

UNIVERSIDADE FEDERAL DO PARÁ  
INSTITUTO DE TECNOLOGIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

**UMA ANÁLISE SÓCIO-DEMOGRÁFICA DA INCIDÊNCIA DE  
HANSENÍASE NA AMAZÔNIA LEGAL BRASILEIRA: ABORDAGEM  
BASEADA EM REDES BAYESIANAS**

JOSÉ MARIA DA SILVEIRA GOMES

DM: 04/2019

UFPA / ITEC / PPGEE  
Campus Universitário do Guamá  
Belém-Pará-Brasil  
2019



UNIVERSIDADE FEDERAL DO PARÁ  
INSTITUTO DE TECNOLOGIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

JOSÉ MARIA DA SILVEIRA GOMES

**UMA ANÁLISE SÓCIO-DEMOGRÁFICA DA INCIDÊNCIA DE  
HANSENÍASE NA AMAZÔNIA LEGAL BRASILEIRA: ABORDAGEM  
BASEADA EM REDES BAYESIANAS**

DM: 04/2019

UFPA / ITEC / PPGEE  
Campus Universitário do Guamá  
Belém-Pará-Brasil  
2019

UNIVERSIDADE FEDERAL DO PARÁ  
INSTITUTO DE TECNOLOGIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

JOSÉ MARIA DA SILVEIRA GOMES

**UMA ANÁLISE SÓCIO-DEMOGRÁFICA DA INCIDÊNCIA DE  
HANSENÍASE NA AMAZÔNIA LEGAL BRASILEIRA: ABORDAGEM  
BASEADA EM REDES BAYESIANAS**

Dissertação de Mestrado submetida à Banca Examinadora do Programa de Pós-Graduação em Engenharia Elétrica como requisito para obtenção do Grau de Mestre em Engenharia Elétrica, na área de Concentração de Computação Aplicada.

UFPA / ITEC / PPGEE  
Campus Universitário do Guamá  
Belém-Pará-Brasil  
2019

Dados Internacionais de Catalogação - na – Publicação (CIP) Sistema de Bibliotecas da UFPA

---

G633a Gomes, José Maria da Silveira, 1952-

Uma Análise sócio-demográfica da incidência de hanseníase na Amazônia Legal brasileira: abordagem baseada em redes bayesianas / José Maria da Silveira Gomes.-2019.

Orientador: Carlos Renato Lisboa Francês  
Dissertação (Mestrado) - Universidade Federal do Pará, Instituto de Tecnologia, Programa de Pós-Graduação em Engenharia Elétrica, Belém, 2019.

1. Teoria bayesiana de decisão estatística. 2, Hanseníases- métodos estatísticos. 3. Saúde pública – aspectos sociais. 4. Indicadores de saúde. I. Título.

CDD 23. ed. 519.542

---

Elaborada por Lucicléa S. de Oliveira – CRB -2/648

UNIVERSIDADE FEDERAL DO PARÁ  
INSTITUTO DE TECNOLOGIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

**“UMA ANÁLISE SÓCIO-DEMOGRÁFICA DA INCIDÊNCIA DE HANSENÍASE NA  
AMAZÔNIA LEGAL BRASILEIRA: ABORDAGEM BASEADA EM REDES  
BAYESIANAS”**

AUTOR: JOSÉ MARIA DA SILVEIRA GOMES

DISSERTAÇÃO DE MESTRADO SUBMETIDA À BANCA EXAMINADORA APROVADA PELO  
COLEGIADO DO PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA, SENDO  
JULGADA ADEQUADA PARA A OBTENÇÃO DO GRAU DE MESTRE EM ENGENHARIA  
ELÉTRICA NA ÁREA DE COMPUTAÇÃO APLICADA.

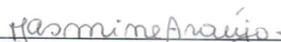
APROVADA EM: 08/02/2019

**BANCA EXAMINADORA:**



**Prof. Dr. Carlos Renato Lisboa Francês**

(Orientador – PPGEE/UFPA)



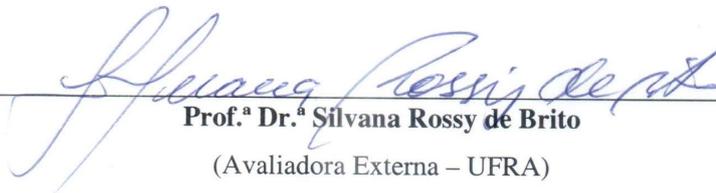
**Prof.<sup>a</sup> Dr.<sup>a</sup> Jasmine Priscyla Leite de Araújo**

(Avaliadora Interna – PPGEE/UFPA)



**Prof.<sup>a</sup> Dr.<sup>a</sup> Regina Fátima Feio Barroso**

(Avaliadora Externa ao Programa – ICS/UFPA)



**Prof.<sup>a</sup> Dr.<sup>a</sup> Silvana Rossy de Brito**

(Avaliadora Externa – UFRA)

**VISTO:**



(Coordenadora do PPGEE/ITEC/UFPA)

Dedico este trabalho à minha mulher Maria Auxiliadora, que sempre esteve comigo em todos os momentos de minha vida. Nas alegrias e principalmente nas dificuldades e que sempre me encorajou e incentivou à realização deste trabalho

## AGRADECIMENTOS

A Deus, por me permitir mais essa oportunidade e ao Espírito Santo pela luz diária.

Agradeço à Nossa Senhora de Nazaré por todas as graças alcançadas.

À minha tia e madrinha Oneide, por me conduzir e direcionar os primeiros passos de minha vida na educação, instrução e principalmente na retidão.

Às minhas filhas Ana Cláudia e Ana Paula razão de nossas vidas e também as minhas netas Ana Júlia e Liv.

Ao meu orientador Prof. Dr. Renato Francês, pela oportunidade, encorajamento e orientações ao longo de toda essa caminhada.

À professora Profa. Dra. Regina Feio Barroso, pelo incentivo, dedicação, disponibilidade e colaboração e principalmente pela “bondosa paciência”.

Às professoras Silvana Rossi de Brito, Aleksandra Silva e a colega do LPRAD Eulália Mata, pelo apoio, colaboração e principalmente pela disponibilidade em ajudar-me para a realização deste trabalho.

Aos professores do LPRAD pela disponibilidade, profissionalismo e dedicação ao PPGEE. Especialmente aos professores Diego e Marcelino pela paciência.

Aos colegas e amigos do LPRAD (Rogério, Alfaia, Anderson, Jonata e Priscila) que me ajudaram e incentivaram nessa caminhada. Principalmente ao Delcinho, sempre atencioso e de sorriso “escancarado”, nos ofertando um cafezinho novo e quentinho.

À Universidade Federal Rural da Amazônia pelo apoio e aos colegas professores do Instituto Ciber Espacial em liberar-me para essa qualificação, principalmente os Diretores Prof. Dr. Pedro Campos e Prof. Dr. Rosemiro Galate por incentivar-me.

À Profa. Dra. Merilene, pelas oportunidades concedidas em minha vida profissional e acima de tudo pela amizade.

À Profa. Dra. Janae, pelo incentivo e oportunidades concedidas.

Ao Prof. Dr. Sodek, amigo de muitos anos, pela valiosa colaboração neste trabalho.

À Universidade Federal do Pará e ao programa de Pós-Graduação em Engenharia Elétrica pela oportunidade de um novo aprendizado.

Por fim, agradeço a todos que de alguma maneira contribuíram para que este trabalho pudesse ser realizado.

## SUMÁRIO

<b>LISTA DE FIGURAS</b> .....	<b>XI</b>
<b>LISTA DE TABELAS</b> .....	<b>XII</b>
<b>LISTA DE QUADROS</b> .....	<b>XIII</b>
<b>LISTA DE ABREVIATURAS</b> .....	<b>XIV</b>
<b>RESUMO</b> .....	<b>XV</b>
<b>ABSTRACT</b> .....	<b>XVII</b>
<b>CAPÍTULO 1 – INTRODUÇÃO</b> .....	<b>19</b>
1.1 MOTIVAÇÕES .....	19
1.2 CONTEXTO .....	20
1.3 HIPÓTESES .....	24
1.4 OBJETIVOS .....	24
1.5 ORGANIZAÇÃO DO TRABALHO .....	24
<b>CAPÍTULO 2 – TRABALHOS CORRELATOS E REVISÃO BIBLIOGRÁFICA</b>	<b>25</b>
2.1 CONSIDERAÇÕES INICIAIS.....	25
2.2 REDES BAYESIANAS .....	26
2.3 TÉCNICAS COMPUTACIONAIS PARA APOIO A DECISÃO EM SAÚDE COM ENFASE NAS REDES BAYESIANAS .....	29
2.4 A HANSENÍASE ESTUDADA E VISTA NESSE CONTEXTO .....	35
2.5 CONSIDERAÇÕES FINAIS .....	44
<b>CAPÍTULO 3 – METODOLOGIA DA COLETA DE DADOS SOBRE A SITUAÇÃO SÓCIO-DEMOGRÁFICA E A ABORDAGEM DA UTILIZAÇÃO DAS REDES BAYESIANAS.....</b>	<b>45</b>
3.1 CONSIDERAÇÕES INICIAIS .....	45
3.2 COLETA DE DADOS .....	45
3.3 METODOLOGIA UTILIZADA PARA ANÁLISE DOS DADOS .....	47
3.3.1 Fase I - Preparação dos dados .....	47
3.3.2 Fase II - Mineração dos dados .....	48
3.3.3 Fase III – Visualização da distribuição espacial .....	49
3.3.4 Fase IV – Análise dos dados .....	49
3.4 CONSIDERAÇÕES FINAIS.....	49
<b>CAPÍTULO 4 – ANÁLISE SÓCIO-DEMOGRÁFICA DA INCIDÊNCIA DA HANSENÍASE UTILIZANDO AS REDES BAYESIANAS</b> .....	<b>50</b>
4.1 CONSIDERAÇÕES INICIAIS .....	50
4.2 AJUSTES NECESSÁRIOS PARA UNIFORMIZAR OS DADOS PARA A ANÁLISE .....	50
4.3 ANÁLISE DOS DADOS .....	51
4.4 CONSIDERAÇÕES FINAIS .....	60
<b>CAPÍTULO 5 – CONCLUSÃO</b> .....	<b>61</b>
5.1 CONTRIBUIÇÕES DESTE TRABALHO .....	63
5.2 TRABALHOS FUTUROS E DESDOBRAMENTOS DA PESQUISA .....	63
5.3 DIFICULDADES ENCONTRADAS .....	64

**REFERÊNCIAS BIBLIOGRÁFICAS.....**  
**ANEXO I – TRABALHO PUBLICADO**

- Figura 1.** Distribuição Global da Hanseníase no mundo
- Figura 2.** Distribuição da Hanseníase em 10 estados Brasileiros em 2017
- Figura 3.** Rede bayesiana utilizada para instanciar as métricas de complexidade.
- Figura 4.** Metodologia em fases para análise dos dados.
- Figura 5.** Casos de Hanseníase nos estados brasileiros
- Figura 6.** Estados com menor número de casos de Hanseníase detectados
- Figura 7.** Rede bayesiana com as 10 variáveis de estudo.
- Figura 8.** Rede bayesiana com 6 variáveis escolhidas.
- Figura 9.** Estrutura da rede bayesiana selecionada.
- Figura 10.** Rede bayesiana (evidências: *EspaçoReferência = "Amazônia"* e *DensidadePopMaior2 > 32,58*).
- Figura 11.** Mapa de casos de Hanseníase nos municípios brasileiros.
- Figura 12.** Comparativo da Hanseníase por Região nos anos de 2010 e 2018.
- Figura 13.** Comparativo da Hanseníase por Estado nos anos de 2016 a 2018.

- Tabela 1** Distribuição da Hanseníase por UF na Amazônia Legal
- Tabela 2** Trabalhos pesquisados e encontrados
- Tabela 3** Categorização da taxa de incidência.
- Tabela 4** Categorização de *IDHM-Renda*, *IDHM-Educação*, *BanheiroAguaEncanada* e *DensidadePopMaior2*.
- Tabela 5** Hanseníase, conforme a Taxa de Incidência - média (Amazônia, Outra) e municípios selecionados.
- Tabela 6** Casos e taxa de incidência de Hanseníase de acordo com a classificação do município.
- Tabela 7** Municípios com menor e maior taxa de incidência de Hanseníase em 2010.

**LISTA DE QUADROS**

- Quadro 1** Distribuição dos estudos analisados de Redes Bayesianas quanto ao autor, ano, e contribuição 2019.
- Quadro 2** Distribuição dos estudos analisados de Hanseníase quanto ao autor, ano, e contribuição 2019.
- Quadro 3** Variáveis selecionadas e fontes de dados

<b>AM</b>	Amazonas
<b>BN</b>	Networks Bayesianas
<b>BR</b>	Brasil
<b>CGHDE</b>	Coordenação Geral de Hanseníase e Doenças em Eliminação
<b>CT</b>	Clique Tables
<b>DAG</b>	Grafo Acíclico Direto
<b>DATASUS</b>	Departamento de Informática do Sistema Único de Saúde
<b>DPC</b>	Distribuição de Probabilidade Conjunta
<b>EPUSP</b>	Escola Politécnica da Universidade de São Paulo
<b>GDA</b>	Gráficos Direcionais Acíclicos
<b>GO</b>	Goiás
<b>HU-USP</b>	Hospital Universitário da USP
<b>IBGE</b>	Instituto Brasileiro de Geografia e Estatística
<b>ID</b>	Diagramas de Influências
<b>IDHM</b>	Índice de Desenvolvimento Humano Municipal
<b>K2</b>	Algoritmo de busca heurística
<b>MGP</b>	Modelos Gráficos Probalilísticos
<b>MN</b>	Networks Markov
<b>MS</b>	Ministério da Saúde
<b>MS</b>	Mato Grosso do Sul
<b>MT</b>	Mato Grosso
<b>NHR</b>	<i>Netherlands Hanseniasis Relief</i>
<b>OMS</b>	Organização Mundial da Saúde
<b>ONU</b>	Organização das Nações Unidas
<b>PA</b>	Pará
<b>PE</b>	Pernambuco
<b>PNAD</b>	Pesquisa Nacional por Amostra de Domicílios
<b>PR</b>	Paraná
<b>RIPSA</b>	Rede Integrada de Informação para a Saúde
<b>RNA</b>	Redes Neurais Artificiais
<b>RR</b>	Roraima
<b>RS</b>	Rio Grande do Sul
<b>SE</b>	Sergipe
<b>SINAN</b>	Sistema de Informação de Agravos de Notificação
<b>SUS</b>	Serviço Único de Saúde
<b>TIC</b>	Tecnologia da Informação e Comunicação
<b>TO</b>	Tocantins
<b>TPC</b>	Tabelas de Probabilidades Condicionais
<b>USP</b>	Universidade de São Paulo
<b>UTI</b>	Unidade de Tratamento Intensivo

A Hanseníase é uma doença infectocontagiosa milenar, de caráter crônico e estigmatizante, desde os tempos mais remotos da humanidade até os dias de hoje. Caracteriza-se por uma doença da pobreza e o Brasil é o segundo país do mundo com a maior incidência. A falta de políticas públicas para a redução da pobreza através da melhoria dos indicadores sócio-econômicos do país, estão diretamente relacionados à incidência da doença no Brasil. As estratégias para o controle e monitoramento devem perpassar por ações inteligentes. Uma das soluções para o monitoramento da doença é a utilização das redes bayesianas como método probabilístico para a tomada de decisões no controle e na tomada de decisão quanto aos procedimentos a serem adotados para a redução da incidência da doença. O trabalho tem como objetivo analisar a associação da incidência da doença Hanseníase em relação aos indicadores do desenvolvimento humano, habitação e nível de renda, considerando os municípios da Amazônia Legal brasileira em relação ao Brasil. Estudo ecológico, baseado em dados obtidos sobre os casos de Hanseníase no Brasil no ano de 2010, disponibilizados pelo Sistema de Informação de Agravos de Notificação (SINAN) através do Departamento de Informática do SUS (DATASUS) e os indicadores socioeconômicos através da base da Pesquisa de Domicílios do Censo Demográfico do Instituto Brasileiro de Geografia e Estatística – IBGE e componentes do Índice de Desenvolvimento Humano Municipal, referentes à educação e renda foram obtidos a partir da plataforma de consulta Atlas do Desenvolvimento Humano no Brasil, também adotando como referência o ano de 2010. A metodologia combina as etapas da mineração de dados obtidos através dos bancos de dados consultados, com a análise da distribuição espacial. Foi utilizada a técnica de redes bayesianas com o objetivo de medir a associação entre as variáveis do domínio do problema e para estabelecer a analogia dos dados entre os municípios estudados com os dados encontrados para os demais municípios brasileiros. Com a aplicação do algoritmo K2, foram encontradas associações relevantes com os seguintes indicadores aplicados no estudo: Amazônia Legal Brasileira, Índice de Desenvolvimento Humano Municipal de Renda e Educação e Condição de Habitação dos Domicílios. No modelo de rede bayesiana encontrado, existe uma associação significativa entre os percentuais de domicílios com densidade maior que 2 e taxa de incidência de Hanseníase. Embora a relação entre a taxa de incidência, fatores socioeconômicos, baixos índices educacionais e de renda já tenha sido evidenciada em vários estudos, a inserção do indicador que considera a densidade populacional do domicílio foi uma proposta inovadora deste trabalho, sendo o indicador mais significativo

neste estudo. A análise da incidência da Hanseníase pela distribuição espacial comparativa entre a Amazônia Legal e o Brasil, contribuiu para demonstrar que a política pública de habitação para a região estudada é quase inexistente, já que a densidade populacional por moradia é muito alta, facilitando o aparecimento de doenças infectocontagiosas, como a Hanseníase.

**DESCRITORES: Saúde Pública, Hanseníase, Amazônia, redes bayesianas.**

## ABSTRACT

Leprosy is a millenarian contagious disease, with chronic and stigmatizing characteristics, from the remotest times of humanity until today. It is characterized as a disease of the poor and Brazil is the second country in the world with the highest incidence. The lack of public policies aimed at reducing poverty through the improvement of socio-economic factors in the country is directly related to the incidence of the disease in Brazil. Strategies for control and monitoring should follow intelligent actions. One of the solutions for monitoring the disease is the use of Bayesian networks as a probabilistic method for taking decisions on both the control and the procedures to adopt in order to reduce the incidence of the disease. The objective of the present study is to analyse the association of leprosy incidence in relation to indicators of human development, habitation and income level, considering the Brazilian Amazon region in relation to the entire country. An ecological study, based on data obtained on cases of leprosy in Brazil for the year 2010, obtained from the Information System of Hardship Notifications (SINAN) through the Informatics Department of the National Health Service (DATASUS) and the socio-economic indicators found in the Demographic Census Research database of the Brazilian Institute for Geographical and Statistical Survey – IBGE, as well as information from the Municipal Human Development Index, regarding education and income, obtained from the website of the Human Development Atlas of Brazil, also for the year 2010. The methodology combined data mining with the analysis of spatial distribution. The Bayesian network technique was used aimed at measuring the association between variables of the domain of the problem as well as to establish the analogy of the data between the municipalities under study with data for all other Brazilian municipalities. Applying the algorithm K2 relevant associations were found for the following indicators applied in the investigation: Brazilian Legal Amazon, Municipal Human Development Index of Income and Education and Household Housing Condition. Using the Bayesian network model adopted, there is a significant association between the percentage of homes with more than 2 inhabitants and the rate of incidence of leprosy. Although the relationship between the rate of incidence, socio-economic factors (no water supply, no toilet, poverty and overcrowding of the home), low educational indices and income has already been reported in several studies, the insertion of the indicators that considers population density of the home was a novel proposition of the

present study and the indicators of greatest most significance of this investigation. The analysis of leprosy incidence with respect to spatial distribution, comparing the Amazon region with the entire country, revealed that public policies for habitation in the studied region were almost non-existent, since the population density of homes is very high, facilitating the appearance of contagious diseases such as leprosy.

**KEYWORDS:** Public Health, Leprosy, Amazon, Bayesian Networks.

## CAPÍTULO 1 – INTRODUÇÃO

Neste primeiro capítulo, serão apresentadas as motivações que levaram à realização do trabalho, uma contextualização sucinta sobre a doença Hanseníase e redes bayesianas, as hipóteses levantadas, os objetivos a serem alcançados pelo estudo e a organização do trabalho.

### 1.1 MOTIVAÇÕES

A Hanseníase é uma doença infecto-contagiosa, de caráter crônico, que apesar dos avanços da Medicina, ainda é um problema de Saúde Pública. Durante milênios, a doença estigmatizante, era considerada como uma doença que atingia as pessoas impuras e que portanto, deveriam ser castigadas com as lesões e as anormalidades funcionais provocadas por seu agente causador.

Na contemporaneidade, sabe-se que a Hanseníase é uma doença da pobreza, causada pelas iniquidades sociais e pelo acesso limitado aos serviços de saúde, o que dificulta o diagnóstico precoce, imprescindível à sua cura.

A Declaração do Milênio, elaborada pela Organização das Nações Unidas, foi elaborada a partir da concepção de oito objetivos firmados entre seus estados membros, por uma série de compromissos que deveriam ser cumpridos nos prazos fixados, com a finalidade de melhorar o destino da humanidade neste século.

Os objetivos, com metas e indicadores, foram assumidos por 164 países. O sexto objetivo trata do combate às doenças que ameaçam acima de tudo as populações mais pobres e vulneráveis, comprometendo-se os países com maior índice de pobreza a reduzir sua incidência, por meio do acesso da população à informação, aos meios de prevenção e aos meios de tratamento, sem descuidar da criação de condições ambientais e nutritivas que estanquem os ciclos de reprodução dessas doenças (ONU, 2000).

As chamadas doenças negligenciadas, dentre as quais a Hanseníase, são agravos que persistem no país e estão relacionadas às condições de pobreza, iniquidades em saúde e exclusão social (AUGUSTO, 2016).

No Brasil, apesar da redução significativa do perfil de mortalidade por doenças infecciosas, algumas doenças ainda se manifestam sob a forma epidêmica como a dengue e outras doenças de contaminação vetorial, até pela persistência de endemias importantes, como a tuberculose e a Hanseníase, as quais se esperavam estar controladas (BARRETO, 2013).

O estabelecimento de uma política pública eficaz de prevenção, controle e tratamento, requer um eficiente processo de registro, tratamento e recuperação de informação para que a conduta diagnóstica, terapêutica e epidemiológica possa ser tomada de maneira correta.

Os avanços das Tecnologias da Informação e Comunicação (TIC's), estimuladas pela Internet, influenciam o modo de vida das sociedades, das organizações, quer privadas ou públicas, através de instrumentos que deveriam ser utilizados como apoio na elaboração das políticas de estado e de governo, em especial na área da saúde para a previsão de epidemias e controle de doenças de caráter endêmico,

A utilização de sistemas computacionais combinada com a aplicação de modelos probabilísticos, como por exemplo o modelo de redes bayesianas (CHEN, 2001; COOPER E HERSKOVITS, 1992) podem se apresentar como soluções eficazes no apoio à tomada de decisão para gestores da área de saúde.

A construção de políticas públicas para o controle da Hanseníase na Amazônia Legal, onde em diversos municípios a doença é considerada endêmica, poderia ser apoiada pela utilização do modelo das redes bayesianas, cujo método probabilístico seria imprescindível para a tomada de decisões no monitoramento da doença nessa região do país, sendo a principal motivação para a realização deste estudo.

