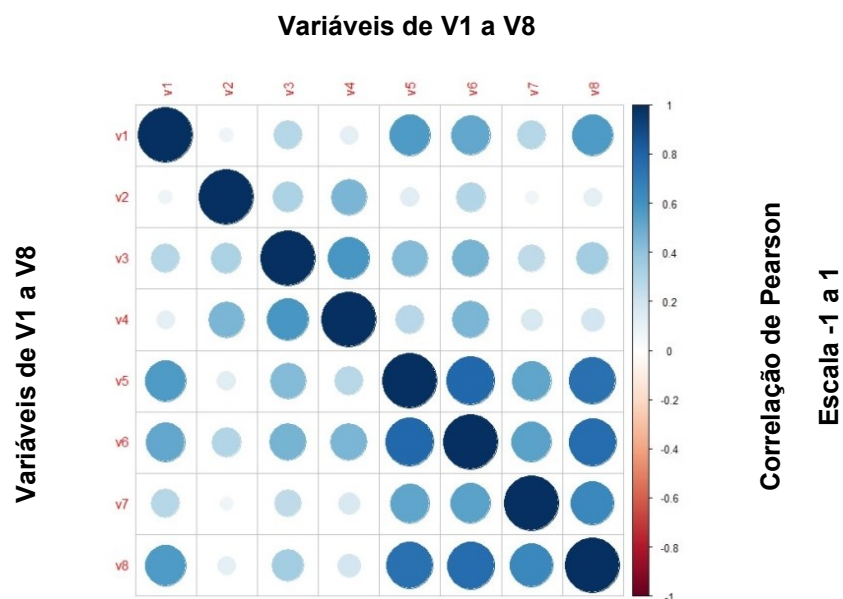
A matriz de correlação entre as variáveis originais correspondentes aos setores censitários de Londrina indicou fortes correlações lineares de Pearson (acima de 0,7) para algumas variáveis, entre elas: V5 (taxa de analfabetismo) e V6 (renda até meio salário mínimo), V5 (analfabetismo) e V8 (raça) e, V6 (renda até meio salário mínimo) e V8 (raça). A Figura 5 ilustra o resultado da matriz de correlação. Quanto mais intensa a correlação maior a área dos círculos; por outro lado, a coloração dos círculos representa a direção da relação linear; relações positivas (tons de azul) e, relações negativas ou inversamente proporcionais (tons de vermelho).

Figura 2 – Resultado do IVS a partir dos pesos originais



Fonte: Elaborado pelas autoras.

Figura 5 – Ilustração da correlação linear entre as variáveis originais.



Fonte: Elaborado pelas autoras.

Em seu artigo “Raízes sociolinguísticas do analfabetismo no Brasil”, Bortoni-Ricardo, *et al* (2008, p.220), afirmam que “o analfabetismo chega a ser 20 vezes maior nas famílias mais pobres. Nos domicílios com renda superior a dez salários mínimos, o índice é de apenas 1,4. Já naqueles cuja renda é inferior a um salário mínimo o índice alcança 29%”. Neste mesmo artigo os autores citam a forte correlação entre analfabetismo e raça, quando apresentam resultados do INEP que indicam haver três vezes mais brancos alfabetizados do que negros e pardos no Brasil.

O número de fatores utilizados no modelo de análise fatorial foi definido a partir do critério dos autovalores, resultado que indica quanto cada fator consegue explicar da variância total dos dados (CORRAR; PAULO; DIAS FILHO, 2012). Desta forma, dois autovalores maiores do que a unidade (um) sugeriram que dois fatores (Fator 1 e Fator 2) deveriam ser considerados para avaliação do universo de dados, ou seja, parte da variância para as variáveis originais está contida nos dois fatores comuns que explicam 60% do universo de dados, destes, 40% representados pelo Fator 1 e 20% pelo Fator 2.

Na Tabela 2 são apresentados os escores (cargas fatoriais) gerados para os dois grupos fatoriais.

Tabela 2 - Resultado do método de rotação *promax*. Escores ajustados para o Fator 1 e Fator 2

Variáveis	Cargas fatoriais	
	Fator 1	Fator 2
V1	0,656	
V2		0,495
V3	0,234	0,537
V4		0,970
V5	0,863	
V6	0,799	0,214
V7	0,671	
V8	0,940	-0,126

Fonte: Elaborado pelas autoras.