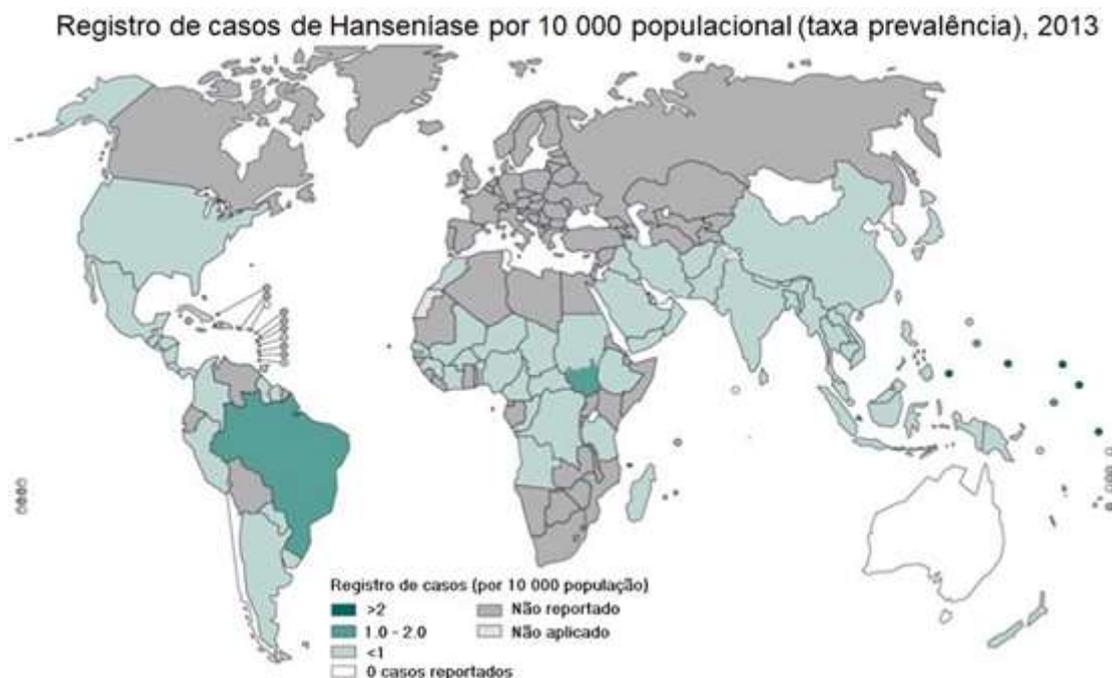
## 1.2 CONTEXTO

A Hanseníase é uma doença infectocontagiosa, crônica, causada pela bactéria *Mycobacterium leprae*. Possui uma dimensão social inigualável porque tem a capacidade de infectar grande número de indivíduos, porém, poucos adoecem. Sua maior incidência se dá em classes socioeconômicas baixas, devido à multiexposição, baixo nível de formação, nutrição e moradia (BRITO *et al.* 2014). Portanto, o domicílio é apontado como importante espaço de transmissão da doença (OLIVEIRA *et al.* 2014).

A Hanseníase ainda se apresenta como um problema de saúde pública no Brasil e necessita de políticas públicas mais eficazes. O diagnóstico tardio, a falta de busca aos serviços de tratamento e a negligência com o autocuidado pode resultar em deformidades e incapacidades físicas que, além do ônus financeiro aos sistemas assistenciais, reflete na autoestima e na forma como a pessoa se percebe no mundo, levando-a a evitar relações e afastar-se do convívio social. As incapacidades refletem na autoestima e participação social. O estigma, mais presente na Hanseníase, tem importância destacada, principalmente em relação ao sexo, em que homens apresentam maiores prejuízos, dificultando a interpretação da cura (MELINSKI *et al.*, 2014).

Em 2013, 105 países/territórios reportaram à Organização Mundial de Saúde (OMS), 215.656 casos novos de Hanseníase. Destes, 72% ocorreram na Ásia e 15% nas Américas. A

Índia com 126.913 casos novos e o Brasil com 31.044 casos novos, conforme observado na Figura 1.



**Figura 1.** Distribuição global da Hanseníase no mundo

Fonte: OMS, 2013

A Hanseníase ainda constitui um relevante problema de saúde pública. O Brasil se mantém como segundo país com maior número de casos novos detectados no mundo, precedido pela Índia. Embora a detecção anual venha declinando, no início de 2010 a incidência global de Hanseníase era de cerca de 212 mil casos. No Brasil, em 2014, dos 31.064 casos novos detectados, 2.341 foram em menores de 15 anos e o coeficiente de detecção geral foi de 15,32 para cada 100 mil habitantes. Alguns municípios no Brasil já alcançaram a meta de eliminação da doença, no entanto, as regiões Norte, Centro-Oeste e Nordeste ainda apresentam altos patamares de casos da doença. Apesar da tendência de estabilização dos coeficientes de detecção no Brasil, os coeficientes apresentam grande variação nas diferentes regiões brasileiras (GRACIE *et al.* 2017).

Entre as doenças diretamente ligadas às condições de vida e às deficiências nutricionais, merecem citação a Hanseníase e a tuberculose, em razão dos indicadores atingidos na região. Conforme o coeficiente de detecção da Hanseníase por 100.000 habitantes para o ano de 2010, os estados do Mato Grosso, Tocantins, Maranhão e Rondônia apresentaram os maiores resultados do país. As taxas confirmaram esses estados como hiperendêmicos (VIANA, *et al.*, 2015), (Tabela 1).

**Tabela 1.** Distribuição da Hanseníase por UF na Amazônia Legal

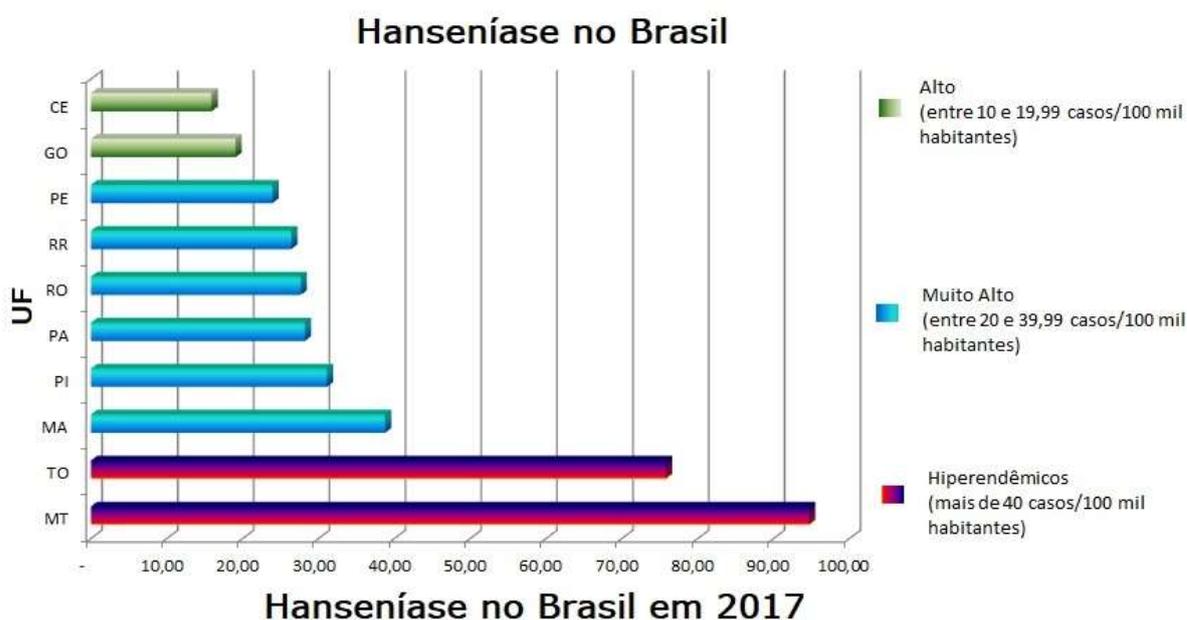
<b>Coefficiente de detecção de Hanseníase na Amazônia Legal, Brasil (2010)</b>	
<b>Região/UF</b>	<b>Coefficiente de Hanseníase por 100.000 habitantes</b>
Brasil	18,75
Amazônia Legal	49,26
Rondônia	58,82
Acre	34,22
Amazonas	20,58
Roraima	32,63
Pará	49,43
Amapá	21,06
Tocantins	79,06
Maranhão	62,45
Mato Grosso	85,07

Fonte: RIPSA (2013). Elaboração dos autores a partir do tratamento estatístico de dados.

Em outro estudo, (BRITO, 2018), cita que Mato Grosso e Tocantins foram os dois estados com maior concentração de casos de Hanseníase em 2017, conforme dados preliminares divulgados pelo Ministério da Saúde. Enquanto Mato Grosso apresentou taxa de detecção de 94,69 casos para cada 100 mil habitantes, a taxa foi de 75,8 para Tocantins. Com estes índices, esses dois estados são considerados hiperendêmicos. Para não ser considerada como problema de saúde pública, a doença deve ter taxa de detecção inferior a 10 casos para cada 100 mil habitantes. Esta taxa estimada para 2017 é de 11,66 casos no Brasil. Outros seis estados das regiões Norte e Nordeste (ver Figura 2 abaixo) apresentaram taxas de detecção acima de 19 casos para cada 100 mil habitantes, sendo incluído no parâmetro “muito alto”, segundo indicadores do monitoramento adotado pelo Ministério. Os dados foram divulgados na manhã desta quarta-feira, 31, durante lançamento da Campanha Nacional de Luta Contra a Hanseníase 2018, realizado no município de Belém (PA). Apesar de uma queda de 37% no registro de novos casos no Brasil desde 2007, o País ainda tem o controle da Hanseníase como um desafio, explicou Carmelita Ribeiro Filha, coordenadora à frente da Coordenação Geral de Hanseníase e Doenças em Eliminação (CGHDE), do Ministério da Saúde. Ela ainda esclareceu que os registros para 2017 são parciais, com consolidação das informações estimada para os meses de maio e junho de 2018 destacado por (BRITO, 2018).

No Brasil, foram 24,2 mil novos casos diagnosticados em 2017. Destes, a coordenadora destacou que 1,5 mil (equivalente a 6%) dos casos atingiram crianças de até 15 anos. A

manifestação da doença em crianças indica que há circuitos ativos de transmissão em áreas mais endêmicas, com pessoas do convívio familiar sem acesso ao diagnóstico e ao tratamento da Hanseníase. Outra preocupação apresentada foi com o público masculino, que apresenta a maior taxa média de detecção de novos casos com grau 2 de incapacidade. Na avaliação clínica dos pacientes, entram no indicador de grau 2 aqueles com deficiências visíveis nos olhos, nas mãos ou nos pés. Diagnosticar novos casos nesta situação mostra que a detecção da doença chegou em momento tardio para que estas incapacidades fossem prevenidas. Dados demonstrados por (BRITO, 2018). (Figura 2).



**Figura 2.** Distribuição da Hanseníase em 10 estados Brasileiros em 2017.

Fonte: (BRITO, 2018). Elaboração dos autores a partir do tratamento estatístico de dados.

Para que se estabeleça uma política pública eficaz de prevenção, controle e tratamento, há a necessidade que se estabeleça um eficiente processo de informação, para que a conduta diagnóstica, terapêutica e epidemiológica possa ser tomada de maneira correta para o aprimoramento dos modelos de causalidades das condições de saúde (SATO *et al.*, 2015).

A utilização da inteligência artificial, por intermédio de sistemas computacionais, através de modelos gráficos probabilísticos (MGP), em especial as redes bayesianas, podem e são utilizadas no apoio da tomada de decisão na área da saúde (SATO *et al.*, 2015).

As redes bayesianas são grafos acíclicos orientados cujos “nós” representam as variáveis e cujos “ramos” associam as relações de dependência entre as variáveis em um modelo

probabilístico. Diz-se também que os ramos ou arestas direcionadas representam as ligações de causa-e-efeito entre variáveis. São denominadas variáveis *Pais* aquelas causais (ou montante), e de nós *Filhos* aquelas “sob efeito de” (ou colocadas a jusante), (BASTOS e OLIVEIRA, 2017). Isso oferece a flexibilidade de incluir dados como quantidades observadas, variáveis latentes, parâmetros desconhecidos ou até mesmo hipóteses levantadas pelo pesquisador, Sato *et al.*, 2015. Os sistemas baseados em redes bayesianas são capazes de gerar automaticamente previsões ou decisões mesmo na situação de inexistência de algumas peças de informação (FRED, 2006).

Como as redes bayesianas são abordagens interpretativa e analítica para probabilidades, é também um método de modelagem e de decisão, daí serem utilizadas em diversas áreas do conhecimento, dentre as quais, a saúde, como proposta de análises do controle e prevenção das enfermidades, a partir das variáveis selecionadas (ARA-SOUZA, 2010).

### 1.3 HIPÓTESES

Neste trabalho serão investigadas as seguintes hipóteses:

- a) A incidência da Hanseníase está diretamente relacionada às variáveis sócio-demográficas em todas as regiões brasileiras;
- b) O uso das redes bayesianas como método probabilístico pode apoiar a tomada de decisões na construção de políticas no controle da Hanseníase.

### 1.4 OBJETIVOS

Considerando-se a importância do estudo da incidência da Hanseníase e a escassez de estudos que avaliem a ocorrência desse fenômeno nos municípios da Amazônia Legal Brasileira em relação ao Brasil, este trabalho tem o objetivo de *analisar, utilizando as redes bayesianas, a associação entre a incidência da Hanseníase e os indicadores municipais de educação, renda e a densidade populacional nos domicílios, tomando como referência os municípios da Amazônia Legal Brasileira em relação aos demais municípios do Brasil*. O termo "Amazônia Legal Brasileira" será referido no restante deste texto como "Amazônia".

### 1.5 ORGANIZAÇÃO DO TRABALHO

Este trabalho está organizado em capítulos, na seguinte ordem:

- a) No primeiro capítulo, introdutório, são caracterizadas a motivação, o objetivo, as hipóteses e realizada uma contextualização relacionada ao tema abordado.

- b) No segundo capítulo é realizada uma revisão bibliográfica, descrevendo referencial teórico sobre as técnicas computacionais aplicadas como apoio nas decisões em saúde, sobre as redes bayesianas e sua aplicação na área da saúde e sobre a Hanseníase no contexto deste estudo.
- c) No terceiro capítulo são descritas as técnicas para a coleta dos dados a partir da seleção dos bancos de dados e realizada a seleção das variáveis a serem coletadas. Ainda neste capítulo são estabelecidas as fases para o tratamento dos dados, bem como a discretização das variáveis a serem analisadas pela rede bayesiana.
- d) No capítulo quatro, os dados são analisados após os ajustes necessários para a construção da rede bayesiana, foram mostrados os resultados encontrados e comparados com outros trabalhos descritos na literatura.
- e) No quinto capítulo são apresentadas as conclusões do trabalho, a dificuldade encontrada para sua realização e sua contribuição para estudos utilizando técnica computacional sobre as decisões em saúde na Amazônia Legal.

## **CAPÍTULO 2 – TRABALHOS CORRELATOS E REVISÃO BIBLIOGRÁFICA**

### **2.1. CONSIDERAÇÕES INICIAIS**

Tendo em vista de trilhar um rumo e um melhor entendimento a esta pesquisa, este capítulo se propõe a apresentar um levantamento dos trabalhos correlatos, pesquisados e selecionados contendo as técnicas de apoio as decisões nas áreas de saúde. Também faz uma revisão sobre as redes bayesianas e sua aplicabilidade de forma geral da área de saúde. E por fim, uma visão do estudo dos questionamentos e comprovações quanto à Hanseníase, enquanto doença que afeta as populações vulneráveis, entre as quais a da Amazônia Brasileira.

A revisão bibliográfica realizada nesse estudo, foi dividida em duas partes. A primeira aborda os trabalhos publicados sobre técnicas computacionais para apoio a decisão em saúde com a utilização das redes bayesianas. A segunda aborda os trabalhos publicados da Hanseníase em diversos contextos.

Para esse estudo, foram realizadas várias pesquisas de busca nas diversas bases de dados, principalmente o Google Acadêmico, tanto na língua mãe (português) como em outros idiomas como inglês, espanhol e francês. Foram encontrados vários trabalhos e também descartados conforme mostra a (Tabela 2):

**Tabela 2**– Trabalhos pesquisados e encontrados

<b>Palavras-Chave</b>	<b>Nº de trabalhos pesquisados</b>	<b>Nº de trabalhos selecionados</b>	<b>Nº de trabalhos descartados</b>
Redes bayesianas	30	26	4
Epidemiologia da Hanseníase	35	28	7

## 2.2 REDES BAYESIANAS

Redes bayesianas (RB) são modelos probabilísticos utilizados para descrever a relação entre variáveis aleatórias. Representam, de forma compacta, a distribuição conjunta de probabilidades de um conjunto de variáveis, (RUSSEL e NORVIG, 2002).

O termo redes bayesianas foi criado por Judea Pearl em 1985, mas as ideias e conceitos vieram de muitas fontes. (PEARL, 1986), também conceitua redes bayesianas como modelos probabilísticos que permitem lidar de forma rigorosa com a representação de conhecimentos em domínios onde existe incerteza. As redes bayesianas oferecem uma abordagem para o raciocínio probabilístico que engloba teoria dos grafos, para o estabelecimento das relações causais entre sentenças: a parte qualitativa, que representa as dependências sobre os nós e a teoria das probabilidades, para atribuição de níveis de confiabilidade, e a parte quantitativa, (tabelas de probabilidades condicionais - TPCs desses nós), que avalia, em termos probabilísticos, essas dependências. (FISCHER, 2010). As probabilidades definem as chances que uma variável possui de assumir determinado valor ou estado. Cada variável possui um conjunto finito de estados que pode assumir. (SEFFRIN e JACQUES, 2014).

Uma Rede Bayesiana consiste, segundo Jensen (in FISCHER, 2010):

- Um conjunto de variáveis e um conjunto de arcos ligando as variáveis.
- Cada variável possui um conjunto limitado de estados mutuamente exclusivos.
- As variáveis e arcos formam necessariamente um DAG (gráfico acíclico direto).

As redes bayesianas, segundo (PEARL, 1988), oferecem duas vantagens importantes para modelar o tipo de conhecimentos em domínio:

- a) permitem uma representação gráfica das relações causais entre as variáveis do domínio, mas que leva em conta o grau ou peso da relação causal através de probabilidades;
- b) oferecem um mecanismo eficiente e exato de cálculo da distribuição de probabilidades das variáveis do domínio.

Os componentes qualitativo e quantitativo que compõem as redes bayesianas, propiciam uma representação eficiente da distribuição de probabilidade conjunta (DPC) do conjunto de

variáveis aleatórias  $X = \{X_1, X_2, \dots, X_n\}$  de um determinado domínio. Essa distribuição de probabilidade é dada pela equação (1):

$$P(x_1, \dots, x_n) = \prod_{i=1}^n P(x_i | Pa_i) \quad (1)$$

na qual  $Pa_i$  são os nós-pais do nó  $X_i$ . Essa representação acarreta uma redução substancial do número de probabilidades a serem manipuladas, a partir da utilização do conceito de independência condicional, expressa por essa equação.

Desta forma, uma rede bayesiana provê uma completa descrição do domínio. Cada entrada na DPC pode ser calculada a partir das informações da rede. Uma entrada genérica da DPC é a probabilidade de uma conjunção de uma particular atribuição de valores para cada variável, assim como  $P(X_1 = x_1 \cap X_2 = x_2 \dots \cap X_n = x_n)$ . Utilizando a notação  $P(x_1, \dots, x_n)$ , os valores das entradas da DPC podem ser obtidos por intermédio da equação 1.

Cada entrada da distribuição de probabilidades conjunta é representada pelo produto dos elementos apropriados (de acordo com o relacionamento entre as variáveis) das tabelas de probabilidades condicionais na RB, ou seja, as TPCs estabelecem uma representação simplificada da DPC. (ROCHA, 2009).

As redes bayesianas podem ser obtidas a partir dos dados, via de regra, considerando dois paradigmas: busca e pontuação e baseado em independência condicional. Um dos métodos que implementam o primeiro paradigma e mais amplamente utilizado é o implementado pelo algoritmo K2. Esse algoritmo permite encontrar a mais provável estrutura de rede de crença  $S$  a partir de um determinado conjunto de dados  $D$ . O algoritmo K2 aplica a pontuação Bayesiana segundo a equação (2).

$$P(S | D) = \prod_{i=1}^n \prod_{j=1}^{q_i} \frac{\Gamma(r_i)}{\Gamma(r_i + N_{ij})} \prod_{k=1}^{r_i} \Gamma(N_{ijk} + 1) \quad (2)$$

onde:

$n$  é o número de nós;

$q_i$  é o número de configurações dos pais da variável  $X_i$ ;

$r_i$  é o número de possíveis valores do nó  $X_i$ ;

$N_{ijk}$  é o número de casos em  $D$  onde o atributo  $X_i$  é instanciado com o seu valor  $k$ , e a configuração dos pais de  $X_i$  é instanciada com o valor  $j$ ;

$N_{ij}$  denota o número de observações em que a configuração dos pais de  $X_i$  é instanciada com o

valor  $j$ , sendo  $N_{ij} = \sum_{k=1}^{r_i} N_{ijk}$ .

Após a construção da rede, são necessários mecanismos de inferência para computar as eventuais probabilidades de interesse. Essas inferências correspondem ao cálculo da distribuição de probabilidade condicional de um conjunto de variáveis de uma rede bayesiana, chamadas variáveis de consulta, dado os estados de outro conjunto de variáveis, denominadas variáveis de evidência, que compõe esse modelo. Os mecanismos de inferência são classificados em métodos exatos, aproximados e simbólicos. A escolha da classe do método depende fundamentalmente do domínio de aplicação, bem como considera, para efeito do levantamento da complexidade e, em consequência, do custo computacional, os seguintes fatores:

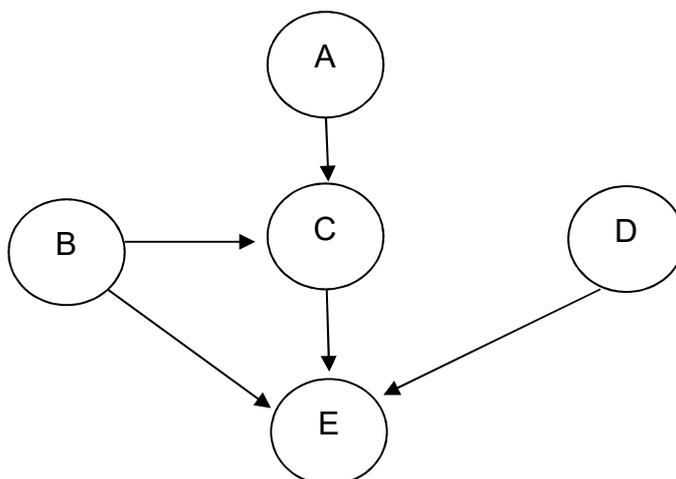
- a estrutura da rede, isto é o número de nós e arcos;
- a cardinalidade das variáveis (número de estados das variáveis);
- o número de *clique tables*( $CT$ ), isto é, a soma do número de entradas das tabelas de probabilidades condicionais associadas a todos os nós da RB. O número de  $CT$  de um nó  $N$  é dado por:

$$CT_N = NE \times PE_1 \times PE_2 \dots PE_m$$

Na qual,  $NE$  é o número de estados da variável  $N$  e  $PE_1, PE_2 \dots PE_m$  é o número de estados de cada uma das variáveis cujos nós, são pai de  $N$ .

Utiliza-se, ainda, como critério de medida de complexidade, o valor máximo de  $CT$ , representado por  $MAX(CT)$  de uma rede bayesiana, isto é, o nó que apresenta o maior valor de  $CT$  na rede bayesiana.

Rocha 2009, elaborou um modelo de rede bayesiana para entendimento da métrica de complexidade, com 5 nós, representados por 4 variáveis binárias  $A, B, C, D$  e pela variável  $E$ , que possui 4 estados.