A configuração do Fator 1, que explica 40% do universo de dados, indica que altas cargas são observadas para as variáveis socioeconômicas, destacando-se as variáveis V8 - Percentual de pessoas de raça/cor preta, parda e indígena, seguida da variável V5 - Percentual de pessoas analfabetas e V6 - Percentual de domicílios particulares com rendimento per capita até ½ Salários mínimos. Já o Fator 2, que explica 20% dos dados, é fortemente representado pelas variáveis de saneamento básico, destacando o escore da variável V4 - Percentual de domicílios particulares permanentes com destino do lixo de forma inadequada ou ausente (Tabela 2).

Na Tabela 2 é possível observar ainda que as variáveis V2 e V4, por exemplo, não apresentam cargas (escores) para o Fator 1, isto significa que as cargas foram tão baixas que a composição do Fator 1 nem as representa. O mesmo ocorre para as variáveis V1, V5 e V7 para o Fator 2. A partir da rotação *promax* executada produziu-se os seguintes modelos de cálculo (Equações 5 e 6) para estabelecimento do IVS alternativo:

$$F1 = V1 * 0,656 + V3 * 0,234 + V5 * 0,863 + V6 * 0,799 + V7 * 0,671 + V8 * 0,940 \quad (5)$$

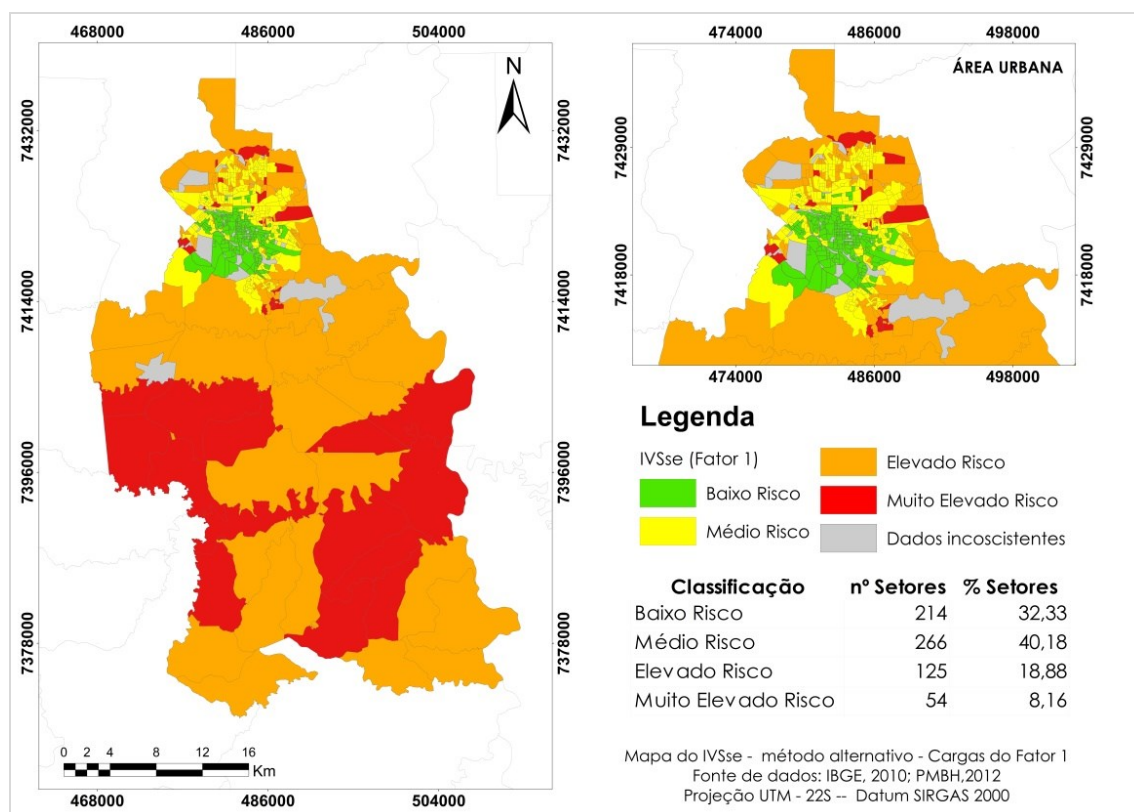
$$F2 = V2 * 0,495 + V3 * 0,537 + V4 * 0,970 + V6 * 0,214 - V8 * 0,126 \quad (6)$$

Considerando que o Fator 1 tem uma carga positiva alta para V8 (raça), V5 (taxa de analfabetismo) e V6 (renda de até meio salário mínimo) este pode ser renomeado como “Índice de Vulnerabilidade em Saúde Socioeconômico - IVSse. Por outro lado, o Fator 2 é quase inteiramente explicado pela variável V4 (resíduos sólidos), e pode ser rotulado como Índice de Vulnerabilidade em Saúde Saneamento Básico - IVSsb.

A princípio considerou-se a possibilidade de formulação de um único índice a partir da média ponderada dos pesos estabelecidos pelos fatores 1 e 2, ou a simplificação das fórmulas, retirando as menores cargas para variáveis atribuídas aos dois fatores (como no caso da V3, V6 e V8), no entanto, pesquisas realizadas em literatura especializada (MINGOTI, 2005; MANLY, 2008) indicaram não ser conveniente este procedimento, dado que os escores fatoriais complementam-se no que diz respeito a variância do universo de dados, neste caso, optou-se pela construção de dois mapas síntese de IVS alternativo considerando o modelo tradicional de cálculo a partir da equação de Manly (2008).

Na Figura 6 apresenta-se o IVSse e na Figura 7 o IVSsb. O IVSse indicou baixa vulnerabilidade em saúde para setores da área centro-sudoeste da área urbana (setores em verde), 214 setores foram classificados nesta categoria, representando 32% do total de setores. 266 setores foram classificados em médio risco (setores em amarelo), totalizando 40% dos setores, 125 setores foram classificados em elevado risco (setores em laranja), aproximadamente 19%, e em muito elevado risco, (setores em vermelho), foram mapeados 54 setores, 8% do total de setores do município.

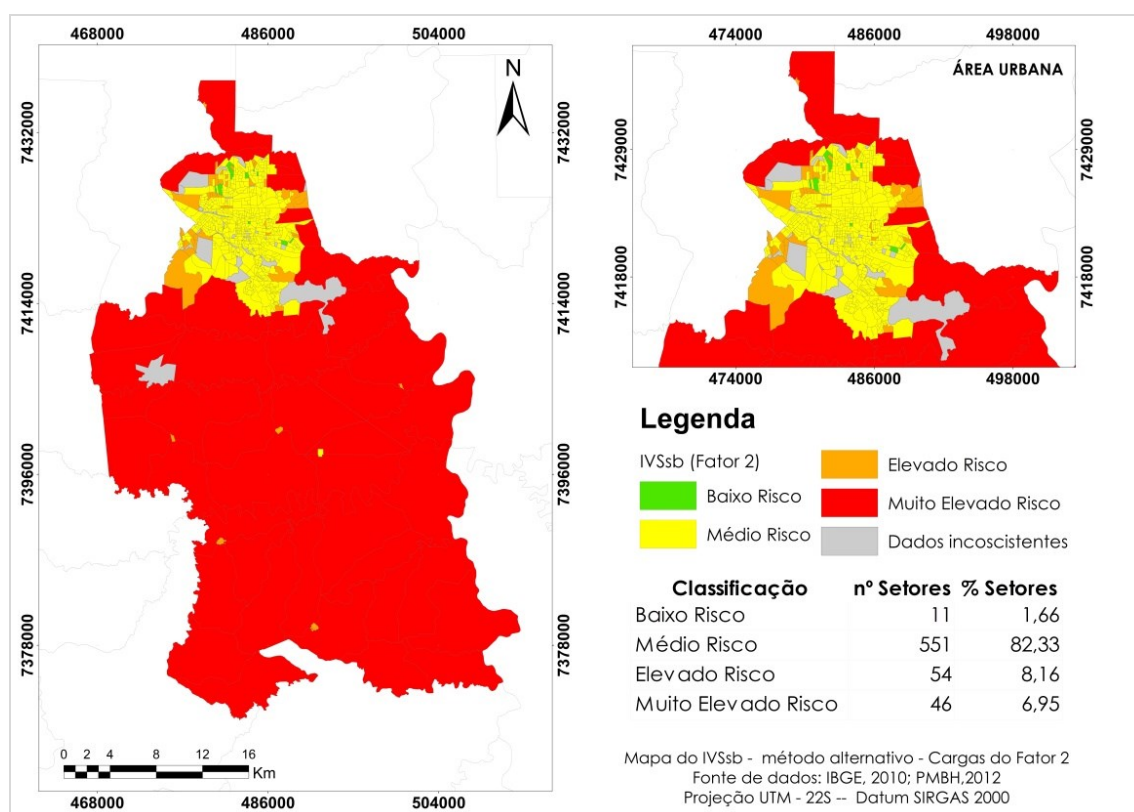
Figura 6 – Mapa do resultado de IVS a partir escores do Fator 1 - IVSse



Fonte: Elaborado pelas autoras.