**Figura 3.** Rede bayesiana utilizada para instanciar as métricas de complexidade.

Ainda segundo (ROCHA, 2009), pode-se observar que, na RB da (**Figura 3**), por exemplo, a variável *C* tem cardinalidade 2 e a variável *E* possui cardinalidade 4. Além disso, é possível dizer que o número de *CT* do nó *C* é 8 ( $2 \times 2 \times 2$ ) e do nó *E* é igual a 32 ( $2 \times 2 \times 2 \times 4$ ).

### 2.3 TÉCNICAS COMPUTACIONAIS PARA APOIO A DECISÃO EM SAÚDE COM ENFASE NAS REDES BAYESIANAS.

A saúde hoje, em razão das necessidades de soluções urgentes para controle ou extermínio de doenças está sendo a cada dia, vem se tornando mais uma fonte de dedicação total das pesquisas no mundo, utilizando como ferramenta as técnicas computacionais para o apoio de tomada de decisões.

(ASSUNÇÃO *et al.*, 1998), desenvolveram métodos estatísticos com a finalidade de análise de mapas de taxas de morbidade quando as unidades geográficas possuem pequenas populações de risco, onde adotaram a abordagem Bayesiana e utilizaram métodos computacionais intensivos para estimação de riscos de cada área. O principal objetivo é separar a variabilidade das taxas devida às diferenças entre regiões do risco subjacente daquela devida à pura flutuação aleatória. As estimativas de risco possuem um erro quadrático médio total menor que as estimativas atuais, esses novos métodos foram aplicados para estimar o risco de mortalidade infantil nos municípios de Minas Gerais, em 1994.

Segundo (WAGNER, 1998), as medidas de frequência de doença se enquadram em duas grandes categorias: prevalência e incidência. A prevalência determina a proporção de indivíduos que tem a doença em um determinado momento no tempo. A incidência prediz o risco de que um indivíduo saudável venha a desenvolver a doença em um período de tempo. A prevalência está relacionada com a incidência na medida em que é proporcional ao produto da incidência com a média da doença. As medidas de frequência, ou seja, prevalência e incidência, são fundamentais para o estudo da ocorrência das doenças.

Segundo (ANDRADE, 1999), a análise Bayesiana seria, pois, uma teoria de decisão estatística para cálculo de probabilidade de uma proposição, com base na probabilidade original e nas novas relevâncias. Expresso em termos clínicos é o conceito segundo o qual o valor preditivo de um teste ou achado clínico depende não apenas da sua sensibilidade e especificidade, mas também da probabilidade prévia (ou seja, da prevalência da doença na população estudada). É de importância histórica ressaltar que o trabalho pioneiro sobre a aplicação das redes bayesianas em Medicina foi publicado em 1959 por um especialista em informática, Lusted, e um médico radiologista, Lusted, este último previamente envolvido nos trabalhos que deram origem aos conceitos médicos de sensibilidade e especificidade. Um aspecto inerente à fórmula bayesiana é a possibilidade de ser aplicada sequencialmente. Em outras palavras, após aplicar ao resultado de um teste à fórmula bayesiana, o valor preditivo obtido passa a ser a nova probabilidade de ocorrência da doença. A partir dela novas probabilidades deverão ser calculadas face a outros testes ou achados. O único pressuposto para essa aplicação sequencial é que os testes sejam independentes entre si, ou seja, que o resultado de um não interfira no do outro.

(DIAS *et al.* 2000), implementaram a Rede Bayesiana na modelagem de dados, a partir do *shell NETICA* da *Norsys*. Os resultados deste demonstraram que na análise das variáveis grau de assertividade do sistema criado foi de cerca de 90% dos casos de identificação de partos prematuros e de recém nascidos de baixo peso. Com isso, foi estimada uma redução nas duas variáveis, além do programa experimentado trazer uma redução financeira em função da diminuição da permanência e utilização de UTI Neonatal e demais custos relacionados a uma gestante de alto risco.

(KOEHLER, 2002), por meio de uma rede bayesiana, modelou um diagnóstico de desnutrição infantil, utilizando as variáveis: idade, dados antropométricos e sinais clínicos.

Observou que a tecnologia de redes bayesianas representa uma promissora técnica para suporte em decisões clínicas, pois tem representação flexível que permite especificar dependências e independências das variáveis, de uma forma simples, através da topologia da rede. A tecnologia ainda oportuniza a utilização da Teoria das Probabilidades, que é uma ferramenta muito importante na área da saúde, que geralmente possui informações incompletas e ao utilizá-la, representando a incerteza, proporciona consistência para fazer interpretações e ser utilizada como um sistema de apoio às tomadas de decisão em saúde.

(VICARI *et al.*, 2002), afirma que com o avanço da pesquisa, foram desenvolvidos sistemas que utilizam o formalismo de redes bayesianas para tratamento da incerteza. Redes bayesianas (BN), redes de Markov (MN) e diagramas de influências (ID) são redes probabilísticas. A característica principal das redes probabilísticas é a habilidade para explorar a estrutura do grafo e reduzir o cálculo (da probabilidade condicional de um evento, dada a evidência disponível) a uma série de cálculos locais, usando somente variáveis obtidas de um nó e seus vizinhos em uma estrutura de grafo, evitando calcular a função de distribuição de probabilidades conjunta global. Esses sistemas representam o conhecimento do domínio da aplicação através de redes bayesianas. Integrou-se ao projeto, visto que o escopo de sua pesquisa refere-se ao desenvolvimento de uma arquitetura para tomada de decisão em ambiente com incerteza, baseada no paradigma de redes probabilísticas, em especial na inferência (avaliação) de redes bayesianas com algoritmos baseados em árvore de junções. Estes sistemas utilizam representações gráficas de dependências probabilísticas.

A utilização das redes bayesianas na aplicação do estudo de doenças e endemias tem se mostrado uma forte parceira na tomada de decisão quanto aos procedimentos, norteando as pesquisas de saúde nas suas tomadas de decisões.

Para (SAHEKI, 2005), existem duas metodologias para a criação dos modelos qualitativo e quantitativo: uma é utilizar-se de conhecimentos de especialistas humanos e traduzir tal conhecimento para uma forma computacional tratável; outra é utilizar-se de uma base de dados de casos anteriores, a qual é usada por um algoritmo de aprendizado para a construção do sistema. A utilização de aprendizado tem como maior desvantagem a necessidade de uma grande base de dados pertinentes ao problema, o que pode ser caro ou impossível. Uma vantagem do aprendizado é limitar a influência de opiniões particulares dos especialistas colaboradores. Também é possível a utilização do conhecimento de especialistas juntamente com a base de dados. Em situações desse tipo, o aspecto qualitativo recebe maior participação

dos especialistas enquanto que o aspecto quantitativo é obtido através da base de dados. O trabalho aqui apresentado se iniciou a partir de interações entre a Escola Politécnica da Universidade de São Paulo (EPUSP) e o Hospital Universitário da USP (HU-USP), resultando em um projeto conjunto entre as duas instituições, integrando conhecimentos de inteligência artificial e de medicina. Essa colaboração Inter unidades na USP visou a avaliar o benefício que a aplicação de tecnologias de inteligência artificial pode trazer à prática médica.

(TENÓRIO, 2011), desenvolveu um sistema de apoio à decisão médica para reconhecimento de casos de doença celíaca utilizando cinco técnicas de inteligência artificial: árvores de decisão; classificador bayesiano; k-vizinhos próximos; máquinas de vetor de suporte e redes neurais artificiais. Os resultados demonstraram que o classificador desenvolvido foi eficaz no diagnóstico de sensibilidade da doença celíaca.

(GIRARDI *et al.*, 2014), realizaram um trabalho comparativo entre uma doença estigmatizante (Hanseníase) e uma não estigmatizante (diabetes), utilizando as redes bayesianas (RB) para analisar dados referentes à participação social e a limitação de atividades em pessoas atingidas por Hanseníase e por diabetes e correlacioná-las no intuito de compreender o desenvolvimento da autoestima. A partir da elaboração e análise das redes bayesianas com os dados de Hanseníase e Diabetes, foi possível associar a variável autoestima às variáveis sexo, pessoas novas e opinião e foi possível identificar que a autoestima está intimamente relacionada com as limitações que o paciente encontra nas doenças estudadas.

(MELINSKI *et al.*, 2014), realizaram estudo comparativo entre diabetes e Hanseníase, doenças que quando não controladas são incapacitantes, utilizando a técnica de redes bayesianas. Foi verificado que, em relação ao total de pacientes, o número de pessoas que desenvolveram algum tipo de incapacidade é pequeno, mas não deixa de ser relevante. As incapacidades refletem na autoestima e participação social. O estigma, mais presente na Hanseníase, tem importância destacada, principalmente em relação ao sexo, em que homens apresentam maiores prejuízos, dificultando a interpretação da cura.

Para (BRITO *et al.*, 2015), existe uma associação entre a gravidez na adolescência e variáveis representativas das dimensões de renda, educação e acesso às TIC nos municípios brasileiros, realizadas através da utilização das redes bayesianas. As análises das variáveis, utilizando-se as redes bayesianas demonstraram que a ocorrência maior da gravidez na

adolescência está associada ao baixo acesso à informação e os autores sugerem a utilização das TIC pelos gestores da saúde na utilização de projetos educativos e de promoção de saúde.

(HÔHER *et al.*, 2015), desenvolveram um sistema baseado na tecnologia de desenvolvimento de sistemas de apoio a decisão médica, baseados no paradigma de redes probabilísticas, em especial na avaliação de redes bayesianas com algoritmos baseados em árvore de junções, usando como exemplo o domínio de conhecimento em cardiopatias congênitas. O sistema desenvolvido demonstrou ser eficaz no apoio à decisão na área médica.

Segundo (SATO *et al.*, 2015), as tomadas de decisões são um aspecto fundamental na conduta de um diagnóstico ou tratamento. A ampla difusão dos sistemas computacionais e dos bancos de dados permite sistematizar, por meio do uso da inteligência artificial, parte dessa tomada de decisão. Neste texto, é apresentada, de modo básico, a possibilidade de uso dos modelos gráficos probabilísticos como ferramenta de análise na causalidade das condições de saúde. Essa metodologia vem sendo utilizada para diagnósticos da doença de Alzheimer, apneia do sono e doenças cardiológicas. Entende-se uma rede bayesiana como um modelo no qual a causalidade é importante, porém não é raro ter-se um entendimento incompleto sobre o que está acontecendo e, por isso, tentar-se descrever probabilisticamente essas relações. Desse modo, o aspecto probabilístico ganha uma importante dimensão nessa rede de relações na tentativa de superar a limitação do conhecimento e auxiliar o processo de tomada de decisão. As redes bayesianas são gráficos direcionais acíclicos (GDA), nos quais os nós, são variáveis aleatórias e as premissas de independência entre as variáveis são mantidas. Os nós do gráfico representam as variáveis aleatórias da rede bayesiana e podem variar quanto à sua natureza. Isso oferece a flexibilidade de incluir dados como quantidades observadas, variáveis latentes, parâmetros desconhecidos ou até mesmo hipóteses levantadas pelo pesquisador.

(SILVA, 2018), utilizou as Redes Neurais Artificiais (RNA), na previsão de novos casos de dengue na região metropolitana de Belém-Pará-Brasil. O estudo se mostra eficiente para previsões de novos casos, usando séries temporais, podendo prever com erros médios de 0.0347, servindo para tomada de decisão dos gestores de saúde pública. A pesquisa realizada, bem como o sistema de monitoramento desenvolvido, mostra-se útil na detecção precoce e alerta no aumento de novos casos de dengue, que podem ser usados por órgãos de saúde pública na tomada de decisão e ações de controle das epidemias em seus municípios.

**Quadro 1** - Distribuição dos estudos analisados de Redes Bayesianas quanto ao autor, ano e contribuição 2019.

<b>Autor(es)</b>	<b>Ano</b>	<b>Contribuição</b>
COOPER E HERSKOVITS	1992	Aplicação do algoritmo de busca heurística K2 para aprendizado da rede bayesiana.
AGRAWAL <i>et al.</i>	1993	Modelo de suporte de confiança que permite avaliar a qualidade da regra de associação.
PONNIGHAUS <i>et al.</i>	1994	Inserção da variável que considera domicílios com densidade maior que 2.
ASSUNÇÃO <i>et al.</i>	1998	Desenvolveram métodos estatísticos com a finalidade de análise de mapas de taxas de morbidade quando as unidades geográficas possuem pequenas populações de risco, onde adotaram a abordagem Bayesiana.
WAGNER	1998	Enquadramento das medidas de frequência em duas categorias: prevalência e incidência.
PEARL	1988	Conceitua redes bayesianas como modelos probabilísticos que permitem lidar de forma rigorosa com a representação de conhecimentos em domínios onde existe incerteza.
ANDRADE	1999	A análise Bayesiana seria como uma teoria de decisão estatística para cálculos de probabilidade de uma proposição, com base na probabilidade original e nas novas relevâncias.
DIAS <i>et al.</i>	2000	Implementaram a Rede Bayesiana na modelagem de dados, a partir do <i>shell NETICA</i> da <i>Norsys</i> .
CHEN	2001	As redes bayesianas permitem construir modelos formados por uma estrutura qualitativa, definida por grafos acíclicos dirigidos, onde são representadas as dependências entre os nós.
KOEHLER	2002	Utilizando a rede bayesiana, modelou um diagnóstico de desnutrição infantil, utilizando as variáveis: idade, dados antropométricos e sinais clínicos.
RUSSEL E NORVIG	2002	Redes bayesianas (RB) são modelos probabilísticos utilizados para descrever a relação entre variáveis aleatórias.
VICARI <i>et al.</i>	2002	Foram desenvolvidos sistemas que utilizam o formalismo de redes bayesianas para tratamento da incerteza.
SAHEKI	2005	Existência de duas metodologias para a criação dos modelos qualitativos e quantitativos.
KERR-PONTES <i>et al.</i>	2006	Inserção da variável que considera domicílios com densidade maior que 2.

ROCHA	2009	Entrada da distribuição de probabilidades conjunta é representada pelo produto dos elementos apropriados (de acordo com o relacionamento entre as variáveis) das tabelas de probabilidades condicionais na RB.
JENSEN IN FISCHER	2010	Uma rede bayesiana consiste em um conjunto de variáveis e arcos ligando as variáveis.
TENÓRIO	2011	Desenvolveu um sistema de apoio à decisão médica para reconhecimento de casos de doença celíaca utilizando cinco técnicas de inteligência artificial.
SALES <i>et al.</i>	2011	Inserção da variável que considera domicílios com densidade maior que 2.
SUZUKI	2012	Inserção da variável que considera domicílios com densidade maior que 2.
MOURA <i>et al.</i>	2013	Inserção da variável que considera domicílios com densidade maior que 2.
GIRARDI <i>et al.</i>	2014	Trabalho comparativo entre uma doença estigmatizante (Hanseníase) e uma não estigmatizante (diabetes), utilizando as redes bayesianas (RB) para analisar dados referentes à participação social.
SEFFRIN E JACQUES	2014	Cada variável possui um conjunto finito de estados que pode assumir
MELINSKI, <i>et al.</i>	2014	Realizaram estudo comparativo entre diabetes e Hanseníase, doenças que quando não controladas são incapacitantes, utilizando a técnica de redes bayesianas.
BRITO <i>et al.</i>	2015	Existência de uma associação entre a gravidez na adolescência e variáveis representativas das dimensões de renda, educação e acesso às TIC nos municípios brasileiros.
KÔHER <i>et al.</i>	2015	Desenvolveram um sistema baseado na tecnologia de desenvolvimento de sistemas de apoio a decisão médica, baseados no paradigma de redes probabilísticas.
SATO <i>et al.</i>	2015	A ampla difusão dos sistemas computacionais e dos bancos de dados permite sistematizar, por meio do uso da inteligência artificial, parte dessa tomada de decisão.
SILVA	2018	Utilizou as Redes Neurais Artificiais (RNA), na previsão de novos casos de dengue na região metropolitana de Belém-Pará-Brasil.

**Fonte:** Características extraídas dos estudos publicados no período de 1992 a 2018.

#### 2.4. A HANSENÍASE ESTUDADA E VISTA NESSE CONTEXTO

De acordo com (ANDRADE *et al.*, 1994), os resultados de indicadores epidemiológicos, o Brasil ainda apresenta altas taxas de detecção de Hanseníase, assim como elevados

coeficientes de prevalência, ao contrário da tendência global, cujos indicadores de transmissão apontam comportamento decrescente da endemia. Em algumas áreas do país, a Hanseníase mostra magnitudes equivalentes às mais altas do mundo, mesmo quando comparadas com alguns países da África e Ásia. Diversos estudos dos prováveis determinantes da Hanseníase foram realizados a partir de 1936, acompanhando a produção científica no campo da epidemiologia desde a análise da distribuição dos atributos, definição dos indicadores e do conceito de risco até a utilização de modelos e técnicas estatísticas mais recentes, por meio do aperfeiçoamento dos desenhos de pesquisa e da utilização intensiva da informática na análise dos dados. Segundo o autor, no Brasil, não foram até hoje desenvolvidos estudos do tipo longitudinal, visando à identificação dos fatores associados. Entretanto, na busca de determinantes dessa doença, quatro grandes estudos transversais, a partir da década de 1940, contribuíram de forma importante para o conhecimento epidemiológico da endemia.

Segundo (MATOS, 1999), o Brasil se inclui entre os países de alta endemicidade de Hanseníase no mundo. Isto significa que apresenta um coeficiente de prevalência médio superior a um caso por mil habitantes (MS, 1989). Embora, geograficamente, esse coeficiente seja desigual, com áreas de alta, média e baixa prevalência distribuídas em todo o território do país, há uma tendência uniforme de crescimento no número de casos diagnosticados desde os anos 70 até agora. E, atualmente, a Hanseníase representa uma das poucas doenças infecciosas com elevação de incidência em nosso país (IBGE, 1995), uma tendência contrária àquela que vem ocorrendo em termos mundiais (OMS, 1998). Diante desse quadro, medidas de prevenção e controle, além do tratamento eficaz dos casos, são absolutamente necessárias. Enquanto uma vacina efetiva para Hanseníase não se encontra disponível, programas de vigilância epidemiológica assumem importância central no controle e prevenção da doença (MS, 1989). Como a transmissão do *mycobacterium leprae* ocorre de indivíduo para indivíduo, os contatos intradomiciliares constituem um grupo de risco.

(LAPA *et al.*, 2001), realizaram um estudo de análise espacial da ocorrência de Hanseníase na cidade de Olinda-PE, obtido por meio da “situação de risco coletivo” e de indicadores de carência social e/ou relacionados à probabilidade de adoecer: renda, educação, abastecimento d’água, habitantes por dormitório, habitações em aglomerados subnormais e densidade demográfica. Os resultados do estudo indicam que a distribuição espacial heterogênea da Hanseníase em Olinda não é aleatória, identificando-se um padrão de agregação

no espaço que está associado às condições de vida da população, expresso mediante o indicador de carência social.

Segundo (BARROS, *et al.*, 2006), os inquéritos de saúde de base populacional constituem o principal instrumento utilizado para conhecer a prevalência de doenças crônicas, de restrições de atividades e de uso de serviços de saúde. Concluiu que no padrão de doze doenças crônicas analisadas com base no PNAD, segundo sexo, cor/raça, nível de escolaridade, região de residência e situação de domicílio e a presenças de limitações das doenças e o uso dos serviços de saúde, significativas desigualdades sociais foram detectadas na magnitude das prevalências e no perfil de morbidade.

(IMBIRIBA *et al.*, 2009), relata que na América Latina, o Brasil registra o maior número de casos de Hanseníase e a eliminação da endemia constitui um desafio à saúde pública, cuja meta é a redução do coeficiente de prevalência para menos de um caso para cada dez mil habitantes. Em comparação aos demais estados da região Norte, o Amazonas registrou os menores coeficientes de prevalência (2,30/10.000) e de detecção (3,23/10.000) em 2004. Porém, de acordo com os parâmetros do Ministério da Saúde, tal situação ainda corresponde à média e alta endemicidade, respectivamente. Em Manaus, a detecção ultrapassou o valor nacional daquele ano, mesmo tendo reduzido de 10,42 para 3,10/10.000 habitantes entre 1987 e 2004. Assim, a cidade persiste como área de alta endemicidade, responsável por 46,2% dos casos novos e por 42,3% dos casos em registro ativo de todo o estado. O estudo concluiu que a distribuição da doença na cidade estudada é heterogênea, atingindo mais intensamente as regiões habitadas por grupos em situação de maior vulnerabilidade.

De acordo com (FUNDAÇÃO PRÓ-HANSEN, 2011), o Brasil ocupa o segundo lugar no ranking geral de casos descobertos anualmente - uma média de 49 mil na última década. Só perde para a Índia, país asiático com densidade populacional 15 vezes maior e duas vezes mais pobre. E até a Índia cumpriu a meta definida pela Organização Mundial da Saúde (OMS) para a eliminação da Hanseníase. Das 193 nações do globo, apenas quatro não reduziram as taxas à média aceitável de um caso a cada 10 mil habitantes - Brasil, Nepal, Moçambique e Congo. A posição do país deve ser motivo de vergonha. Não dos vitimados. Dos governantes. Presidentes, governadores e prefeitos das últimas cinco décadas.

A situação do Brasil é alarmante, principalmente porque há muitos registros da doença em crianças e adolescentes com menos de 15 anos, totalizando 2.420 casos. De acordo com

dados do Ministério da Saúde, em 2012 foram detectados quase 29 mil casos de Hanseníase no país, dos quais 1.936 em menores de 15 anos.

De acordo com o Ministério da Saúde, os casos de Hanseníase diminuíram 26% entre 2001 e 2011. No entanto, a queda da doença no resto do mundo foi muito mais acentuada, já que, segundo a OMS, em um período de seis anos (entre 2004 e 2010) houve uma redução de 40% nos casos da doença em todo o mundo. Para reduzir essas estatísticas é preciso que os três níveis de governo intensifiquem suas ações na prevenção e no tratamento da doença.

A Hanseníase apresenta tendência de redução de casos no Paraná nos últimos anos, foram notificados no Sistema de Informação de Agravos de Notificação – SINAN em 2003, 1.722 casos novos e 996 em 2012, tendo uma redução significativa em número de casos.

A Fundação Pró-Hansen 2011, em parceria com as Secretarias Estadual e Municipal de Saúde tem dado atenção especial às ações voltadas a expansão do acesso às oportunidades de diagnóstico, tratamento e vigilância. O controle da Hanseníase é baseado no diagnóstico precoce, tratamento e a alta. A intensificação das ações de controle da Hanseníase está relacionada ao comprometimento dos profissionais de saúde, da vontade política dos gestores nas três esferas de governo e da mobilização da população.

A estratégia prioritária adotada pela Fundação Pró-Hansen é a atenção integral e humanizada para as pessoas atingidas pela Hanseníase, além de desenvolver articulações ampliando as parcerias, destacando-se o desenvolvimento de campanhas informativas e educativas e de capacitações para profissionais de saúde.

Para (FREITAS *et. al.*, 2014), a maior taxa de incidência da Hanseníase, foi encontrada nos municípios com pior status socioeconômico, onde recomenda que os esforços dos programas de controle, sejam direcionados para estas áreas que são consideradas altamente endêmico. Existe claramente a necessidade de intensificar as medidas visando o diagnóstico precoce da doença, bem como ações para preveni-lo e controlá-lo, levando em consideração desigualdades sociais existentes. Sugere que o serviço de saúde e ações de programas sociais sejam integradas. Melhorando as condições de vida da população, estas são fundamentais para interromper a transmissão contínua da lepra e, conseqüentemente, sua incidência poderia ser reduzida.