Quanto aos resultados do IVSsb é possível observar que as piores condições estão mapeadas para os setores rurais. 11 setores foram classificados em baixo risco (setores em verde), representando aproximadamente 2% dos setores, 551 setores (setores em amarelo), foram classificados em médio risco, representando mais de 80% dos setores do município, 54 setores foram classificados em elevado risco, (setores em laranja), aproximadamente 8% dos setores e 46 setores foram classificados em muito elevado risco (setores em vermelho), correspondendo a 7% dos setores, destes, todos localizados em área rural.

Figura 7 – Mapa do resultado de IVS a partir escores do Fator 2 - IVSsb



Fonte: Elaborado pelas autoras.

Sabe-se que as condições locais dos domicílios da área rural implicam em soluções diferenciadas no que diz respeito ao saneamento básico. No entanto, analisando especificamente os dados censitários das áreas rurais do município, verificou-se que boa parte dos domicílios ainda utiliza de fossas rudimentares para disposição do esgotamento sanitário, modelo considerado inadequado do ponto de vista sanitário, além de manterem condições inadequadas quanto à disposição dos resíduos sólidos, como na opção por queima de resíduos na propriedade.

Na Tabela 3 apresentam-se as cargas fatoriais geradas pela análise fatorial e que foram utilizados como “pesos” para estabelecimento do IVS alternativo e também, os pesos estabelecidos pela proposição original (SS-BH). No que diz respeito às cargas estipuladas na metodologia proposta para Belo Horizonte, as atribuições mais significativas de pesos foram direcionadas para as variáveis V2 (abastecimento de água) e V3 (esgotamento sanitário) diferente dos resultados da análise fatorial que indicaram que para o universo de dados de Londrina essas mesmas variáveis estão mais bem representadas pelo Fator 2, ou seja, essas variáveis representam a maior parte da variância que não foi explicada pelo Fator 1, e que explicam 20% do universo de dados (Tabela 3).

Do ponto de vista das condições de saneamento básico, os dois municípios aproximam-se com relação ao índice de qualidade de atendimento da população. Se consultado o relatório de saneamento básico do TrataBrasil para o ano de 2010, observa-se que os dois municípios estão entre os 20 primeiros no *ranking* de melhores condições de saneamento do Brasil, Londrina em 10º e Belo Horizonte em 16º. Destaca-se ainda que Belo Horizonte foi classificado em 2º lugar no que tange ao abastecimento de água e em 1º na coleta de esgotamento sanitário, de forma que na metodologia original foram atribuídos pesos mais significantes à variáveis que apresentavam baixa deficiência no município.

Tabela 3 - Cargas fatoriais estabelecidas pela análise fatorial e pesos estabelecidos pela proposição original da SS-BH

Variáveis	Cargas fatoriais		Proposição SS-BH
	Fator 1	Fator 2	Pesos
V1	0,656		0,073
V2		0,495	0,424
V3	0,234	0,537	0,375
V4		0,970	0,201
V5	0,863		0,283
V6	0,799	0,214	0,288
V7	0,671		0,173
V8	0,940	-0,126	0,185

Fonte: Elaborado pelas autoras.