Na 62ª Jornada Mundial da Hanseníase, o infectologista francês (JALIER, 2015), revela que mais de 200.000 novos casos de Hanseníase são diagnosticados a cada ano no mundo. O documento afirma que segundo a OMS, a Índia é o país mais afetado, seguido do Brasil,

Indonésia, Etiópia e República Democrática do Congo. Causada por uma microbactéria transmitida pelas vias respiratórias, a Hanseníase é uma doença infecciosa conhecida desde a Antiguidade, mas que desapareceu espontaneamente dos países ocidentais paralelamente à melhoria da qualidade de vida e do nível de higiene. Sua evolução extremamente lenta - a incubação dura entre um e dez anos, às vezes mais - faz com que seja uma doença difícil de controlar nos países mais pobres, no sul da Ásia, África ou América Latina, onde é endêmica.

Para (OLIVEIRA *et al.*, 2015), os indicadores epidemiológicos sinalizaram redução no diagnóstico da Hanseníase nos municípios de Foz do Iguaçu e Londrina, no estado do Paraná. O comportamento desses coeficientes pode indicar a redução da transmissibilidade da doença, visto que o coeficiente de detecção elevado entre menores de 15 anos nos países endêmicos assinala continuidade da transmissão do bacilo e inconsistência das atividades de controle. Ao final da série histórica analisada de 2001 a 2010, coeficiente elevado de detecção de casos novos, indicando a necessidade de potencializar as atividades de controle da Hanseníase. Os indicadores operacionais apontaram discreta melhoria na qualidade do atendimento aos portadores da doença. Embora as cidades estudadas tenham apresentado melhora em alguns indicadores, todavia não é esse o resultado para o controle da doença no país.

De acordo com (VIANA *et al.*, 2015), com base na análise de indicadores socioeconômicos, ambientais e sanitários, os resultados demonstraram uma região dinâmica e muito heterogênea, com uma população crescente e concentrada principalmente em áreas urbanas, combinando crescimento econômico e degradação ambiental a um rápido e precário processo de urbanização e aos projetos de desenvolvimento voltados para a exploração de recursos naturais. O resultado é um quadro complexo de problemas de saúde, com expressividade das doenças infecciosas e parasitárias relacionadas às intensas mudanças ambientais e às difíceis condições de vida, decorrentes principalmente da persistência de imensas desigualdades nos indicadores sociais e econômicos. Tais vulnerabilidades produzem efeitos negativos no quadro ambiental e sanitário da região, principalmente na Amazônia Legal.

Conforme (AUGUSTO, 2016), as doenças negligenciadas são agravos que persistem no país e estão relacionadas às condições de pobreza, iniquidades em saúde e exclusão social. São necessárias ações intersetoriais para a prevenção e controle dessas doenças. Nesse sentido, o ambiente escolar pode ser um local privilegiado para ações educativas em saúde, sendo que o professor do ensino fundamental tem um papel importante na construção de saberes dos

educandos, podendo contribuir na realização de atividades educativas em saúde sobre as doenças negligenciadas. Augusto (2016) conclui que há necessidade de um maior investimento em educação e saúde sobre as doenças negligenciadas com foco nos professores do município para que estes possam contribuir de forma efetiva e segura na prevenção e controle destes agravos.

(NEGRÃO *et al.*, 2016), a Hanseníase se distribui de forma aleatória e focal não obedecendo a critérios pré-estabelecidos, sendo importante conhecer a geografia da doença. O estudo da distribuição espacial da Hanseníase permite a identificação de grupos residentes em áreas que oferecem maior risco de adoecer, de forma a poder auxiliar no planejamento, implementação, monitoramento e avaliação de ações voltadas para a prevenção e controle da doença. A aglomeração domiciliar como variável ecológica, associada às precarização das condições de saneamento básico é um indicador socioeconômico direto na ocorrência da Hanseníase, onde a baixa qualidade da moradia nas periferias do município de Guarapuava-PR, caracterizadas por escassa infraestrutura urbana mantém a possível persistência da endemia nos bolsões de pobreza urbana, favorecendo a manutenção do bacilo viável no ambiente por maior tempo.

Para (CHAVES *et al.*, 2017), a Hanseníase é uma doença infectocontagiosa, de evolução lenta, que se manifesta principalmente por meio de sinais e sintomas dermatoneurológicos. Quando diagnosticada e tratada tardiamente, pode gerar incapacidades físicas nas mãos, pés e olhos, resultantes do comprometimento dos nervos periféricos; em estágios mais avançados de manifestação clínica, pode-se observar comprometimento neural troncular, capaz de trazer repercussões como parestesias e plegias musculares. Estima-se que entre 1 a 2 milhões de pessoas no mundo possuam deformidades e deficiências resultantes da Hanseníase. Em virtude dessas deformidades e deficiências, frequentemente, pessoas com Hanseníase são vítimas de estigma e discriminação. O estigma, um fator importante relacionado ao atraso do diagnóstico, facilita a transmissão da infecção na família e entre a comunidade, razão porque foi introduzido na “Estratégia Global para a Hanseníase 2016-2020” como um indicador para monitorar a discriminação das pessoas afetadas pela doença. Mesmo sendo uma doença antiga, a Hanseníase permanece como um problema de Saúde Pública, uma vez que o controle da transmissão é difícil de ser alcançado em alguns países, incluindo o Brasil. Em 2015, foram diagnosticados 210.758 casos novos em 136 países, constituindo uma taxa de detecção mundial

de 3,2 por 100 mil habitantes. O Sudeste Asiático foi a região que notificou maior número de casos novos (156.118), seguido pelas Américas (28.808) e África (20.004).

Ainda segundo os autores acima, 3.358 novos casos de Hanseníase foram relatados no Pará, com 17,5% dos municípios do estado sendo considerados hiperendêmicos e tendo uma relação direta entre o índice de privação social e o índice de detecção da Hanseníase.

Segundo (FREITAS *et al.*, 2017), no Brasil, a distribuição espacial da Hanseníase é heterogênea: os Estados mais desenvolvidos socioeconomicamente da Região Sul alcançaram a meta de eliminação da Hanseníase como problema de saúde pública — prevalência de menos de 1 caso para 10.000 habitantes. Porém, bolsões de alta carga da doença ainda permanecem nas Regiões Norte, Centro-Oeste e Nordeste do Brasil, consideradas as áreas de maior transmissão da doença no país. Uma análise de grupo, realizada em 2009 pelo Ministério da Saúde, mostrou que os Estados do Mato Grosso, do Tocantins, de Rondônia, do Pará e do Maranhão pertencem a uma área com alto risco de transmissão persistente de Hanseníase e que a persistência geográfica e temporal aponta para a necessidade de buscar novas estratégias de controle nessas áreas onde existe sobreposição de riscos.

(GRACIE *et al.*, 2017), cita que a Hanseníase ainda constitui um relevante problema de saúde pública. O Brasil se mantém como segundo país com maior número de casos novos detectados no mundo, precedido pela Índia. Embora a detecção anual venha declinando, no início de 2010 a prevalência global de Hanseníase era de cerca de 212 mil casos. No Brasil, em 2014, dos 31.064 casos novos detectados, 2.341 foram em menores de 15 anos e o coeficiente de detecção geral foi de 15,32 para cada 100 mil habitantes. Alguns municípios no Brasil já alcançaram a meta de eliminação da doença, no entanto, as regiões Norte, Centro-Oeste e Nordeste ainda apresentam altos patamares de casos da doença. Apesar da tendência de estabilização dos coeficientes de detecção no Brasil, os coeficientes apresentam grande variação nas diferentes regiões brasileiras. Fatores como densidade populacional, hábitos de vida, aspectos culturais, condições sanitárias e de moradia devem ser ponderados em um cenário desfavorável como o apresentado pelo estado do Rio de Janeiro. Estudos demonstram que a Hanseníase é fortemente relacionada com as condições de vida e pobreza. Evidencia-se uma tendência de concentração dos doentes em camadas da sociedade menos favorecidas.

Para (BARRETO, 2018), no Brasil, são diagnosticados normalmente de 20 a 30 mil novos casos por ano. É uma doença extremamente comum. O Brasil é o país que ocupa o

segundo lugar no mundo em matéria da Hanseníase. Nós temos uma população desfavorecida muito grande. É uma doença ligada drasticamente a 70% de uma classe social desfavorecida. São pessoas que normalmente não têm acesso à saúde de qualidade e que estão em um nível socioeconômico cultural muito baixo. Elas não procuram assistência médica na fase inicial, que é facilmente confundida com o famoso “pano branco”. Quando uma criança apresenta sintomas de “pano branco” em uma área onde há Hanseníase ou que tenha um parente com Hanseníase, é preciso verificar se não se trata de uma lesão inicial de Hanseníase. O problema é que muitos médicos não aprendem sobre Hanseníase na faculdade. Eu me formei na Universidade Federal de Santa Catarina e nem sabia que a Hanseníase existia no Brasil. Como pensar no diagnóstico se nós achamos que não existe? Além do mais, a medicação é gratuita e está disponível no SUS.

Segundo (RODRIGUEZ, 2018), os estados de Mato Grosso, Pará, Maranhão, Tocantins, Rondônia e Goiás são as áreas com maior risco de transmissão, concentrando mais de 80% do total de casos diagnosticado, reforçando a desigualdade social entre as regiões do país e a incidência da Hanseníase.

**Quadro 2** - Distribuição dos estudos analisados de Hanseníase quanto ao autor, ano e contribuição 2019.

<b>Autor(es)</b>	<b>Ano</b>	<b>Contribuição</b>
ANDRADE <i>et al.</i>	1994	Os resultados de indicadores epidemiológicos, o Brasil ainda apresenta altas taxas de detecção de Hanseníase, assim como elevados coeficientes de prevalência, ao contrário da tendência global, cujos indicadores de transmissão apontam comportamento decrescente da endemia.
FREEDMAN	1999	A falácia ecológica consiste em pensar que as relações observadas para grupos necessariamente se mantêm para os indivíduos.
MATOS	1999	O Brasil se inclui entre os países de alta endemicidade de Hanseníase no mundo. Isto significa que apresenta um coeficiente de prevalência médio superior a um caso por mil habitantes.
ONU	2000	Objetivos do combate as doenças ameaçam acima de tudo as populações mais pobres.
LAPA <i>et al.</i>	2001	Realizaram um estudo de análise espacial da ocorrência de Hanseníase na cidade de Olinda-PE, obtido por meio da “situação de risco coletivo”.

BARROS <i>et al.</i>	2006	Inquéritos de saúde de base populacional constituem o principal instrumento utilizado para conhecer a prevalência de doenças crônicas, de restrições de atividades e de uso de serviços de saúde.
FRED	2006	Os sistemas baseados em redes bayesianas são capazes de gerar automaticamente previsões ou decisões mesmo na situação de inexistência de algumas peças de informação.
IMBIRIBA <i>et al.</i>	2009	Na América Latina, o Brasil registra o maior número de casos de Hanseníase e a eliminação da endemia constitui um desafio à saúde pública, cuja meta é a redução do coeficiente de prevalência para menos de um caso para cada dez mil habitantes.
ARA-SOUZA	2010	Como as redes bayesianas são abordagens interpretativa e analítica para probabilidades, é também um método de modelagem e de decisão.
PRÓ-HANSEN	2011	O Brasil ocupa o segundo lugar no ranking geral de casos descobertos anualmente - uma média de 49 mil na última década.
SALES <i>et al.</i>	2011	a inserção da variável que considera Domicílios com Densidade maior que dois.
BARRETO	2013	No Brasil, apesar da redução significativa do perfil de mortalidade por doenças infecciosas, algumas doenças ainda se manifestam sob a forma epidêmica.
BRITO <i>et al.</i>	2014	A maior incidência da Hanseníase se dá em classes socioeconômicas baixas, devido à multiexposição, baixo nível de formação, nutrição e moradia.
FREITAS <i>et al.</i>	2014	A maior taxa de incidência da Hanseníase, foi encontrada nos municípios com pior status socioeconômico.
MELINSKI E GIRARDI	2014	O estigma, mais presente na Hanseníase, tem importância destacada, principalmente em relação ao sexo, em que homens apresentam maiores prejuízos, dificultando a interpretação da cura.
OLIVEIRA <i>et al.</i>	2014	Na Hanseníase, o domicílio é apontado como importante espaço de transmissão da doença.
JALIER	2015	Revela que mais de 200.000 novos casos de Hanseníase são diagnosticados a cada ano no mundo.
OLIVEIRA <i>et al.</i>	2015	Os indicadores epidemiológicos sinalizaram redução no diagnóstico da Hanseníase nos municípios de Foz do Iguaçu e Londrina, no estado do Paraná.
VIANA <i>et al.</i>	2015	Com base na análise de indicadores socioeconômicos, ambientais e sanitários, os resultados demonstraram uma região dinâmica e muito heterogênea, com uma população crescente e concentrada principalmente em áreas urbanas.

AUGUSTO	2016	As doenças negligenciadas são agravos que persistem no país e estão relacionadas às condições de pobreza, iniquidades em saúde e exclusão social.
NEGRÃO <i>et al.</i>	2016	Hanseníase se distribui de forma aleatória e focal não obedecendo a critérios pré-estabelecidos, sendo importante conhecer a geografia da doença.
CHAVES <i>et al.</i>	2017	Hanseníase é uma doença infectocontagiosa, de evolução lenta, que se manifesta principalmente por meio de sinais e sintomas dermatoneurológicos.
BASTOS E OLIVEIRA	2017	São denominadas variáveis <i>Pais</i> aquelas causais (ou montante), e de nós <i>Filhos</i> aquelas “sob efeito de” (ou colocadas a jusante).
FREITAS <i>et al.</i>	2017	No Brasil, a distribuição espacial da Hanseníase é heterogênea: os Estados mais desenvolvidos socioeconomicamente da Região Sul alcançaram a meta de eliminação da Hanseníase como problema de saúde pública.
GRACIE <i>et al.</i>	2017	Hanseníase ainda constitui um relevante problema de saúde pública. O Brasil se mantém como segundo país com maior número de casos novos detectados no mundo, precedido pela Índia.
BARRETO	2018	No Brasil, são diagnosticados normalmente de 20 a 30 mil novos casos por ano. É uma doença extremamente comum.
BRITO	2018	Mato Grosso e Tocantins foram os dois estados com maior concentração de casos de Hanseníase em 2017, conforme dados preliminares divulgados pelo Ministério da Saúde
RODRIGUEZ	2018	Os estados de Mato Grosso, Pará, Maranhão, Tocantins, Rondônia e Goiás são as áreas com maior risco de transmissão.

**Fonte:** Características extraídas dos estudos publicados no período de 1994 a 2018.

## 2.5. CONSIDERAÇÕES FINAIS

O presente capítulo demonstrou, através da bibliografia analisada, um grande número de pesquisas em saúde que utilizam cada vez mais a tecnologias das redes bayesianas, com o objetivo de nortear os caminhos que deverão ser adotados no planejamento, monitoramento, avaliação e diagnóstico de diversas doenças e principalmente das negligenciadas por fatores socioeconômicos, e as endêmicas nas regiões periféricas do Brasil, como é o caso da Hanseníase, objeto deste estudo.

## **CAPÍTULO 3 – METODOLOGIA DA COLETA DE DADOS SOBRE A SITUAÇÃO SÓCIO-DEMOGRÁFICA E A ABORDAGEM DA UTILIZAÇÃO DAS REDES BAYESIANAS**

### **3.1 - CONSIDERAÇÕES INICIAIS**

O objetivo deste capítulo é descrever a metodologia utilizada para a coleta dos dados sócio-demográficos dos municípios da Amazônia Legal e da incidência da Hanseníase para definir variáveis, utilizando as redes bayesianas para estabelecer a analogia dos dados entre os municípios estudados com os dados encontrados para os demais municípios brasileiros.

### **3.2 – COLETA DE DADOS**

Este estudo utilizou como fontes de dados para a seleção das variáveis, os microdados da base da Pesquisa de Domicílios do Censo Demográfico do Instituto Brasileiro de Geografia e Estatística – IBGE<sup>1</sup>, tendo como referência os municípios brasileiros e o ano de 2010, por este ser o Censo mais recente. A partir do IBGE, foram recuperadas variáveis que discriminam as condições de habitação dos domicílios. No total, a amostra abrangeu 57.320.474 domicílios em 5565 municípios, sendo 771 localizados na Amazônia Legal brasileira.

Foram recuperados 42.017 registros de incidência da doença Hanseníase nos municípios a partir do SINAN - Sistema de Informação de Agravos de Notificação, disponibilizados através do DATASUS – Departamento de Informática do Sistema Único de Saúde<sup>2</sup>, tendo como referência os municípios brasileiros e o ano de 2010, para manter a consistência metodológica.

Adicionalmente, os componentes do Índice de Desenvolvimento Humano Municipal (IDHM), referentes à educação (IDHM-Educação) e renda (IDHM-Renda) foram obtidos a partir da plataforma de consulta Atlas do Desenvolvimento Humano no Brasil<sup>3</sup>, também adotando como referência o ano de 2010.

A base de dados geoespaciais dos municípios brasileiros foi obtida a partir do IBGE<sup>4</sup> em formato de arquivo *shapefile*, que é manipulável pela maioria dos sistemas de informações geográficos.

---

<sup>1</sup> <https://sidra.ibge.gov.br/>

<sup>2</sup> <http://www2.datasus.gov.br>

<sup>3</sup> <http://atlasbrasil.org.br>

<sup>4</sup> [http://downloads.ibge.gov.br/downloads\\_geociencias.htm](http://downloads.ibge.gov.br/downloads_geociencias.htm)

A partir dessas diferentes fontes de dados, foram selecionadas dez variáveis das dimensões de Hanseníase, habitação, renda e educação, apresentadas no (Quadro 3).

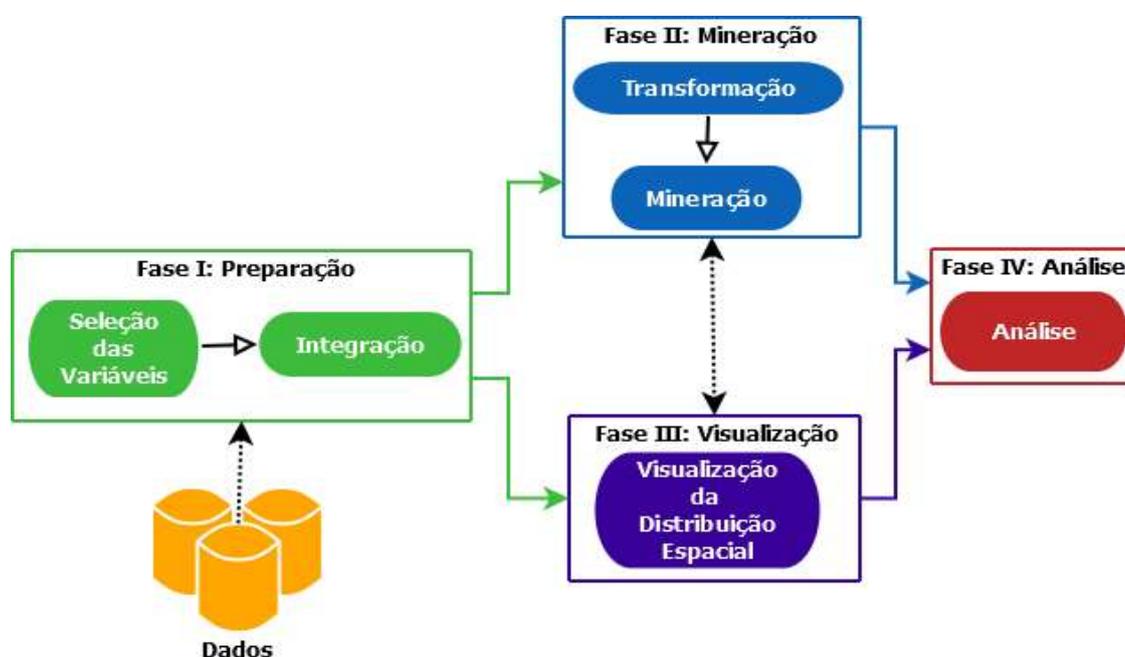
**Quadro 3.** Variáveis selecionadas e fontes de dados.

Variável	Descrição	Fonte de dados
<i>AguaEncanada</i>	Percentual de domicílios com Água Encanada. Calculado como sendo a razão entre a população que vive em domicílios particulares permanentes com água canalizada para um ou mais cômodos e a população total residente em domicílios particulares permanentes, multiplicado por 100. A água pode ser proveniente de rede geral, de poço, de nascente ou de reservatório abastecido por água das chuvas ou carro-pipa	IBGE
<i>BanheiroAguaEncanada</i>	Percentual de domicílios com Banheiro e Água Encanada. Calculado como sendo a razão entre a população que vive em domicílios particulares permanentes com água encanada em pelo menos um de seus cômodos e com banheiro exclusivo e a população total residente em domicílios particulares permanentes multiplicado por 100. A água pode ser proveniente de rede geral, de poço, de nascente ou de reservatório abastecido por água das chuvas ou carro-pipa. Banheiro exclusivo é definido como cômodo que dispõe de chuveiro ou banheira e aparelho sanitário	IBGE
<i>ColetaLixo</i>	Percentual de domicílios com coleta de lixo. Calculado como sendo a razão entre a população que vive em domicílios com coleta de lixo e a população total residente em domicílios particulares permanentes multiplicados por 100. Estão incluídas as situações em que a coleta de lixo realizada diretamente por empresa pública ou privada, ou o lixo é depositado em caçamba, tanque ou depósito fora do domicílio, para posterior coleta pela prestadora do serviço. São considerados apenas os domicílios particulares permanentes localizados em área urbana;	IBGE
<i>EnergiaEletrica</i>	Percentual de domicílios com Energia Elétrica. Calculado como sendo a razão entre a população que vive em domicílios particulares permanentes com iluminação elétrica e a população total residente em domicílios particulares permanentes multiplicado por 100. Considera-se iluminação proveniente ou não de uma rede geral, com ou sem medidor.	IBGE
<i>DensidadePopMaior2</i>	Percentual de domicílios com densidade superior a 2 pessoas por dormitório. Razão entre a população que vive em domicílios particulares permanentes com densidade superior a 2 e a população total residente em domicílios particulares permanentes multiplicado por 100. A densidade do domicílio é dada pela razão entre o total de moradores do domicílio e o número total de cômodos usados como dormitório.	IBGE
<i>Hanseníase</i>	Incidência da Hanseníase no município. Calculada como sendo o quociente entre o número de casos multiplicado por 100.000 habitantes e o total da população.	DATASUS
<i>IDHM-Longevidade</i>	Corresponde à dimensão "longevidade" do IDHM. É medida pela expectativa de vida ao nascer; mostra o número médio de anos que uma pessoa nascida em determinado município viveria a partir do nascimento, mantidos os mesmos padrões de mortalidade.	ATLAS BRASIL
<i>IDHM-Renda</i>	Corresponde a componente "renda" do IDHM. É medido pela renda municipal <i>per capita</i> , ou seja, a renda média dos residentes de um determinado município.	ATLAS BRASIL

<i>IDHM-Educação</i>	Corresponde a componente "educação" do IDHM. É calculado a partir da média geométrica de dois sub-índices: a frequência de crianças e jovens na escola (com peso de 2/3) e a educação da população adulta (com peso de 1/3).	ATLAS BRASIL
<i>EspaçoReferência</i>	Indica se o município está localizado na região Amazônica ou em outra região do país, assumindo os valores respectivamente como "Amazônia" ou "Outra".	IBGE

### 3.3 – METODOLOGIA UTILIZADA PARA ANÁLISE DOS DADOS

A metodologia aplicada neste estudo ecológico, combina as etapas da mineração de dados com a análise da distribuição espacial, em quatro fases, conforme demonstrado na Figura 4, e descrita a seguir.