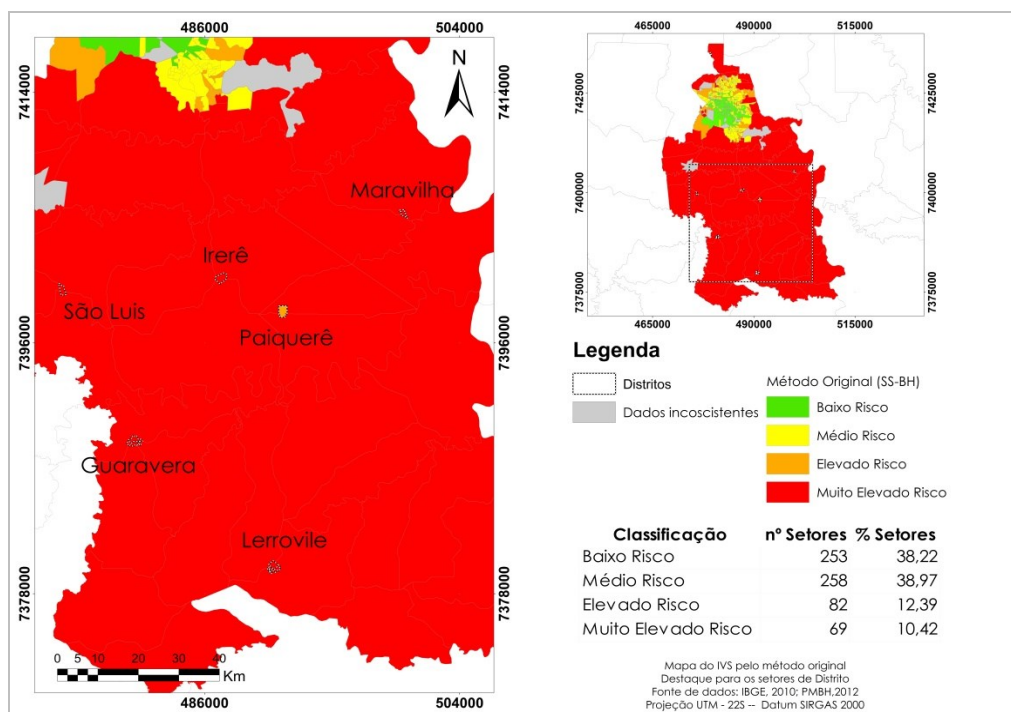
Nesse sentido, a análise fatorial em suas etapas de formulação, permitiu chegar a uma classificação de importância para as oito variáveis, indicando que para Londrina, as variáveis de saneamento básico foram menos preditivas no que diz respeito às deficiências, além disso, o estudo de correlação indicou a relação positiva entre as deficiências em analfabetismo, renda e vulnerabilidade de raças, permitindo a identificação de condições socioeconômicas específicas do Município.

Quando se observa apenas a espacialização dos resultados tais peculiaridades dos métodos não são perceptíveis. Poderia-se concluir que os dois métodos apresentaram resultados semelhantes. Se comparado o resultado dado pelo método do IVS original com o IVSse (proposta alternativa) por exemplo, os setores centrais e localizados mais a sudoeste do município apresentaram baixa vulnerabilidade em ambos modelos. Resultados semelhantes também são observados para os setores localizados nas áreas rurais que foram classificados em muito elevado risco tanto pelo IVS original quanto pelo IVSsb.

De fato, era provável que isso acontecesse considerando que as melhores condições socioeconômicas e de infraestrutura estão localizadas na área central de Londrina e os pesos atribuídos pelos dois métodos foram anulados devido à baixa deficiência observada nas variáveis para estes setores. No entanto, ressalta-se que quando observado os resultados para os setores mais periféricos e distritais do município, diferenciações de classificação de vulnerabilidade são observadas tanto para o IVSse quanto para o IVSsb.