**Figura 4.** Metodologia em fases para análise dos dados (Adaptado de FAYYAD *et al.* 1996).

#### 3.3.1 – Fase I – Preparação dos dados

Na fase de preparação de dados, as variáveis foram selecionadas e integradas a uma única base de dados, resultando em uma base de dados contendo 5565 registros identificados de municípios (código e nome do município) com 10 atributos associados (variáveis). Para a integração, foi necessário um pré-processamento, ou seja, uma preparação dos dados, onde foram realizadas todas as tarefas para eliminar redundância e inconsistência nos dados.

### 3.3.2 – Fase II – Mineração dos dados

Nesta fase foram realizadas duas etapas: de transformação e de mineração de dados. Para esta fase, foram selecionadas apenas as variáveis do estudo, descartando-se a identificação do município (código e nome do município) o que resultou em uma matriz de 5565 linhas (municípios) por 10 colunas (variáveis).

A transformação envolveu a aplicação de algoritmos de discretização com o objetivo de colocar valores em intervalos para que haja um número limitado de estados possíveis. Essa etapa é importante porque na busca por associação não é possível utilizar valores contínuos como entrada, pois uma única variável pode conter grande quantidade de valores distintos, de tal forma que seja impossível identificar as associações relevantes; assim, as variáveis contínuas foram transformadas em variáveis categóricas. O método de discretização utilizado, foi baseado na frequência, ou seja, buscou-se um equilíbrio quanto ao número de registros em cada faixa de valores, sendo que idealmente, cada faixa discretizada deveria conter a mesma quantidade de registros. Assim, todas as variáveis contínuas (*AguaEncanada*, *BanheiroAguaEncanada*, *ColetaLixo*, *EnergiaEletrica*, *DensidadePopMaior2*, *Hanseníase*, *IDHM-Longevidade*, *IDHM-Renda*, *IDHM-Educação*) passaram pelo processo de discretização por frequência.

Na sequência, ainda na fase de mineração, foi utilizada a técnica de redes bayesianas com o objetivo de medir a associação entre as variáveis do domínio do problema. As redes bayesianas permitem construir modelos formados por uma estrutura qualitativa, definida por grafos acíclicos dirigidos, onde são representadas as dependências entre os nós (variáveis) e uma estrutura quantitativa (tabelas de probabilidades condicionais dos nós), possibilitando medir essas dependências em termos probabilísticos (CHEN, 2001). Essas estruturas oferecem uma representação eficiente da distribuição de probabilidades do conjunto de variáveis em estudo. Assim, para cada nó, existe uma distribuição de probabilidade local, que depende do estado de seus pais.

Para o aprendizado da estrutura da rede bayesiana foi aplicado o algoritmo de busca heurística K2 (COOPER e HERSKOVITS, 1992), tendo como entrada a matriz 5565 x 10 colunas. O K2 aprende a estrutura da rede bayesiana a partir da matriz de entrada adotando um método que seleciona a estrutura com máxima verossimilhança, através de uma busca iterativa que inicia com uma estrutura de rede bem simples, onde todas as variáveis são independentes uma das outras. Em seguida são avaliadas as verossimilhanças marginais de cada rede resultante de possíveis mudanças (e.g. adicionar uma nova ligação entre dois nós dessa rede), e então é aplicada a melhor das alterações antes de uma nova iteração. Esse processo continua até que o algoritmo não consiga encontrar nenhuma alteração simples que melhore (aumente) o valor de

máxima verossimilhança. A rede resultante pode conter um número de variáveis dependentes inferior ao número de variáveis da matriz de entrada. Esse algoritmo, classificado como de busca e pontuação, foi selecionado devido ao seu desempenho em termos de complexidade computacional e resultados precisos, obtidos a partir de uma ordenação de variáveis adequada. Nesse estudo, a variável alvo (*Hanseníase*) foi posicionada na última coluna da matriz de entrada.

Uma vez encontrada a rede bayesiana, foi utilizada a inferência estatística a fim de estimar a posterior distribuição dos parâmetros. A tarefa básica da inferência consistiu em computar a distribuição da probabilidade condicional utilizando as informações quantitativas da rede bayesiana, ou seja, para um conjunto de variáveis de consulta (*Consulta*), dado os valores de um evento observado (*Evidências*), o sistema computa  $P(\text{Consulta}|\text{Evidências})$ . Para avaliar a relevância das regras de associação produzidas, foi utilizado o modelo suporte-confiança de (AGRAWAL *et al.*, 1993), que permite avaliar a qualidade da regra de associação encontrada.

### 3.3.3 – Fase III - Visualização da distribuição espacial

Na terceira fase, utilizando o software Qgis<sup>5</sup>, a base dos 5565 municípios foi integrada ao arquivo *shapefile*, que contém as informações georeferenciadas dos municípios, para apresentar a distribuição espacial da incidência da Hanseníase no Brasil.

### 3.3.4 – Fase IV - Análise dos dados

Nesta fase, foram produzidos mapas em conjunto com os resultados da mineração de dados, que serão analisados no capítulo referente aos resultados encontrados.

## 3.4 – CONSIDERAÇÕES FINAIS

Este capítulo mostrou a metodologia utilizada, desde a coleta dos dados, a seleção das variáveis e as fases que foram utilizadas para a análise dos dados, a ser descrita no próximo capítulo.

---

<sup>5</sup> <http://qgis.org>

## CAPÍTULO 4 – ANÁLISE SÓCIO-DEMOGRÁFICA DA INCIDÊNCIA DA HANSENÍASE UTILIZANDO AS REDES BAYESIANAS

### 4.1 – CONSIDERAÇÕES INICIAIS

Neste capítulo será mostrado os resultados obtidos a partir dos dados sócio demográficos coletados e da incidência da Hanseníase nos municípios estudados.

Vários softwares para aprender redes bayesianas estão disponíveis no mercado para a serem utilizados em trabalhos a partir de dados e raciocínio usando inferência probabilística, como exemplo: Weka, Arena, NS3, Bayesware e outros. Para construção da rede bayesiana deste trabalho, foi utilizado o software *Bayesware Discoverer* versão 1.0 *Professional Edition for Microsoft Windows*, por ter se mostrado um software mais amigável.

### 4.2 – AJUSTES NECESSÁRIOS PARA UNIFORMIZAR OS DADOS PARA A ANÁLISE

Conforme descrito no capítulo anterior, os dados foram selecionados e integrados em uma única base de dados. Todas as variáveis contínuas passaram pelo processo de discretização, que considerou intervalos de mesma frequência para a distribuição de cada valor nas categorias criadas.

No caso da variável *Hanseníase*, em função do elevado número de municípios (2.193) sem registro da doença, foi necessário preliminarmente categorizar esses municípios como “SemRegistro” e na sequência aplicar o método de discretização por frequência para os demais municípios (3.372).

Assim, os municípios com incidência da doença foram categorizados em três faixas de valores, de frequência igual (Tabela 3), ou seja, 1.124 municípios em cada categoria de incidência (“baixa”, “média” e “alta”).

**Tabela 3.** Categorização da taxa de incidência.

<b>Categoria da Taxa de Incidência</b>	<b>Número de Municípios</b>	<b>(%) de municípios na categoria</b>	<b>Intervalo da categoria</b>
SemRegistro	2.193	39,40%	[0]
Baixa	1.124	20,20%	(0 – 14,09]
Média	1.124	20,20%	(14,09 – 37,46]
Alta	1.124	20,20%	(37,46– 840,97]

Para as demais variáveis, o processo de discretização resultou também em quatro faixas de valores (categorização), conforme apresentado na Tabela 4.

**Tabela 4.** Categorização de *IDHM-Renda*, *IDHM-Educação*, *BanheiroAguaEncanada* e *DensidadePopMaior2*

Variável	Categoria	(%) de municípios na categoria	Intervalo da categoria
<i>IDHM-Renda</i>	Muito Baixo	24,9%	[0,400 – 0,572]
	Baixo	25,0%	(0,572 – 0,654]
	Médio	24,9%	(0,654 – 0,707]
	Alto	25,2%	(0,707 – 0,891]
<i>IDHM-Educação</i>	Muito Baixo	24,9%	[0,207 – 0,490]
	Baixo	24,9%	(0,490 – 0,560]
	Médio	25,2%	(0,560 – 0,631]
	Alto	25,0%	(0,631 – 0,825]
<i>BanheiroAguaEncanada</i>	Muito Baixo	24,6%	[3,26 – 67,78]
	Baixo	25,6%	(67,78 – 91,25]
	Médio	25,2%	(91,25 – 98,00]
	Alto	24,6%	(98,00 – 100,0]
<i>DensidadePopMaior2</i>	Muito Baixo	24,7%	[0,65 – 15,41]
	Baixo	25,2%	(15,41 – 23,07]
	Médio	25,5%	(23,07 – 32,58]
	Alto	24,7%	(32,58 – 88,64]

#### 4.3 – ANÁLISE DOS DADOS

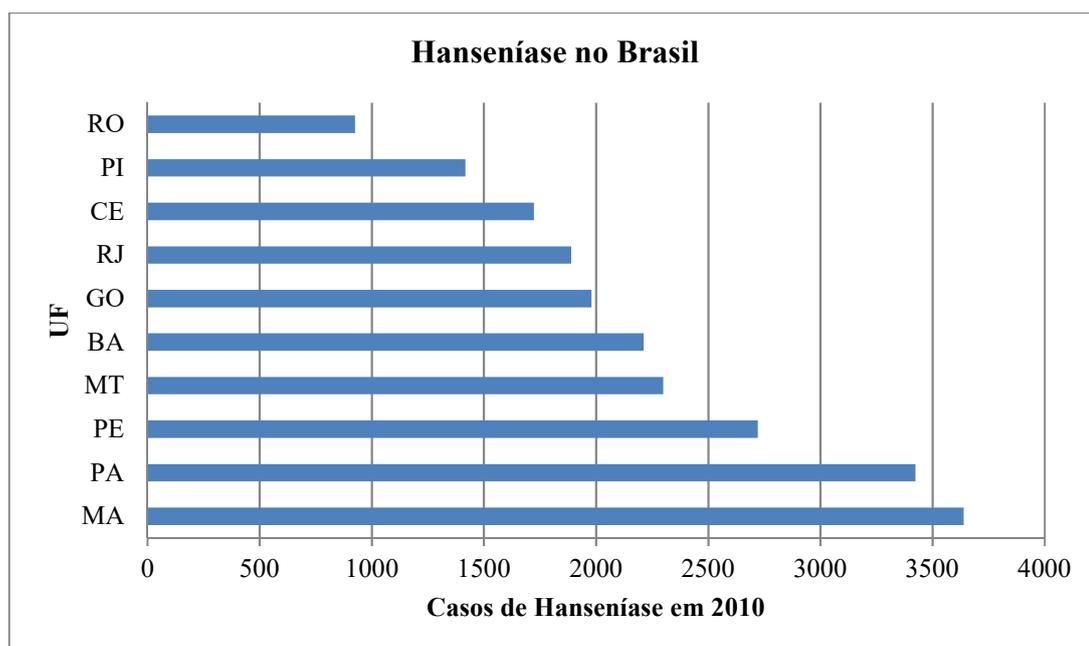
Na (Tabela 5), são apresentadas as médias da taxa de incidência de Hanseníase para municípios da Amazônia e outras regiões, incluindo a relação dos dez primeiros municípios do Brasil com maior taxa de incidência da doença. Embora a média da Amazônia seja maior em relação às outras regiões do país, nos dez primeiros municípios há, apenas, cinco municípios da Amazônia, sendo três do Mato Grosso (MT), um de Tocantins (TO) e um do Pará (PA). Na Amazônia, o município com maior taxa de incidência é Araguaiana (MT) com 563,03 e o segundo é Jacareacanga (PA) com 354,53. O valor máximo desse indicador (840,97) é encontrado no município brasileiro de Pedro Gomes (MS).

**Tabela 5.** Hanseníase, conforme a Taxa de Incidência - média (Amazônia, Outra) e municípios selecionados.

<b>Território</b>	<b>Município</b>	<b>Taxa de Incidência</b>
<b>Amazônia</b>	(média)	73,23
<b>Outra região</b>	(média)	16,98
MS	Pedro Gomes	840,97
GO	Campos Verdes	697,21
MT	Araguaiana	563,03
SE	Amparo de São Francisco	483,52
GO	Padre Bernardo	458,96
TO	Novo Jardim	366,30
PA	Jacareacanga	354,53
MT	Itaúba	349,73
MT	Nova Nazaré	330,14
GO	Jussara	313,27

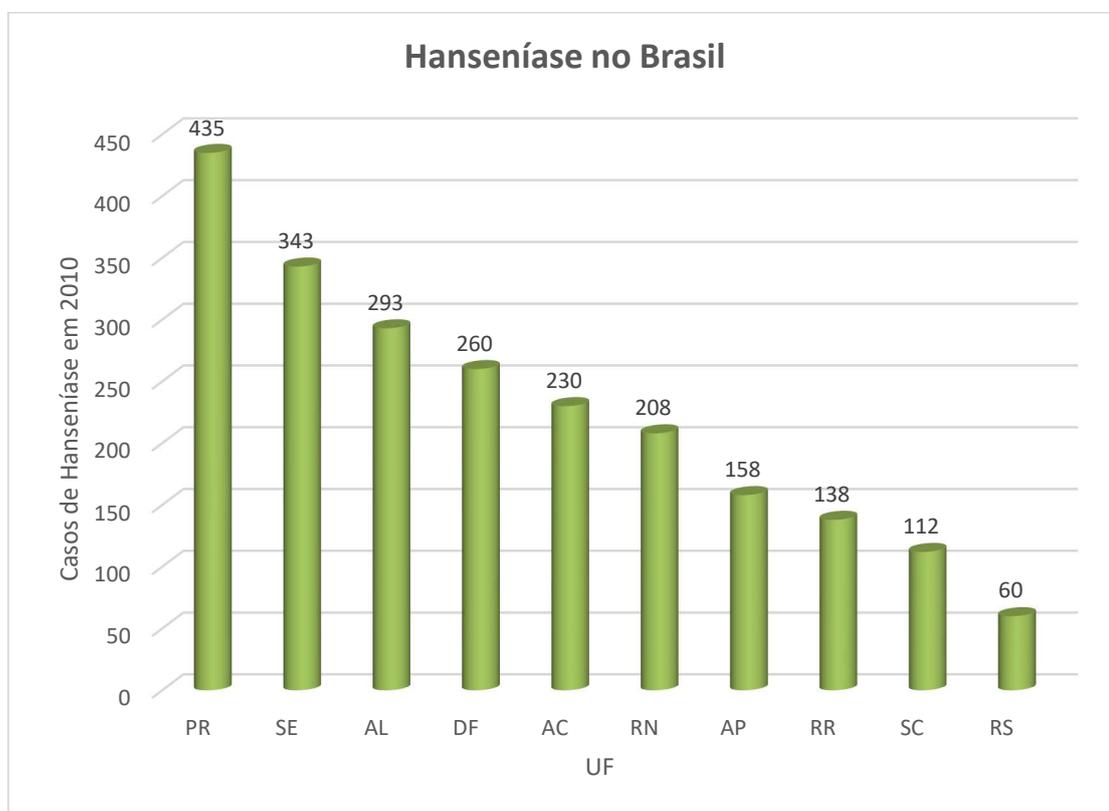
Expandindo a (Tabela 5) para os cinquenta primeiros municípios com maior taxa de incidência, todos são municípios de pequeno porte (até 50 mil habitantes) e 31 desses municípios estão localizados na região amazônica. Do total de municípios da Amazônia (771), 498 municípios apresentam alta taxa de incidência de Hanseníase (acima de 37,46), o que representa 64,59% dos municípios da Amazônia; quando o espaço de referência são outras regiões do Brasil, apenas 13,06% dos municípios estão nessa categoria.

Quando o espaço de referência são os estados brasileiros, os dez estados com mais casos de Hanseníase são o Maranhão e Pará, que, juntos, apresentam mais de 7000 casos (Figura 5). A alta e a média taxa de incidência de hanseníase no estado do Maranhão são encontradas, respectivamente, em 60,84% e 29,37% dos municípios. No estado do Pará a alta taxa de incidência da doença é encontrada em 58,06% dos municípios e a média em 22,58%.



**Figura 5.** Casos de Hanseníase nos estados brasileiros.

Ainda levando em conta o espaço referência serem os estados brasileiros, os 10 estados com menor número de incidência registrado no ano de 2010 são (**Figura 6**), onde o estado do Rio Grande do Sul, com 60 casos, é o que apresenta menor número de casos de Hanseníase.



**Figura 6.** Estados com menor número de casos de Hanseníase detectados.

Na sequência, a técnica de redes bayesianas foi utilizada com o objetivo de medir a associação entre as variáveis estudadas.

Inicialmente foi utilizado para este estudo, 10 variáveis das dimensões de Hanseníase sendo: *Aguaencanada*, *Banheiroaguaencanada*, *Coletadelixo*, *Energiaeletrica*, *Idhm-Longevidade*, *Idhm-Renda*, *Idhm-Educacao*, *Densidadepopmaior2*, *Espacoreferencia* e *Hanseníase*.

Nesta primeira amostragem, percebe-se as inferências mostrada na (Figura 7), pela Rede Bayesiana, que a Hanseníase está diretamente relacionada ao Espaço Referência e a Densidade populacional maior que 2.

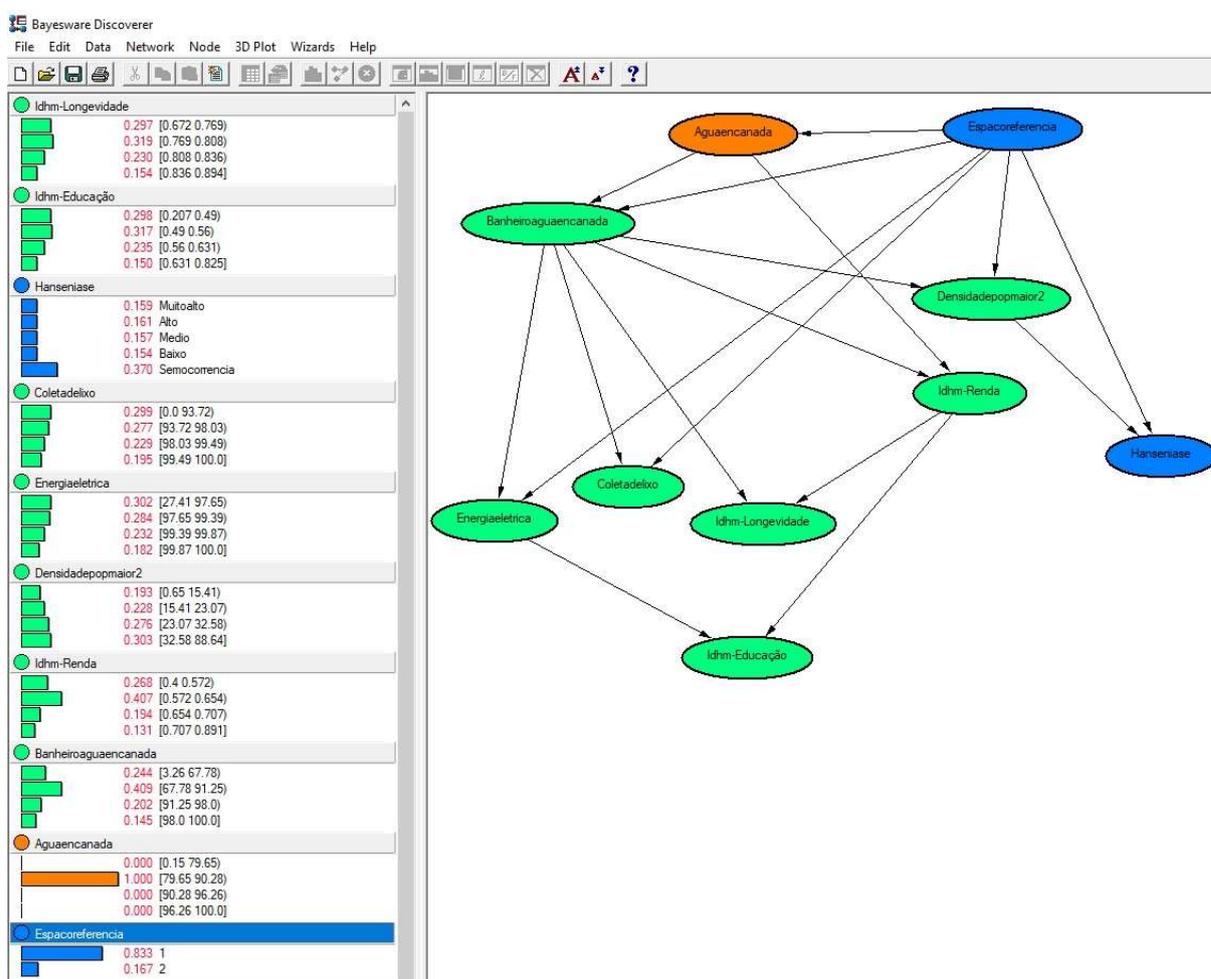
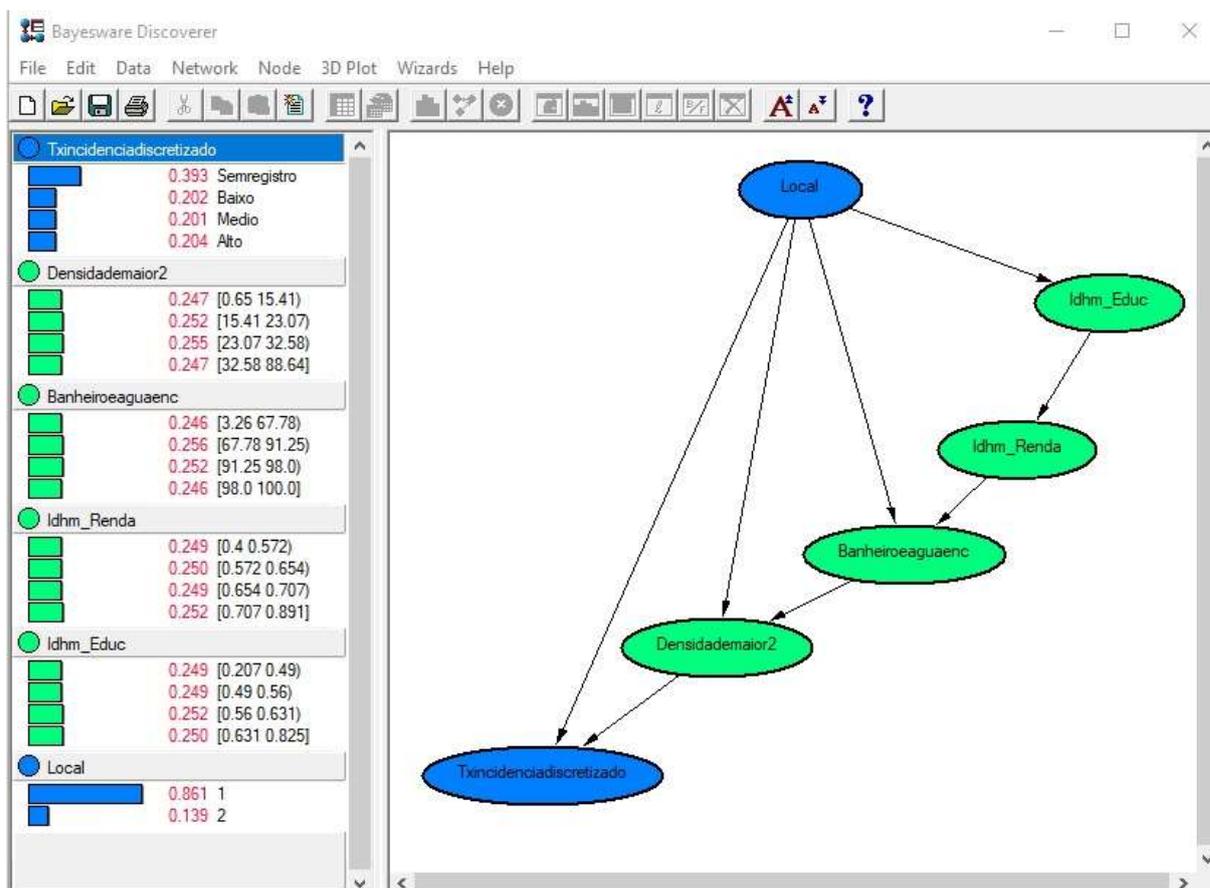


Figura 7. Rede bayesiana com as 10 variáveis iniciais de estudo

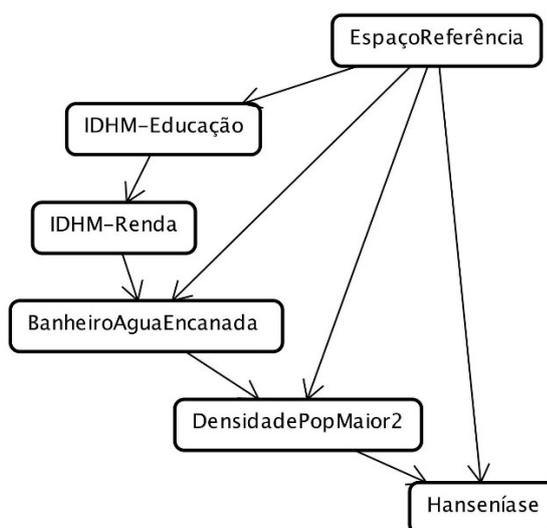
Numa nova etapa com o uso da Rede Bayesiana, com a aplicação do algoritmo K2, foram encontradas associações relevantes com as seguintes variáveis: *EspaçoReferência*,

*IDHM-Renda, IDHM-Educação, BanheiroAguaEncanada e DensidadePopMaior2*. Nesta etapa, as demais variáveis foram descartadas do estudo da associação (**Figura 8**).



**Figura 8.** Rede bayesiana com 6 variáveis

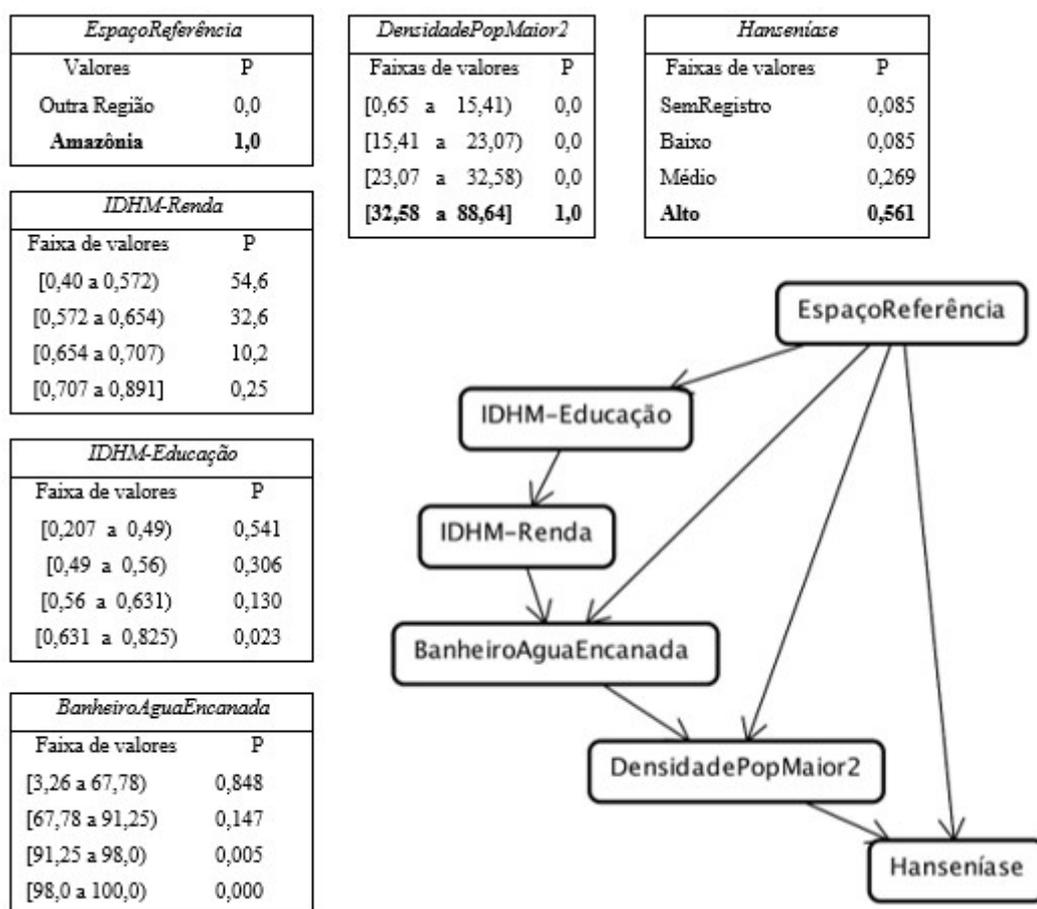
Na rede bayesiana resultante (**Figura 9**), a variável *Hanseníase* é diretamente dependente das variáveis *DensidadePopMaior2* e *EspaçoReferência*.



**Figura 9.** Estrutura da rede bayesiana selecionada.

As inferências apresentadas a seguir foram obtidas a partir dos resultados dos algoritmos de propagação.

Quando a evidência é que o município pertence à Amazônia, a probabilidade de apresentar alta taxa de incidência de Hanseníase é de 65,7%. Esse valor cai para 13,1% quando o espaço de referência são outras regiões do país. Quando a evidência é que um município apresenta a faixa mais baixa de *DensidadePopMaior2* (<15,41%), a probabilidade desse município não apresentar registro da doença é de 60%; por outro lado, quando apresenta a faixa mais alta (acima de 32,58%), essa probabilidade cai para 22,7%. Quando a consulta considera as duas variáveis (*EspaçoReferência* = *Amazônia* e *DensidadePopMaior2* < 15,41), esse percentual, em termos probabilísticos, passa a ser de 56,1% (**Figura 10**). Em se tratando de outras regiões do país, esse percentual cai para 17,9%.



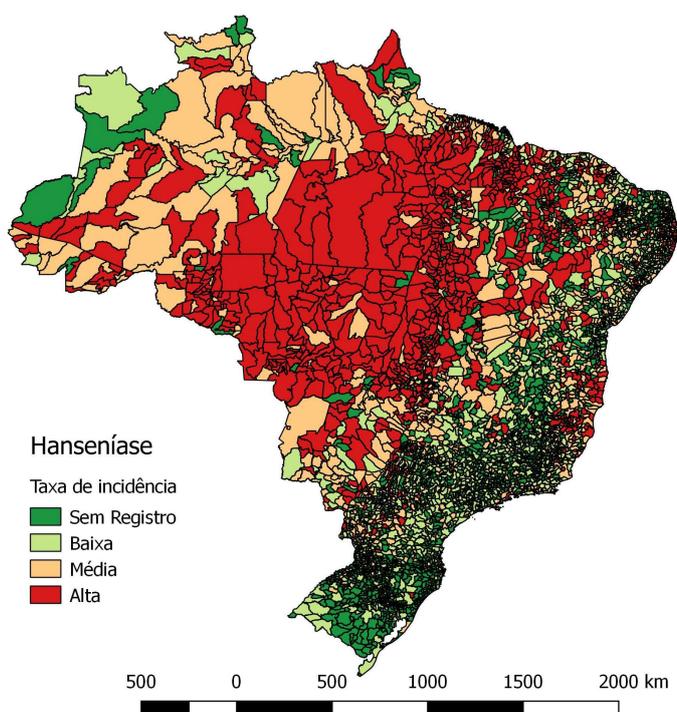
**Figura 10.** Rede bayesiana (evidências: *EspaçoReferência* = "Amazônia" e *DensidadePopMaior2* > 32,58).

A existência de banheiro com água encanada, segundo os dados do Censo, é um problema característico da Amazônia: para os municípios da Amazônia, a probabilidade de apresentar os menores percentuais de existência de banheiro com água encanada é de 56,1%;

esse percentual cai para 19,6% quando se trata de outras regiões. De fato, o cenário onde a evidência é de *Hanseníase* alta, maior faixa de *DensidadePopMaior2* e menor faixa de *BanheiroAguaEncanada* é um cenário típico da Amazônia (probabilidade de 73% contra 27% de outras regiões).

As variáveis *IDHM-Renda* e *IDHM-Educação* não apresentaram associação direta com *Hanseníase*. Entretanto, *IDHM-Renda* apresenta associação direta com *BanheiroAguaEncanada*: na faixa mais baixa do *IDHM-Renda*, a probabilidade do município apresentar os menores percentuais de existência de *BanheiroAguaEncanada* é de 74,2%. Por outro lado, quando se trata da faixa mais elevada do *IDHM-Renda*, essa probabilidade cai para zero. As variáveis *IDHM-Renda* e *IDHM-Educação* influenciam na densidade populacional do domicílio: quando a evidência é que um município apresenta as faixas mais baixas dos indicadores de renda e educação, a probabilidade desse município apresentar a faixa mais elevada de *DensidadePopMaior2* é de 54,2%; essa probabilidade cai para 4,7% quando a evidência é da faixa mais elevada para os indicadores de renda e educação.

Para visualizar, em termos percentuais, a incidência de Hanseníase nos municípios brasileiros, o mapa da (Figura 11) apresenta a distribuição espacial desse fenômeno, conforme as faixas de valores definidas no processo de discretização. Por exemplo, para as faixas de valores de *Taxa de Incidência*, 60,6% de municípios brasileiros apresentaram ocorrência da doença que foram classificados em baixo, médio e alto.



**Figura 11.** Mapa de casos de Hanseníase nos municípios brasileiros.

A distribuição espacial evidenciou que o padrão de alta incidência predomina nos Estados do Pará, Maranhão, Mato Grosso, Goiás, Pernambuco e Bahia e que há, principalmente, alguma relação de contiguidade entre os municípios com alta incidência da doença.

Além disso, observou-se que esse fenômeno é característico de municípios de menor tamanho populacional, porém o número máximo de casos da doença são registrados em municípios de grande porte e em metrópoles (Tabela 6).

**Tabela 6.** Casos e taxa de incidência de Hanseníase de acordo com a classificação do município.

Tamanho do Município	Nº Casos		Taxa de incidência	
	Mínimo	Máximo	Mínimo	Máximo
<b>Metrópolis</b>	<b>737</b>	<b>988</b>	<b>72,62</b>	64,25
<b>Grande Porte</b>	<b>45</b>	<b>703</b>	<b>43,73</b>	86,34
Médio Porte	21	140	40,66	189,08
<b>Pequeno Porte 1</b>	1	67	37,55	<b>840,97</b>
<b>Pequeno Porte 2</b>	8	127	38,75	<b>458,96</b>

Os cinco estados com maiores *taxas de incidência de Hanseníase* são na respectiva ordem: Goiás, Mato Grosso, Tocantins, Maranhão e Pará, destacando-se que 4 estados pertencem a Amazônia Legal, esses quatro estados formam área endêmica no Brasil, acrescido do estado de Rondônia (FREITAS *et al.* 2017).

No modelo de rede bayesiana, existe uma associação significativa entre os percentuais de domicílios com densidade maior que 2 e taxa de incidência de Hanseníase. Embora a relação entre a taxa de incidência, fatores socioeconômicos (indisponibilidade de abastecimento de água, existência de banheiro, pobreza e aglomeração de pessoas nos domicílios), baixos índices educacionais e de renda já tenha sido evidenciada em vários estudos, Ponnighaus *et al.* 1994; Kerr-Pontes *et al.* 2006; Sales *et al.* 2011; Suzuki *et al.* 2012; Moura *et al.* 2013, a inserção da variável que considera Domicílios com Densidade maior que dois acrescenta elementos na discussão do fenômeno, não contrapondo conclusões desses estudos. De fato, os municípios cujos domicílios apresentam os mais elevados percentuais de densidade tendem a ser aqueles que também concentram domicílios com pior padrão de rendimento domiciliar (e vice-versa), e, igualmente, pior padrão de educação (e vice-versa).

Assim como no estudo de Freitas *et al.* 2017, este estudo ecológico investiga a associação de taxa de incidência de Hanseníase em sua ocorrência coletiva (municípios), e não individual, ou seja, uma associação observada para o município não obrigatoriamente ocorre em nível de indivíduo, o que pode levar a erros de falácia ecológica. A falácia ecológica consiste em pensar que as relações observadas para grupos necessariamente se mantêm para os indivíduos (FREEDMAN, 1999).

Neste estudo, as inferências realizadas são apoiadas pelos dados agregados e portanto, não podem ser tomadas em nível do indivíduo. Apesar dessa limitação, as regras de associação encontradas aqui corroboram estudos realizados, cujos argumentos relacionam esse fenômeno com o ambiente em que os indivíduos portadores de Hanseníase, estão inseridos.

O cenário encontrado na análise dos dados revela a periferização no Brasil, o que representa uma barreira para garantir a equidade no acesso aos serviços e informações de saúde. Três Arroios (RS), por exemplo, é um município que possuía 0,65% de densidade maior que dois; no outro extremo, Uiramutã (RR) 88,64%, Santa Isabel do Rio Negro (AM) 87,59%, Melgaço (PA) 86,43% e Maraã (AM) 82,75%, todos da Amazônia.

A tabela abaixo demonstra que a Hanseníase é uma doença que tem maior incidência em áreas com menor acesso a serviços de saúde e piores indicadores sócio econômicos, onde coincidentemente, tais fatores encontram-se localizados em áreas rurais. Se porventura, as áreas rurais forem municípios bem estruturados nas redes de saúde e sócio econômica, a incidência é pequena, como é mostrado na tabela. Ainda vale a pena observar que as menores incidências não estão localizadas nas capitais, à exceção de Porto Alegre. Isso confirma os achados deste estudo, que relacionam a Hanseníase como uma doença da pobreza.

10 Municípios com Menor Taxa			10 Municípios com Maior Taxa		
Município	UF	Incidência	Município	UF	Incidência
Pelotas	RS	0,61	Pedro Gomes	MS	840,97
Canoas	RS	0,62	Campos Verdes	GO	697,21
Porto Alegre	RS	0,64	Araguaiana	MT	563,03
Itapetininga	SP	0,69	Amparo de São Francisco	SE	483,52
Jaú	SP	0,76	Padre Bernardo	GO	458,96
Atibaia	SP	0,79	Novo Jardim	TO	366,30
Viamão	RS	0,84	Jacareacanga	PA	354,53
Santa Cruz do Sul	RS	0,84	Itaúba	MT	349,73
Conselheiro Lafaiete	MG	0,86	Nova Nazaré	MT	330,14
Ribeirão Pires	SP	0,88	Jussara	GO	313,27

**Tabela 7.** - Municípios com menor e maior taxa de incidência de Hanseníase em 2010

#### 4.4 – CONSIDERAÇÕES FINAIS

Os resultados obtidos e a análise das redes bayesianas mostraram a associação entre a taxa de incidência de Hanseníase e as variáveis representativas das dimensões de renda, educação, banheiro e água encanada e domicílios com densidade maior que 2, nos municípios brasileiros no ano de 2010. O estudo ainda demonstrou que os percentuais de incidência de Hanseníase apresentaram-se maiores nos municípios da Amazônia do que no Brasil.

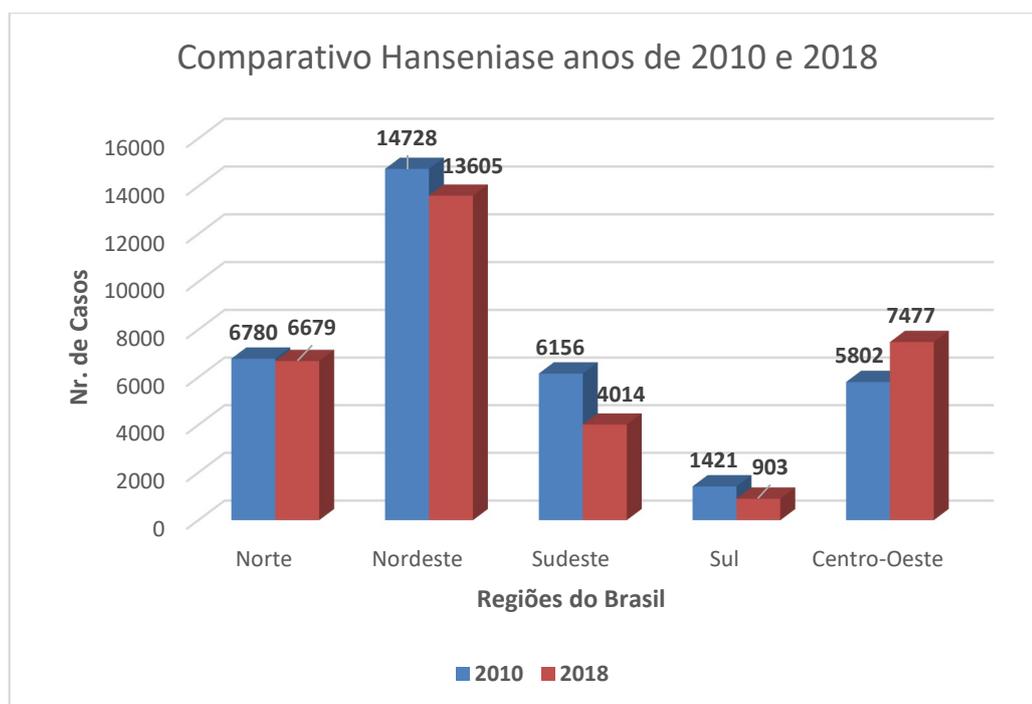
## 5. CONCLUSÃO

O conhecimento da vulnerabilidade das pessoas às doenças transmissíveis auxilia nas tomadas de decisões considerando aspectos das necessidades de saúde, devido a exclusão social e o medo. Identificar as variáveis e os fatores implicam no desenvolvimento de estratégias que possam diminuir o avanço de doenças como a Hanseníase.

O estudo mostrou isso quando utilizou a técnica de redes bayesianas com o objetivo de medir a associação entre as variáveis estudadas. Foram encontradas associações relevantes com as seguintes indicadores: Amazônia Legal Brasileira, Índice de Desenvolvimento Humano Municipal de Renda e Educação e Condição de Habitação dos Domicílios. .

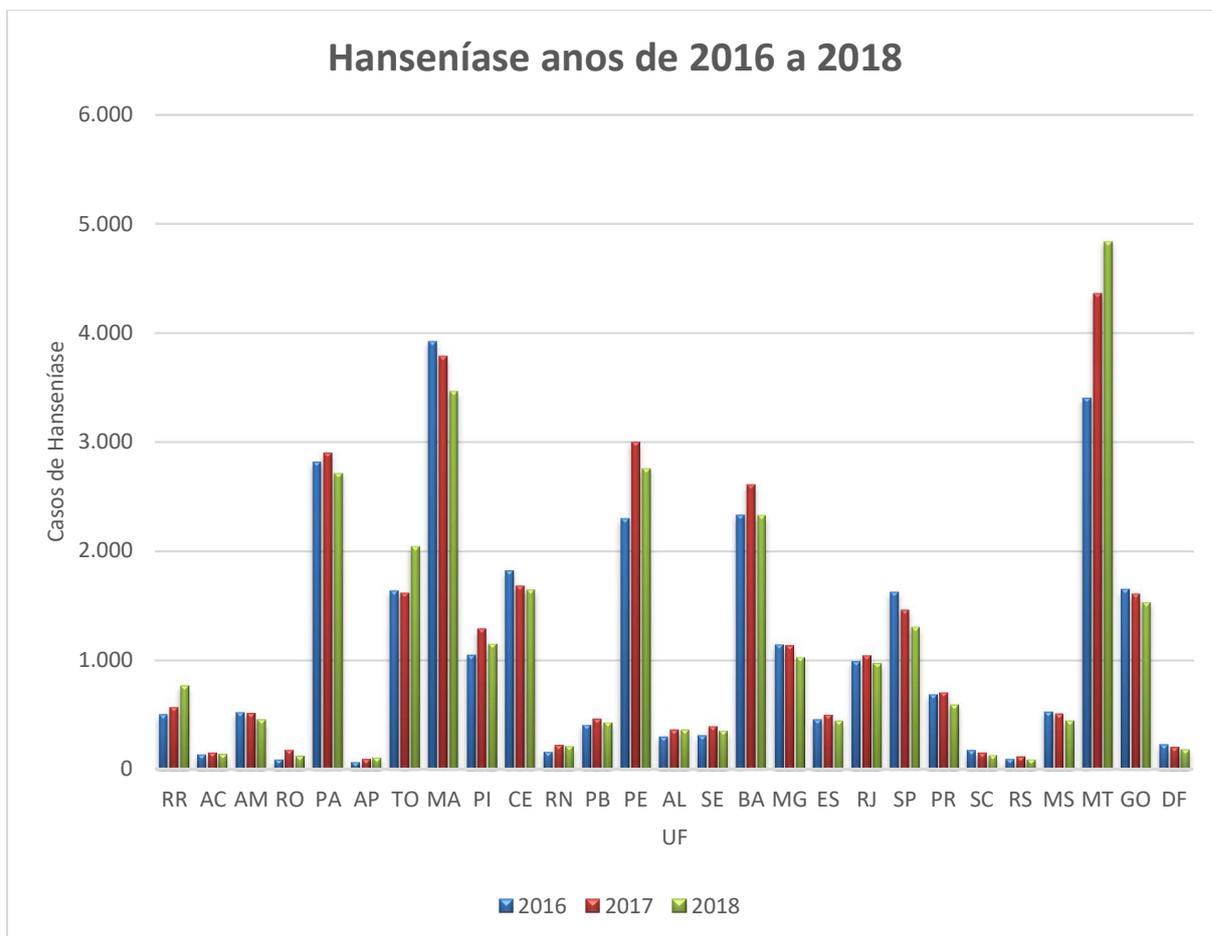
Neste estudo, a rede bayesiana resultante, mostrou que o indicador Hanseníase é diretamente dependente dos indicadores, Condição de Habitação dos Domicílios e Amazônia Legal Brasileira O estudo ainda demonstrou que os percentuais de incidência de Hanseníase apresentaram-se maiores nos municípios da Amazônia do que no Brasil.

De acordo com indicadores do SINAN obtidos em fevereiro de 2019, comparando os anos de 2010 e 2018 por regiões do Brasil (**Figura 12**), a única região em que houve crescimento da Hanseníase foi a região Centro-Oeste, que subiu de 5802 em 2010 para 7477 em 2018.



**Figura 12** – Comparativo da Hanseníase por Região nos anos de 2010 e 2018.

Ainda de acordo com os dados do SINAN obtidos também em fevereiro de 2019, analisando os anos de 2016 a 2018, é possível perceber aumento dos casos de hanseníase no estado de Mato Grosso(MT), que pertence à Amazônia Legal, espaço referência deste estudo, onde respectivamente foram registrados 3403 casos em 2016, 4362 casos em 2017 e 4835 casos em 2018. (**Figura 13**).



**Figura 13** – Comparativo da Hanseníase por Estado nos anos de 2016 a 2018.

Embora a taxa de incidência da Hanseníase envolva elementos sociais relacionados à vulnerabilidade, na prevenção precoce é fundamental prover informações e orientações adequadas. Intensificando as ações de combate a Hanseníase, diagnóstico precoce, prevenção, tratamento, reabilitação física e social, buscando ampliar o acesso ao diagnóstico e ao tratamento nos municípios com maior endemia.