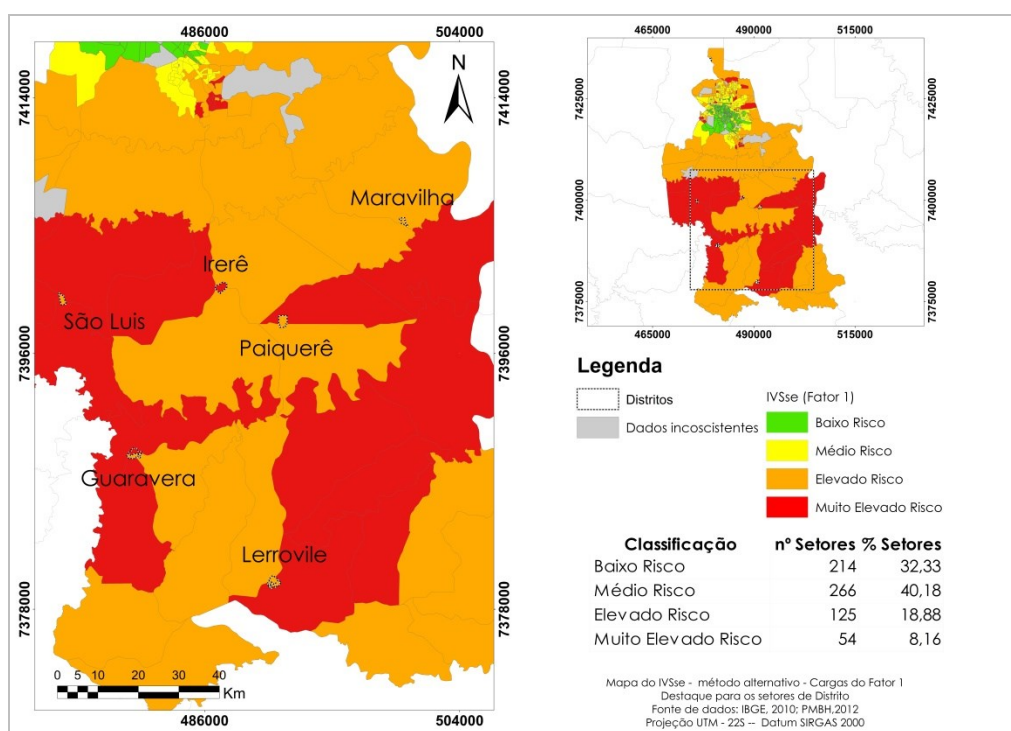
Se considerado os resultados do IVS original, a exceção de um dos distritos (Paiquerê), todos os demais seriam classificados em muito elevado risco (Figura 8), no entanto, pela avaliação gerada a partir da análise fatorial, que resultou no IVSse teríamos apenas 2 distritos em muito elevado risco e um deles parcialmente (Guaravera e Irerê) (Figura 9). No caso do IVSsb cinco distritos foram classificados em elevado risco, indicando melhores condições de infraestrutura em saneamento básico se comparado a setores limítrofes aos distritais (Figura 10).

Figura 8 – Mapa do resultado de IVS a partir do método original. Destaque para os Distritos Municipais.



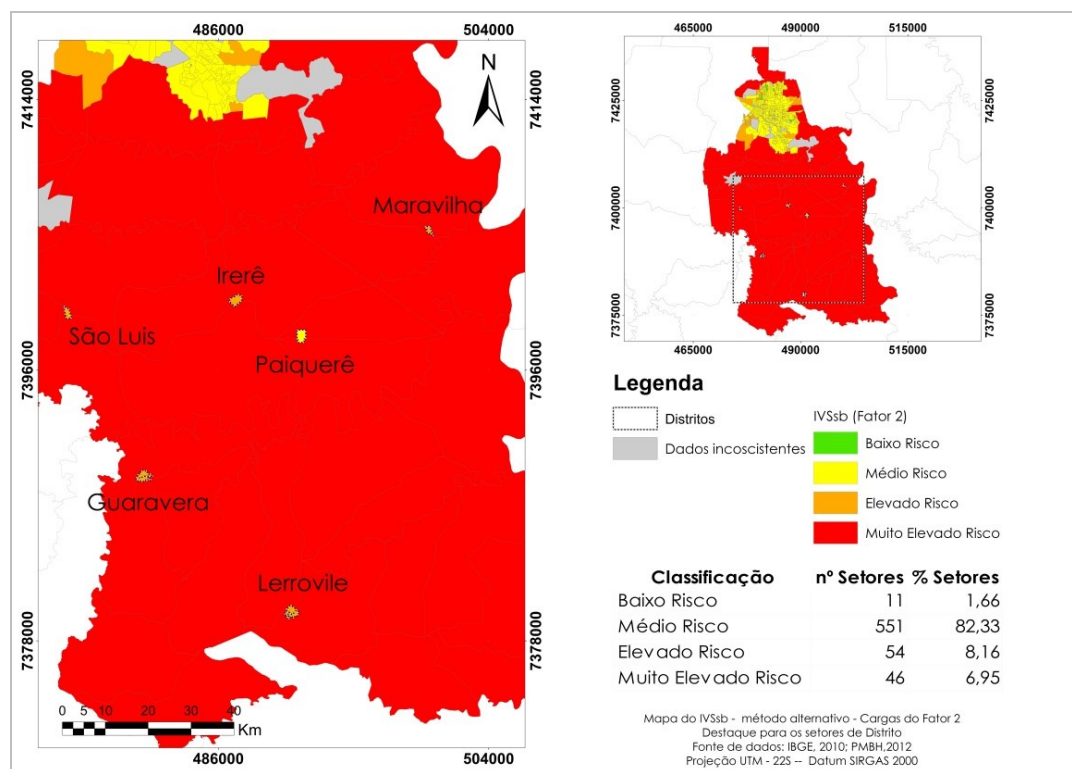
Fonte: Elaborado pelas autoras.