## 5.1 CONTRIBUIÇÕES DESTE TRABALHO

A utilização das redes bayesianas como ferramenta de controle no monitoramento de doenças, quer endêmicas ou epidêmicas, já é uma estratégia exitosa, principalmente na previsão de controle das epidemias. Neste trabalho, pode-se citar as seguintes contribuições:

- a) A analogia da incidência da doença com relação à distribuição espacial Amazônia Legal e Brasil. Os estudos encontrados relacionam Estados e alguns em Regiões fazendo analogia com o Brasil.
- b) Desenvolvimento de um modelo de rede bayesiana com variáveis socioeconômicas da Amazônia Legal relevantes para a incidência da Hanseníase.
- c) Introdução da variável densidade populacional como a de maior relevância na incidência da doença estudada no espaço referência.
- d) O trabalho permitiu validar, através da significância da densidade populacional, a necessidade de uma política pública de habitação com qualidade, para a população amazônica.
- e) A divulgação dos resultados desta pesquisa ocorreu a partir da publicação do artigo intitulado “*INCIDENCE OF LEPROSY IN THE BRAZILIAN AMAZON REGION AND ITS RELATION TO BRAZILIAN MUNICIPALITIES: A SOCIODEMOGRAPHIC ANALYSIS USING BAYESIAN NETWORKS*”, publicado no *International Journal of Development Research*, Vol. 08, Issue, 11, pp.24189-24194, November, 2018. e consolidar-se nesta dissertação.

## 5.2 TRABALHOS FUTUROS E DESDOBRAMENTOS DA PESQUISA

Futuras pesquisas poderão ter continuidade a partir deste trabalho utilizando o modelo de rede bayesiana selecionada, através de dados a serem obtidos de séries temporais de incidência e prevalência da Hanseníase, no sentido de ampliar o trabalho e oferecer novas perspectivas de monitoramento para a problemática do controle da doença.

Também a realização de um estudo georreferenciado para diagnóstico mais preciso das áreas mais atingidas pela doença no espaço referência, objetivando a elaboração de políticas públicas, principalmente habitacionais, nos municípios mais atingidos pelo acúmulo populacional domiciliar.

Além disso, a metodologia utilizada no trabalho contribui para monitoramento de outras doenças infecciosas negligenciadas, relacionadas com as variáveis estudadas e de caráter endêmico na Amazônia Legal, como a tuberculose, a leishmaniose e as hepatites virais.

### **5.3 DIFICULDADES ENCONTRADAS**

Uma das dificuldades encontradas neste, que pode ser classificada como limitação do estudo, foi a obtenção dos dados tendo como referência o ano de 2010, por ser o ano do último censo, e para manter a consistência metodológica, os dados da incidência da doença também foram coletados para o ano de 2010, o que não implica na fidedignidade dos resultados encontrados.

Apesar da limitação do estudo descrita acima, o trabalho demonstrou a importância da utilização das redes bayesianas na elaboração das políticas públicas em saúde e a importância da aglomeração domiciliar na incidência da Hanseníase. O cenário encontrado na análise dos dados revela a periferização no Brasil, o que representa uma barreira para garantir a equidade no acesso aos serviços e às informações de saúde.

## REFERÊNCIAS BIBLIOGRÁFICAS

- AGRAWAL, R.; IMIELINSKI, T.; SWAMI, A. Mining association rules between sets of items in large databases. **IBM Almaden Research Center 650 Harry Road**. San Jose, CA 95120. 1993.
- ANDRADE, P.J.N. Sistemas Especialistas de Apoio ao Diagnóstico em Medicina. Relações com o Teorema de Bayes e com a Lógica do Raciocínio Diagnóstico. **Hospital Universitário Walter Cantídio da UFC**. Fortaleza-CE, 73(6);537-544. Jun 1999.
- ANDRADE, V.L.G.; SABROZA, P.C.; ARAUJO, A.J.G. Fatores Associados ao Domicílio e à Família na Determinação da Hanseníase. **Cad. Saúde Pública**. Rio de Janeiro-RJ. 10(supl. 2): p. 281-292. 1994.
- ARA-SOUZA, A.L. Redes bayesianas: Uma introdução aplicada a credit scoring. **Relatório apresentado à comissão julgadora do concurso de Iniciação Científica do 19º Simpósio Nacional de Probabilidade e Estatística (SINAPE)** São Carlos-SP. 2010.
- ASSUNÇÃO, R.M.; BARRETO, S.M.; GUERRA, H.L.; SAKURAI, E. Mapas de taxas epidemiológicas: uma abordagem Bayesiana. **Cad. Saúde Pública**, Rio de Janeiro-RJ. 14(4):713-723. out-dez, 1998.
- AUGUSTO, D.S. Doenças Negligenciadas: Estudo sobre Conhecimentos e Práticas dos Professores do Ensino Fundamental no Município do Jaboatão dos Guararapes/PE. **Dissertação: (Centro de Pesquisas Aggeu Magalhães) | Fundação Oswaldo Cruz**. 105p. 2016.
- BARRETO, J.A. O Brasil ocupa o segundo lugar no mundo em hanseníase: Depoimento. [18 de fevereiro, 2018]. São Paulo-SP. **Jornal O São Paulo**. Entrevista concedida à Cleide Barbosa/radio 9 de julho. 2018.
- BARRETO, M.L. Esboços para um cenário das condições de saúde da população brasileira 2022/2030. In: Cruz, FO. (Ed.). **A saúde no Brasil em 2030: prospecção estratégica do sistema de saúde brasileiro - população e perfil sanitário**. Rio de Janeiro: Fiocruz / Ipea / Ministério da Saúde. p.97-120. 2013.

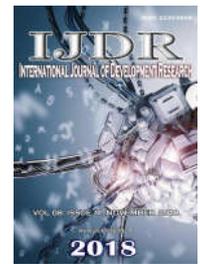
- BARROS, M.B.A.; CÉSAR, C.L.G.; CARANDINA, L.; TORRE, G.D. Desigualdades sociais na prevalência de doenças crônicas no Brasil, PNAD-2003. *Ciência & Saúde Coletiva*. São Paulo-SP. p. 911-926. 2006.
- BASTOS, R.F.M.; OLIVEIRA, D.F.S. Redes Bayesianas – aplicações em confiabilidade e no diagnóstico de perdas não técnicas. **Editora Appris**. Curitiba-PR. 2017.
- BRITO, K.K.G.; ARAUJO, D.A.L.; UCHÔA, R.E.M.N.; FERREIRA, J.D.L.; SOARES, M.J.G.O.; LIMA, J.O. Epidemiologia da hanseníase em um estado do nordeste brasileiro. **Rev.Enferm UFPE**. 8(8): 2686-93. Ago 2014.
- BRITO, S.R.; SILVA, A.S.S.; CRUZ, A.G.; BARROSO, R.F.F.; MONTEIRO, M.A.; COSTA, J.C.W.A.; FRANCES, C.R.L. Gravidez na adolescência e o acesso às Tecnologias de Informação e Comunicação na Amazônia. **Mundo Amazônico**. 6(2):87-103. 2015.
- BRITO, T. Oito estados brasileiros tiveram parâmetro muito alto ou hiperendêmico para hanseníase em 2017. **Jornalista da NHR Brasil (Netherlands Hanseniasis Relief – Brasil)**- Fortaleza-CE; 31, Jan 2018.
- CHAVES, E.C.; COSTA, S.V.; FLORES, R.L.R.; NEVES, E.O.S. Social deprivation index and leprosy in Pará State, Brazil, in 2013: spatial analysis. *Epidemiol. Serv. Saude, Brasilia*, 26(4). p. 90-99. Oct-Dec 2017.
- CHEN, Z. Data Mining and Uncertain Reasoning: An Integrated Approach. **New York: John Wiley & Sons Inc**; 2001.
- COOPER, G.F.; HERSKOVITS, E. A Bayesian method for the induction of probabilistic networks from data. **Machine Learning**. 9(4): 309-347. 1992.
- DIAS, J.S.; ABICALAFFE, C.L.L.; AMARAL, V.F. Aplicação da Rede Bayesiana na Prevenção da Gestação de Alto Risco. **Programa de Pós-Graduação em Tecnologia em Saúde**. Programa de Pós-Graduação em Ciências da Saúde. Pontifícia Universidade Católica do Paraná. 4p. 2000.
- FAYAAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. **The KDD process for extracting useful knowledge from volumes of data**. Communications of the ACM, New York: ACM Press, v.39, n.11, p.27-34, Nov. 1996.

- FISCHER, R.B. Desenvolvendo o conceito de redes bayesianas na construção de cenários prospectivos. **FGV RJ. Diss. Mestrado**. 2010.
- FRED, A.L.N. Redes Bayesianas. [**Material Didático**]. 2006.
- FREEDMAN, D.A. Ecological inference and the ecological fallacy. **International encyclopedia of the social & behavioral sciences**. (6): 4027-4030. 1999.
- FREITAS, L.R.S.; DUARTE, E.C.; GARCIA, L.P. Análise da situação epidemiológica da hanseníase em uma área endêmica no Brasil: distribuição espacial dos períodos 2001 - 2003 e 2010 - 2012. **Rev. Bras. Epidemiol.** ( 4): 702-713. Dez 20. 2017.
- FREITAS, L.R.S.; DUARTE, E.C; GARCIA, L.P. Leprosy in Brazil and its association with characteristics of municipalities: ecological study, 2009-2011. **Tropical Medicine and International Health**. volume 19, nr. 10. p.1216-1225. October 2014.
- FUNDACAO PRÓ-HANSEN. Hanseníase no Mundo, Hanseníase no Brasil e Hanseníase no Paraná. **Indicadores da Hanseníase**. Curitiba-PR. p. 1-2. 2011.
- GIRARDI, D.R.; TAKAHARA, A.L. Autoestima em Portadores de Hanseníase e Diabetes: Variáveis das Escalas Salsa e Participação com apoio das Redes Bayesianas. **Programa de Apoio à Iniciação Científica-PAIC 2013-2014. FAE-Centro Universitário | Núcleo de Pesquisa Acadêmica-NPA**. Pontifícia Universidade Católica do Paraná. p.483-504. 2014.
- GRACIE, R.; PEIXOTO, J.N.B.; SOARES, F.B.R.; HACKER, M.A.V.B. Análise da distribuição geográfica dos casos de hanseníase. Rio de Janeiro, 2001 a 2012. **Ciência & Saúde Coletiva**. 22(5):1695-1704. 2017.
- HÔHER, C.L.; FLORES, C.D; LADEIRA, M.; VICARI, R.M. Uma Experiência do Uso de Redes Probabilísticas no Diagnóstico Médico. In: **Argentine Symposium on Artificial Intelligence**. p.1-10. 2015.
- IMBIRIBA, E.N.B.; NETO, A.L.S.; SOUZA, W.V.; PEDROSA, V.; CUNHA, M.G.; GARNELO, L. Desigualdade social, crescimento urbano e hanseníase em Manaus: abordagem espacial. **Revista de Saúde Pública**. vol. 43, núm. 4. São Paulo-SP. p. 656-665. Agosto 2009.

- JALIER, V. Brasil ocupa o segundo lugar no mundo em casos de hanseníase: depoimento. *Revista Saúde Plena. 62ª Jornada Mundial de Hanseníase*. Entrevista concedida a *AFP-Agence France Presse*. p.1-5. 22 janeiro 2015.
- JALIER, V. Hanseníase, uma doença que diminui, mas não desaparece: depoimento. *Revista Ciência e Saúde. 62ª Jornada Mundial de Hanseníase*. Entrevista concedida a *AFP-Agence France Presse*. p.1-3. 22 janeiro 2015.
- KERR-PONTES, L.R.S.; BARRETO, M.L.; EVANGELISTA, C.M.N.; RODRIGUES, L.C.; HEUKELBACH, J.; FELDMEIERS, H. Fatores de risco socioeconômicos, ambientais e comportamentais para a hanseníase no Nordeste do Brasil: resultados de um estudo de caso-controle. *Intern Jour of Epidemio*. 35 ( 4): 994–1000. Ago 2006.
- KOEHLER, C. Desenvolvimento de um Sistema Inteligente para Apoio à Decisão em Saúde. **Universidade de Caxias do Sul, Campus Universitário de Vacaria, Laboratório de Informática Médica**. Vacaria-RS. p.1-8; 2002.
- LAPA, T.; XIMENES, R.; SILVA, N.N.; SOUZA, W.; ALBUQUERQUE, M.F.M.; CAMPOZANA, G. Vigilância da hanseníase em Olinda, Brasil, utilizando técnicas de análise espacial. *Caderno Saúde Pública*. Rio de Janeiro-RJ, 17(5): 1153-1162, set-out, 2001.
- MATOS, H.J.; DUPPRE, N.; ALVIM, M.F.S.; VIEIRA, L.M.M.; SARNO, E.N.; STRUCHINER, C.J. Epidemiologia da hanseníase em coorte de contatos intradomiciliares no Rio de Janeiro (1987-1991). *Caderno Saúde Pública*. Rio de Janeiro-RJ, 15(3): 533-542, jul-set, 1999.
- MELINSKI, M.C.; GIRARDI, D.R. Diabetes e Hanseníase: estigma, cura, análise por meio de Redes Bayesianas. **FAE - Centro Universitário | Núcleo de Pesquisa Acadêmica**. PUC-PR. 2014.
- MOURA, L.T.R.; LUNA, I.C.F.; VIEIRA, M.C.A. Perfil clínico-epidemiológico da hanseníase em menores de 15 anos no município de Juazeiro-Ba. *Rev. Bras. Promoc Saude*. 26(2): 208-215. Abr/Jun 2013.

- NEGRÃO, G.N.; VIEIRA, I.R.; KATAYAMA, E.M.Y.; BORECKI, M.T. Variáveis Epidemiológicas Intervenientes na Ocorrência da Hanseníase no Município de Guarapuava, PR. *Geografia (Londrina)*. v. 25. n.2. p. 110-129, jul/dez, 2016.
- OLIVEIRA, J.C.F.; LEÃO, A.M.M.; BRITTO, F.V.S. Análise do perfil epidemiológico da hanseníase em Maricá, Rio de Janeiro: uma contribuição da enfermagem. *Rev Enferm UERJ*. 22(6):815-21. Nov/dez 2014.
- OLIVEIRA, K.S.; SOUZA, J.; CAMPOS, R.B.; ZILLY, A.; SILVA-SOBRINHO, R.A. Avaliação dos indicadores epidemiológicos e operacionais para a hanseníase em municípios prioritários no estado do paran , 2001 a 2010. *Epidemiol.Serv. Saude*, Brasilia, 24(3):507-516, jul-set 2015.
- ONU. Declara o do Mil nio das Na es Unidas. 2000.
- PEARL, J. Probabilistic reasoning in Intelligent System, Morgan Kaufmann Publishers (1988).
- ROCHA, C.A.J. Estrat gia de otimiza o para a melhoria da interpretabilidade de redes bayesianas: aplica es em sistemas el tricos de pot ncia. *Tese Dout. UFPA*. 2009.
- RODRIGUEZ, B. Hansen ase: Brasil   o  nico pa s que n o conseguiu eliminar sua propaga o. *Portal Fiorocruz. Funda o Osvlado Cruz*. Rio de Janeiro-RJ. p. 1-3. 2018.
- RUSSEL, S.J.; NORVIG, P. Artificial Intelligence – A Modern Approach. Prentice Hall, Upper Saddle River, NJ, 2a. ed. (2002).
- SAHEKI, A.H. Constru o de uma Rede Bayesiana Aplicada ao Diagn stico de Doen as Card acas. **Disserta o. Departamento de Engenharia Mecatr nica e de Sistemas Mec nicos. Escola Polit cnica da Universidade de S o Paulo**. S o Paulo-SP.70p. 2005.
- SALES, A.M.; LEON, A.P.; DUPPRE, N.C.; HACKER, M.A.; NERY, J.A.C.; SARNO, E.N.; PENNA, M.L.F. Hansen ase entre os contatos com pacientes: um estudo multin vel sobre fatores de risco. *PLOS Neglected Tropical Diseases* 5 (8): 10.137. 2011.
- SATO, R.C.; SATO, G.T.K. Modelos probabil sticos gr ficos aplicados   identifica o de doen as. **Instituto de Ci ncias e Tecnologia. Universidade Federal de S o Paulo**. S o Paulo-SP. *Revendo Ci ncias B sicas*. Einstein. 13(2):330-3. 2015.

- SEFFRIN, H.M.; JAQUES, P. Modelando o conhecimento algébrico do estudante através de Redes Bayesianas Dinâmicas. **III Congresso Brasileiro de Informática na Educação. XXV Simpósio Brasileiro de Informática na Educação**. 2014.
- SILVA, W.R.S. Monitoramento de epidemia de dengue na Amazônia usando Redes Artificiais. **Dissertação. (Programa de Pós-Graduação em Engenharia Elétrica) | Instituto de Tecnologia. Universidade Federal do Pará**. Belém-PA. 91p. 2018.
- SUZUKI, K.; ARAMA, T.; KAWASHIMA, A.; YOSHIHARA, A.; YOTSU, R.R.; ISHII, N. Estado atual da hanseníase: epidemiologia, ciência básica e perspectivas clínicas. **The Journal of dermatol.** 39 (2): 121–9. Fev 2012.
- TENÓRIO, J.M. Aplicação de Técnicas de Inteligência Artificial ao desenvolvimento de um sistema de apoio à decisão para doença celíaca. **Dissertação. (Programa de Pós-Graduação em Informática em Saúde) | Escola Paulista de Medicina**. Universidade Federal de São Paulo. São Paulo-SP. 129p. 2011.
- VIANA, R.L.; FREITAS, C.M.; GIATTI, L.L. Saúde ambiental e desenvolvimento na Amazônia legal: indicadores socioeconômicos, ambientais e sanitários, desafios e perspectivas. **Saúde Soc. São Paulo**, v.25, n.1, p.233-246. 2015.
- VICARI, R.M.; FLORES, C.D.; PEROTTO, F. Sistemas Baseados em Conhecimento para a Área da Saúde. **Pós-Graduação em Computação. Instituto de Filosofia e Ciências Humanas**. Universidade Federal do Rio Grande do Sul. Porto Alegre-RS. 9p. 2002.
- WAGNER, M.B. Medindo a ocorrência da doença: prevalência ou incidência?. **Departamento de Medicina Social. Faculdade de Medicina**. Universidade Federal do Rio Grande do Sul. *Jornal de Pediatria*. 74:157-162. 1998.



## INCIDENCE OF LEPROSY IN THE BRAZILIAN AMAZON REGION AND ITS ANALOGY IN RELATION TO BRAZILIAN MUNICIPALITIES: A SOCIODEMOGRAPHIC ANALYSIS USING BAYESIAN NETWORKS

<sup>\*1,2</sup>José Maria da Silveira Gomes, <sup>2</sup>Aleksandra do Socorro da Silva, <sup>2</sup>Silvana Rossy de Brito, <sup>1,3</sup>Eulália Carvalho da Mata, <sup>4</sup>Regina Fatima Feio Barroso and <sup>1</sup>Carlos Renato Lisboa Francês

<sup>1</sup>Instituto de Tecnologia, Universidade Federal do Pará, Pará, Brazil

<sup>2</sup>Instituto Ciberespacial, Universidade Federal Rural da Amazônia, Pará, Brazil

<sup>3</sup>Campus de Capitão Poço, Universidade Federal Rural da Amazônia, Pará, Brazil

<sup>4</sup>Instituto de Ciências da Saúde, Universidade Federal do Pará, Pará, Brazil

### ARTICLE INFO

#### Article History:

Received 28<sup>th</sup> August, 2018

Received in revised form

25<sup>th</sup> September, 2018

Accepted 19<sup>th</sup> October, 2018

Published online 30<sup>th</sup> November, 2018

#### Key Words:

Data mining  
Incidence of leprosy  
Public health  
Bayesian networks  
Brazilian Amazon.

### ABSTRACT

Considering the importance of studying the incidence of leprosy and the scarcity of studies that evaluate the occurrence of this phenomenon in the Brazilian Amazon region in relation to the Brazil, this study aimed to analyse the association between leprosy incidence and the municipal indicators of education, income and domicile population density, comparing municipalities of the Amazon region to those of the whole country. We used data obtained on cases of leprosy in Brazil, available by the Information System of Hardship Notifications through the Informatics Department of the National Health Service and the socio-economic indicators found in the last Demographic Census of the Brazilian Institute for Geographical and Statistical Survey, information from the Municipal Human Development Index, regarding education and income, obtained from the website of the Human Development Atlas of Brazil. The methodology combined data mining with the analysis of spatial distribution. We found that municipalities of the Brazilian Amazon region present a high rate of leprosy incidence (65.7%); but this value declines (13.1%) when the analysis contemplated other regions of the country. Using the Bayesian network model, was a significant association between the percentage of homes with more than 2 inhabitants and the rate of incidence of leprosy.

Copyright © 2018, José Maria da Silveira Gomes et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Citation:** José Maria da Silveira Gomes, Aleksandra do Socorro da Silva, Silvana Rossy de Brito, Eulália Carvalho da Mata, Regina Fatima Feio Barroso and Carlos Renato Lisboa Francês, 2018. "Incidence of leprosy in the Brazilian Amazon region and its analogy in relation to Brazilian municipalities: a sociodemographic analysis using Bayesian networks", *International Journal of Development Research*, 8, (11), 24189-24194.

### INTRODUCTION

The Millennium Declaration, elaborated by the United Nations Organization, was based on eight Millennium Development Goals, established by the member states, through a series of assurances to be met within a fixed timescale, in order to improve the future of the human race in this century. The goals, with targets and indicators, were adopted by 164 countries. The sixth goal concerns the combat against diseases that threaten above all the poor and most vulnerable populations, where such countries should strive to reduce poverty through access of the population to information and

provide means of prevention and treatment, without neglecting environmental care and nutrition that can stop the spread of these diseases (UN, 2000). In Brazil, in spite of the significant reduction in mortality due to infectious diseases, some diseases are still epidemic such as dengue and other diseases spread by vectorial contamination, and even endemic diseases are persistent, such as tuberculosis and leprosy, which were expected to be under control (Barreto, 2013). The so-called neglected diseases, which include leprosy, are serious hardships that persist in the country and are related to poverty, poor health and social exclusion (Augusto, 2016). Leprosy is a chronic contagious disease caused by the bacterium *Mycobacterium leprae*. It constitutes an enormous social problem in view of its capacity to infect a massive number of individuals, although few become ill. Its incidence is greatest among the economically lowest classes, due to overexposure, low educational level, nutrition and living conditions (Brito et

**\*Corresponding author:** José Maria da Silveira Gomes  
Instituto Ciberespacial, Universidade Federal Rural da Amazônia, Pará, Brazil.

al. 2014). Indeed, the home is regarded as an important location for the transmission of the disease (Oliveira et al. 2014). Leprosy is still a serious problem in public health. Brazil is ranked as the second country with the greatest number of new cases detected in the world, after India. Although the annual detection rate is in decline, in 2010 the global incidence of leprosy was around 212 thousand cases. In Brazil, in 2014, of the 31,064 new cases, 2,341 involved children under 15 and the general detection rate was 15.32 cases per 100 thousand inhabitants. Some municipalities in Brazil have been able to eliminate the disease, however, the north, central-west and northeast regions still record high levels of the disease. In spite of an overall tendency for the detection rate stabilizing in Brazil, there are large regional differences (Gracie et al. 2017). Leprosy is still a public health problem in Brazil and requires more effective public policies. Late diagnosis, the refusal to seek services for treatment and negligence with self-care can lead to deformities and physical incapacities that, besides the financial burden for health services, result in poor self-respect and interfere in an individual's perception of the world, culminating in avoiding social relations. The stigma suffered, especially with leprosy, is particularly severe, where men often are the hardest hit and more difficult to cure (Melinski e Girardi, 2014). The establishment of effective public policies for prevention, control and treatment, requires an efficient system for the registration, treatment and recovery of information so that the diagnosis, therapeutics and epidemiology may be correctly conducted. The use of computational systems combined with the application of probabilistic models, such as the Bayesian network model could be an effective way to reach efficient solutions as an aid for health administrators in decision making. Considering the importance of studying the incidence of leprosy and the scarcity of studies that evaluate the occurrence of this phenomenon in the Brazilian Amazon region in relation to the whole country, this study aims to analyse the association between leprosy incidence and the municipal indicators of education, income and domicile population density, comparing municipalities of the Amazon region to those of the whole country

## MATERIALS AND METHODS

As a source of data, we used the Research on Homes within the database of the Demographic Census of the Brazilian Institute for Geographical and Statistical Survey – IBGE<sup>1</sup>, with regard to the Brazilian municipalities for the year 2010, the latest census carried out. Data were collected from this database for variables concerning the dwelling conditions of the inhabitants. Overall, the sample covered 57,320,474 homes in 5565 municipalities, where 771 were located in the Brazilian Amazon region. Some 42,017 records of leprosy incidence in the municipalities were found in the DATASUS database (Informatics Department of the National Health Service – SUS<sup>2</sup>) of the Information System of Hardship Notifications (SINAN), again regarding Brazilian municipalities for the year 2010. In addition, components of the Municipal Human Development Index (IDHM), regarding education (IDHM-Education) and income (IDHM-Income) were obtained from the website Human Development Atlas of Brazil<sup>3</sup>, also for the year 2010. The geospatial database for

Brazilian municipalities was obtained from IBGE<sup>4</sup> in *shape file* format, that is operable by the majority of geographic information systems. From these various databases ten variables were selected with regard to the dimensions of leprosy, habitation, income and education, shown in Table 1.