Figura 9 – Mapa do resultado do IVSse a partir do método alternativo. Destaque para os Distritos Municipais



Fonte: Elaborado pelas autoras.

Figura 10 – Mapa do resultado do IVSsb a partir do método alternativo – Cargas do Fator 2. Destaque para os Distritos Municipais



Fonte: Elaborado pelas autoras.

Do ponto de vista metodológico, observa-se que, o emprego de pesos a partir da percepção pessoal poderia subestimar algumas características específicas de municípios. No caso de Londrina, considerando as oito variáveis integradoras do IVS, as variáveis mais significativas na abordagem *data-driven* foram de caráter social e econômico, com exceção da área rural que ainda apresenta vulnerabilidade significativa para as variáveis de saneamento básico, de forma que, neste caso, se confirma a hipótese inicial de pesquisa.

Quanto a possibilidade de simplificação do modelo, alternativas foram encontradas em Hair Junior *et al* (2009): eliminar as variáveis cruzadas, usar outras técnicas de rotação ou reduzir o número de fatores. Neste caso, uma opção seria eliminar as variáveis cruzadas, o que alteraria a proposta de análise inicial ao retirar do modelo de IVS Original variáveis pré-estabelecidas pela SS-BH, de forma que como proposta de avanço na pesquisa uma alternativa seria testar outros modelos de rotação.

Por fim, o método para construção do IVS a partir da análise fatorial resultou em dois mapas, diferente da proposta original. A proposição de mais de um modelo espacial de IVS pode ser uma alternativa para o planejamento de ações em saúde, tendo em vista que, conforme recomendações do Ministério da Saúde (BRASIL, 2014), ações em saúde devem ser intersetoriais, ou seja, devem abranger diversos setores públicos de um município.

CONCLUSÕES

O objetivo desta pesquisa foi comparar dois métodos de construção de um IVS considerando uma perspectiva metodológica utilizada em geoprocessamento baseada em *knowledge-driven* e uma alternativa, por meio de um método *data-driven*, através da técnica estatística de análise fatorial.

Ao contrário do modelo originalmente proposto, elaborado pela SS-BH (MG) e cujos pesos mais significativos foram atribuídos às variáveis do sistema de saneamento básico, o modelo alternativo, considerando a estatística de análise fatorial, indicou que, para o caso de Londrina (PR), as variáveis mais preditivas e com pesos mais significativos seriam: raça, renda e analfabetismo.

Tal resultado deriva justamente dos fundamentos inerentes ao método fatorial aliado às condições socioeconômicas observadas no município: a análise fatorial reuniu em um primeiro fator (Fator 1) cuja representatividade atingiu 40% da variância dos dados, as variáveis raça, renda e analfabetismo porque estas apresentaram indicadores negativos mais significativos do que as condições de saneamento básico.

Neste sentido, entende-se que o uso do modelo original como instrumento de planejamento aplicado para o caso de Londrina, poderia negligenciar questões sociais, econômicas e de infraestrutura. Um exemplo foram os resultados apresentados pelo modelo de análise fatorial para as áreas distritais, que indicaram condições de vulnerabilidade inferiores aos resultados da proposição original.

Destaca-se, ainda, a deficiência dos modelos aplicados nesta pesquisa ao depender de dados secundários disponíveis em um único banco de dados, no caso, o censo demográfico do IBGE. A pré-análise do universo de dados de Londrina mostrou a necessidade de dispensar do procedimento metodológico 55 setores censitários devido a inconsistências de dados, o que não limita a aplicação dos dois modelos, mas impede que o índice possa ser observado para todos os setores censitários do município. Uma revisão com o próximo censo do IBGE, a ser realizado em 2021, pode ser uma alternativa na busca por resultados que compreendam todos os setores do município.

Considerando a possibilidade de avanços nas análises propostas, poderiam ser realizados novos testes de rotação, avançando para testes de correlação espacial com casos de doenças, mapeamento de áreas de probabilidade de ocorrência por meio do uso de algoritmos de *machine learning*, ou mesmo a avaliação de uma possível integração dos mapas resultantes do IVSsb e IVSse por meio de outras técnicas estatísticas.

AGRADECIMENTOS

A presente pesquisa foi realizada com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior-Brasil (CAPES) - (Número do Processo: 1814698)

REFERÊNCIAS

BRASIL, Ministério da Saúde. **Abordagens espaciais em saúde pública**. Brasília: Ministério da Saúde, 2004.

_____. Ministério da Saúde. **Portaria nº. 2.446 de 11 de novembro de 2014**. Redefine a Política Nacional de Promoção da saúde. Brasília: Ministério da Saúde, 2014. Disponível em: http://bvsmis.saude.gov.br/bvs/saudelegis/gm/2014/prt2446_11_11_2014.html. Acesso em: 04 nov.2020.

BOUSQUAT, A; COHN, A. A dimensão espacial nos estudos sobre saúde: uma trajetória histórica. **História, Ciências, Saúde**, Rio de Janeiro, v. 11, n.3, p.549-680, set./dez. 2004. <https://doi.org/10.1590/S0104-59702004000300002>

BORTONI-RICARDO, S. M. *et al.* Raízes sociolinguísticas do analfabetismo no Brasil. **Acolhendo a Alfabetização nos Países de Língua Portuguesa**, São Paulo, v.2, n.4, p. 215-304, 2008. Disponível em: <https://doi.org/10.11606/issn.1980-7686.v2i4p215-234>. Acesso em: 23 ago.2018. <https://doi.org/10.11606/issn.1980-7686.v2i4p215-234>

BUSS, P. M.; PELLEGRINI FILHO, A. A saúde e seus determinantes sociais. **Physis**, Rio de Janeiro, v. 17, n. 1, p.77-93, 2007. <https://doi.org/10.1590/S0103-73312007000100006>

CORRAR, L.J; PAULO, E; DIAS FILHO, J.M. **Análise Multivariada. Para cursos de administração, ciências contábeis e economia**. São Paulo: Editora Atlas S.A, 2012.

HAIR JR, J.F *et al.* **Análise Multivariada de dados**. Tradução: Adonai Schlup Sant'Anna. Porto Alegre: Bookman, 2009. 668 p. Título original: Multivariate Data Analysis. ISBN 0-13-032929-0.

HUSSON, F; JOSSE, J; LÊ S. **FactoMiner**. Disponível em: <http://factominer.free.fr/>. Acesso em: 06 jun. 2017.

IBGE – Instituto Brasileiro de Geografia e Estatística. **Censo demográfico de 2010**. Disponível em: <http://censo2010.ibge.gov.br/>. Acesso em: 19 set. 2016.

IÑIGUEZ, L.; BARCELLOS, C. La Cartografía em salud Pública: Viejos Problemas Y Nuevas Oportunidades. **Revista do Departamento de Geografia da USP**, São Paulo, Volume Especial da Cartogeo, p.390-412, 2014. Disponível em: <https://www.revistas.usp.br/rdg/article/view/85561/88350>. Acesso em: 23 ago.2018. <https://doi.org/10.11606/rdg.v0i0.539>

MANLY, B. J. F. **Métodos Estatísticos Multivariados: uma introdução**. 3. Ed. Porto Alegre: Bookman, 2008.

MEIRELLES, M.S.P; CAMARA, G. ALMEIDA, C.M de. **Geomática. Modelos e Aplicações Ambientais**. Brasília: Embrapa Informação Tecnológica, 2007.

MINGOTI, S.A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada**. Belo Horizonte: Editora UFMG, 2005.

NOGUEIRA, H; REMOALDO, P.C. **Olhares Geográficos sobre a Saúde**. Lisboa: Colibri, 2010.

PMBH - PREFEITURA MUNICIPAL DE BELO HORIZONTE. Secretaria Municipal de Saúde. **Índice de Vulnerabilidade da Saúde**: relatório técnico, Belo Horizonte, 2013. Disponível em: https://prefeitura.pbh.gov.br/sites/default/files/estrutura-de-governo/saude/2018/publicacoes-da-vigilancia-em-saude/indice_vulnerabilidade2012.pdf. Acesso em: 04 nov. 2020.

R- PROJECT. Disponível em: <https://cran.r-project.org/>. Acesso em: 10 jan. 2018.

SANTANA, P. Objeto e métodos em Geografia da Saúde. *In*: SANTANA, Paula. **Introdução a Geografia da Saúde: território, saúde e bem-estar**. Coimbra: Imprensa da Universidade de Coimbra, 2014. p.11-30. <https://doi.org/10.14195/978-989-26-0727-6>

TRATABRASIL. **Ranking de saneamento básico no Brasil**, 2010. Disponível em: <http://www.tratabrasil.org.br/saneamento/principais-estatisticas>. Acesso em: 15 maio. 2018.