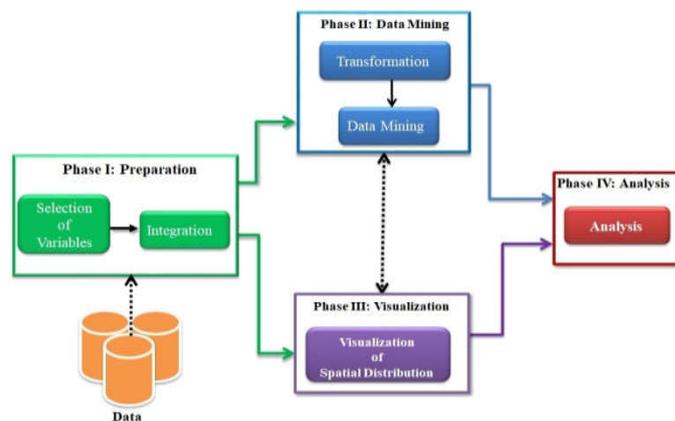


Figure 1. Methodology in phases for data analysis

In phase I (Figure 1), data preparation, the variables were selected and integrated in a single database, resulting in 5565 records of identified municipalities (code and name of municipality) with 10 associated attributes (variables). For integration, pre-processing was necessary, that is, the data were scrutinized so as to eliminate cases of redundancy and inconsistency. In the second phase the two following steps were undertaken: transformation and data mining. For this purpose, only those variables under study were selected, ignoring the identification of the municipality (code and name). This resulted in a matrix of 5565 rows (municipalities) by 10 columns (variables). Transformation involved the application of algorithms of discretization aimed at placing values in intervals such that a limited number of possible states were obtained. This step is important because in a search for association it is not possible to use continuous values since a single variable can be made up of a large quantity of different values, such that it would be impossible to identify relevant associations. The discretization method used was based on frequency, that is, an equilibrium was sought as to the number of records in each range of values, so that ideally each discretization range would contain the same number of records. Thus, all the continuous variables (*Piped Water Supply, Bathroom Piped Water Supply, Rubbish Collection, Electricity Supply, PopDensityAbove2, Leprosy, IDHM-Life Expectancy, IDMH-Income, IDHM-Education*), were put through the process of discretization by frequency. Following this, still in the data-mining phase, the Bayesian network technique was applied in order to measure the association between the studied variables. Bayesian networks permit the construction of models formed by a qualitative structure, defined by directed acyclic graphs, where the dependencies between nodes (variables) and a quantitative structure (conditional probability tables of the nodes) could be represented, enabling these dependencies to be expressed in probabilistic terms (Chen, 2001). These structures offer an efficient representation of the probability distribution of the set of variables under study. Thus, for each node, a distribution of locality probability exists, depending on the status of their parents.

<sup>1</sup> <http://sidra.ibge.gov.br>

<sup>2</sup> <http://www2.datasus.gov.br>

<sup>3</sup> <http://atlasbrasil.org.br>

<sup>4</sup> [http://downloads.ibge.gov.br/downloads\\_geociencias.htm](http://downloads.ibge.gov.br/downloads_geociencias.htm)

Table 1. Selected variables and data source

Variable	Description	Data source
<i>Piped Water Supply</i>	Percentage of homes with piped water supply. Calculated as the ratio between the number of people living in permanent private homes with water on supply to one or more rooms and the total population resident in permanent private homes, multiplied by 100. The water may be from a general pipe system, a well, a spring or reservoir supplied with rainwater or a tanker.	IBGE
<i>Bathroom Piped Water Supply</i>	Percentage of homes with a bathroom and piped water supply. Calculated as the ratio between the number of people living in a permanent private home with water on supply to one or more rooms and a separate bathroom and the total population resident in permanent private homes, multiplied by 100. The water may be from a general pipe system, a well, a spring or reservoir supplied with rainwater or a tanker. A separate bathroom is defined here as a room with a shower or bath and a toilet bowl.	IBGE
<i>Rubbish Collection</i>	Percentage of homes with rubbish collection. Calculated as the ratio between the number of people living in homes with rubbish collection and the total population resident in permanent private homes, multiplied by 100. This includes situations where rubbish is collected directly by a public or private firm, or rubbish deposited in a dump bucket of any kind outside the residence for posterior collection. Only permanent private homes in an urban area are considered.	IBGE
<i>Electricity Supply</i>	Percentage of homes with electricity supply. Calculated as the ratio between the number of people living in a permanent private home with electricity and the total population resident in permanent private homes, multiplied by 100. Electricity supply is considered whether coming from a general network or not, and with or without a meter.	IBGE
<i>Pop Density Above2</i>	Percentage of homes with a population density above 2 people per bedroom. This is the ratio between the number of people living in a permanent private home with a population density above 2 and the total population resident in permanent private homes, multiplied by 100. The population density of the home is the ratio between the total number of inhabitants of the home and the total number of rooms used as bedrooms.	IBGE
<i>Leprosy</i>	Incidence of leprosy in the municipality. Calculated as being the quotient between the number of cases multiplied by 100,000 habitants and the total population.	DATASUS
<i>IDHM-Life Expectancy</i>	Corresponds to the life expectancy according to IDHM. Calculated from the life expectancy at birth; shows the number of years that a person born in a determined municipality will live, maintaining the same mortality rate.	ATLAS BRAZIL
<i>IDHM-Income</i>	Corresponds to the income component of the IDHM. It is measured as the municipal income <i>per capita</i> , or, in other words, the average income of the residents of a determined municipality.	ATLAS BRAZIL
<i>IDHM-Education</i>	Corresponds to the education component of the IDHM. Calculated from the geometric mean of two sub-indexes: the frequency of children and young people in school (weight factor 2/3) and the educational level of the adult population (weight factor 1/3).	ATLAS BRAZIL
<i>Spatial Reference</i>	Indicates whether the municipality is located in the Amazon region or in another region of the country, assigned values as "Amazon" or "Other", respectively.	IBGE

In order to test the Bayesian network structure, an algorithm for a K2 heuristic search (Cooper e Herskovits, 1992) was applied, based on a matrix of 5565 rows x 10 columns. The K2 algorithm learns the structure of the Bayesian network from the matrix applied using a method that selects the structure with maximum true similarity, through a repetitive search that initiates with a very simple network structure where all the variables are dependent on one another. Then, an evaluation is made of marginal true similarities of each network resulting from possible changes (such as, the addition of a new link between two nodes of the network) followed by the application of the best alterations before starting a new repeat process. This procedure continues until the algorithm is no longer able to find any single alteration that increases the maximum similarity. The resulting network may contain a number of dependent variables lower than the number of variables in the matrix. This algorithm, classified as a search tool, was selected according to its performance in terms of computational complexity and precision, obtained from an adequate sorting of variables. In this study, the target variable (leprosy) was placed in the last column of the matrix. Once the Bayesian network was established, statistical analysis was used to estimate the posterior distribution of the parameters. The basic task of analysis consisted in computing the conditional probability distribution using quantitative information of the Bayesian network, that is, for any set of consultation variables (*Consult*), given the values of the observed event (*Evidences*), the system computes  $P(\text{Consult} | \text{Evidences})$ . In order to evaluate the relevance of the association rules produced, the model support-confidence (Agrawal *et al.*, 1993) was used that permits evaluation of the quality of the association found. In the third phase, the software Qgis<sup>5</sup> was employed and the 5565 municipalities

integrated in the *shap efile* file, that contains the georeferenced information of the municipalities, in order to present the spatial distribution of leprosy incidence in Brazil. The maps obtained in this phase, together with the results of data mining, were analyzed in phase IV.

## RESULTS

The data were selected and compiled in a single database, as described in the Methods section for the first phase. In the second phase, all the continuous variables were subjected to discretization, where intervals of the same frequency were considered for distribution of each value in the created categories. In the case of the variable *Leprosy*, due to the elevated number of municipalities (2,193) without any recorded cases of the disease, it was necessary to first classify these municipalities as "NoCases" and then apply the method of discretization by frequency for the remaining municipalities (3,372). Thus, the municipalities with cases of the disease were classified in three ranges of values, of equal frequency (Table 2), that is, 1,124 municipalities in each category of disease incidence ("low", "medium", "high") Other variables were discretized in four range of values (Table3). The means for the rate of incidence of leprosy for the municipalities of the Amazon and other regions are presented in Table 4 and include a list of the top ten Brazilian municipalities with the highest incidence of the disease. Although the average for the Amazon region is greater than for other regions of the country, the top ten list contains only five municipalities of the Amazon. The others are three from the state of Mato Grosso (MT), one from the state of Tocantins (TO) and one from the state of Pará (PA). In the Amazon, the municipality with the greatest incidence is Araguaiana (state of MT), with 563.03 and second is Jacareacanga (state of PA) with 354.53.

<sup>5</sup> <http://qgis.org>

Table 2. Classification of incidence rates

Rate of Incidence Category	Number of Municipalities	(%) municipalities in each category	Interval of the category
NoCases	2,193	39.40%	[0]
Low	1,124	20.20%	(0 – 14.09]
Medium	1,124	20.20%	(14.09 – 37.46]
High	1,124	20.20%	(37.46– 840.97]

Table 3. Classification of IDHM-Income, IDHM-Education, Bathroom Piped Water Supply and Pop Density Above2

Variable	Category	(%) municipalities in category	Interval of category
<i>IDHM-Income</i>	Very Low	24.9%	[0.400 – 0.572]
	Low	25.0%	(0.572 – 0.654]
	Medium	24.9%	(0.654 – 0.707]
	High	25.2%	(0.707 – 0.891]
<i>IDHM-Education</i>	Very Low	24.9%	[0.207 – 0.490]
	Low	24.9%	(0.490 – 0.560]
	Medium	25.2%	(0.560 – 0.631]
	High	25.0%	(0.631– 0.825]
<i>BathroomPipedWaterSupply</i>	Very Low	24.6%	[3.26 – 67.78]
	Low	25.6%	(67.78 – 91.25]
	Medium	25.2%	(91.25 – 98.00]
	High	24.6%	(98.00 – 100.0]
<i>PopDensityAbove2</i>	Very Low	24.7%	[0.65 – 15.41]
	Low	25.2%	(15.41 – 23.07]
	Medium	25.5%	(23.07 – 32.58]
	High	24.7%	(32.58 – 88.64]

Table 4. Leprosy, as the Rate of Incidence - mean (Amazon, Other) and selected municipalities

Region	Municipality	Rate of Incidence
Amazon	(mean)	73.23
Other regions	(mean)	16.98
MS	Pedro Gomes	840.97
GO	Campos Verdes	697.21
MT	Araguaiana	563.03
SE	Amparo de São Francisco	483.52
GO	Padre Bernardo	458.96
TO	Novo Jardim	366.30
PA	Jacareacanga	354.53
MT	Itaúba	349.73
MT	Nova Nazaré	330.14
GO	Jussara	313.27

The maximum value for this indicator (840.97) belongs to the municipality of Pedro Gomes (state of MS). On expanding Table 4 for the first fifty municipalities with the highest incidence, all are municipalities small in size (up to 50 thousand inhabitants) and 31 of these are located in the Amazon region. Of the total municipalities of the Amazon (771), 498 municipalities have a high incidence rate of leprosy (above 37.46), which represents 64.59% of the Amazonian municipalities; when considering other regions of Brazil only 13.06% of the municipalities fall into this category. When the spatial reference refers to the Brazilian states, the ten states with the most cases of leprosy are headed by Maranhão and Pará, which, together, present more than 7000 cases (Figure 2). The high and mean leprosy incidence rate for the state of Pará are found, respectively, in 60.84% and 29.37% of the municipalities. In the state of Maranhão the high incidence rate of the disease is found in 58.06% of the municipalities and the mean in 22.58%. The Bayesian network technique was then used in order to measure the association between the studied variables. Applying the algorithm K2, relevant associations were found for the following variables: *Spatial Reference*, *IDHM-Income*, *IDHM-Education*, *Bathroom Piped Water Supply*, *PopDensityAbove2*. At this stage the remaining variables were discarded from the association study. The resulting Bayesian network (Figure 3) revealed that the variable *Leprosy* is directly dependent on the variables *Pop Density Above2* and *Spatial Reference*.

The inferences presented as follows were obtained from the results of the propagational algorithms. When the data concern a municipality belonging to the Amazon region, the probability of presenting a high rate of leprosy incidence is 65.7%; this value falls to 13.1% when the location is for other regions of the country. When the data show a municipality with a lower *PopDensityAbove2* (<15.41%), the probability that this municipality has no recorded disease is 60%; on the other hand when the highest range is presented (above 32.58%) this probability drops to 22.7%. When looking at the two variables (*Spatial Reference* = Amazon and *PopDensityAbove2* <15.41), this probability becomes 56.1% (Figure 4); otherwise, when considering other regions of the country, this value is reduced to 17.9%. The presence of a bathroom with a piped water supply, according to the Census data, is a typical problem in the Amazon region: for such municipalities, the probability of lower percent values for the presence of such a bathroom is 56.1%; this percentage drops to 19.6% for other regions. Indeed, the case where the incidence of *Leprosy* is high, the *PopDensityAbove2* is also high and the *Bathroom Piped Water Supply* incidence range is low, is a typical scenario for the Amazon region (probability of 73% against 27% for other regions). The variables *IDHM-Income* and *IDHM-Education* did not show a direct association with *Leprosy*. However, *IDHM-Income* was directly associated with *Bathroom Piped Water Supply*: in the lowest range for *IDHM-Income*, the probability of a municipality having the lowest incidence of

Bathroom Piped Water Supply was 74.2%; on the other hand, for the high range of *IDHM-Income* this probability drops to zero. The variables *IDHM-Income* and *IDHM-Education* were related to the population density of the home: when the indicators of income and education were in the lowest range, the probability that the municipality would have the highest range of *PopDensityAbove2* was 54.2%; this probability decreases to 4.7% when the indicators of income and education are in the highest range. In order to illustrate, as a percentage, the incidence of Leprosy in Brazilian municipalities, a map showing the geographical distribution of this disease is presented in Figure 5, according to the classification ranges defined by the process of discretization. For example, for the different value ranges used in the *Incidence Rate*, 60.6% of the Brazilian municipalities present a classification for the disease in the ranges low, medium and high. The geographical distribution revealed that a pattern for high incidence of the disease predominated in the states of Pará, Mato Grosso, Goiás, Pernambuco and Bahia and that, above all, there was a relationship of contiguity among the municipalities with high incidence rates for the disease.

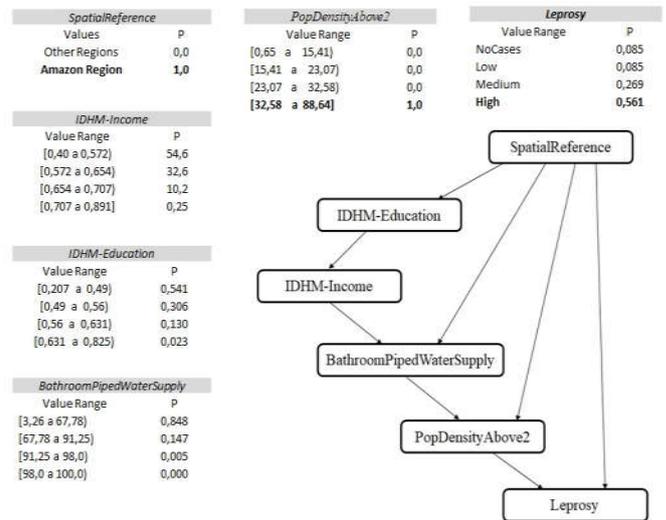


Figure 4. Bayesian network (data: *SpatialReference*="Amazon" and *PopDensityAbove2* >32.58).

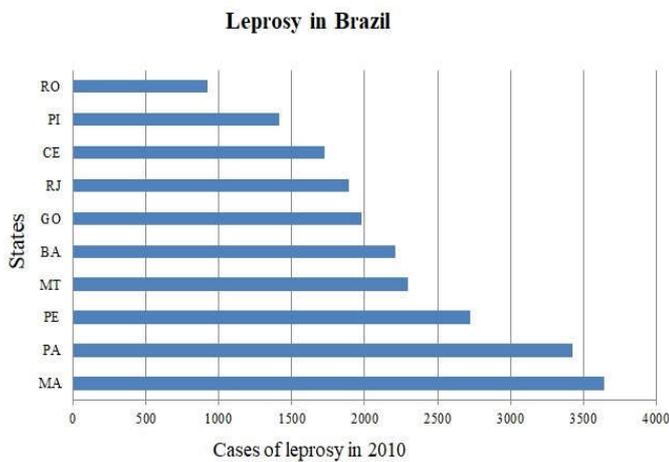


Figure 2. Cases of leprosy in Brazilian states.

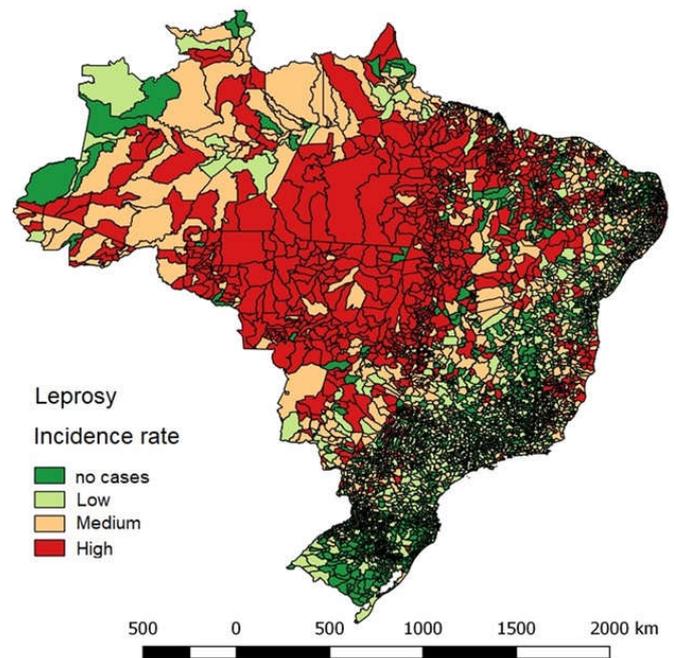


Figure 5. Map of leprosy cases in Brazilian municipalities.

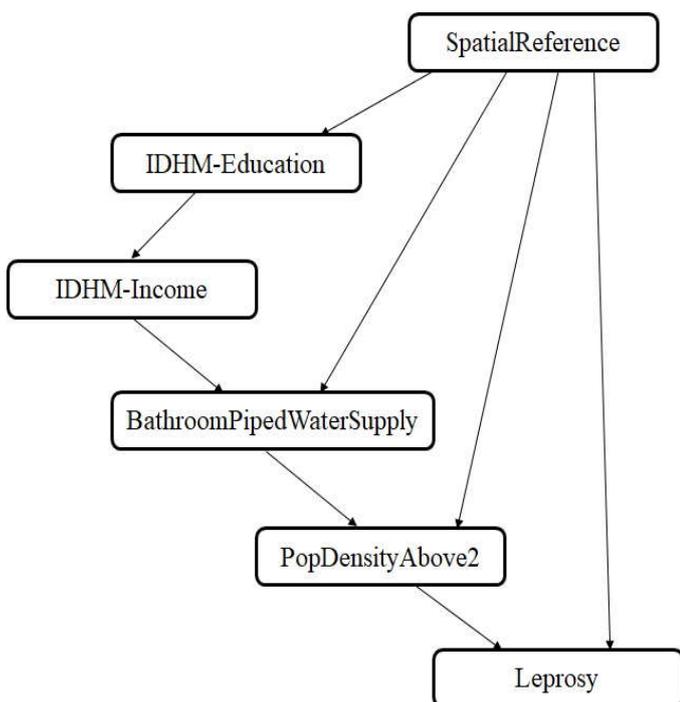


Figure 3. Structure of the selected Bayesian network

### DISCUSSION

The present study demonstrates an association between the incidence rate of leprosy and variables that represent the dimensions of income, education, bathrooms with a piped water supply and homes with a population density greater than 2, using data for Brazilian municipalities for the year 2010. The percentages of the *Leprosy Incidence Rate* were greater for the municipalities of the Amazon region than for the whole country (Table 4). Furthermore, this phenomenon was observed to be typical of municipalities with the lowest populations, although the greatest number of cases are recorded in the largest municipalities and metropolises (Table 5). The five states with the greatest *leprosy incidence rates* are respectively: Goiás, Mato Grosso, Tocantins, Maranhão and Pará, and of these, four are located in the Amazon region forming an area, together with Rondônia, where the disease is endemic in Brazil (Freitas *et al.* 2017). In the Bayesian network model a significant association exists between the percentage of homes with a population density above 2 and the incidence of leprosy.

**Table 5. Cases and incidence rate of leprosy according to the size classification of the municipality**

Size of Municipality	N° Cases		Incidence rate	
	Minimum	Maximum	Minimum	Maximum
Metropolis	737	988	72.62	64.25
Large	45	703	43.73	86.34
Medium	21	140	40.66	189.08
Small 1	1	67	37.55	840.97
Small 2	8	127	38.75	458.96

Although the relationship between the incidence rate and socio-economic factors (no water supply, bathroom in the home, poverty and the agglomeration of people in the home), low educational indexes and income have been reported in several studies (Ponnighaus *et al.* 1994; Kerr-Pontes *et al.* 2006; Queiroz *et al.* 2009; Sales *et al.* 2011; Suzuki *et al.* 2012; Moura *et al.* 2013), the inclusion of the variable that considers homes with a population density greater than 2 adds a further element to the problem without conflicting with the conclusion of those studies. Indeed, the municipalities whose homes have the greatest population density tend to be the ones that also concentrate the worst level of family income (and vice versa), and, equally, the worst level of education (and vice versa). As carried out in the study of Freitas *et al.* 2017, our study investigated the association between the incidence of leprosy in a collective sense (municipalities) and not individual, that is, an association observed for the municipality not necessarily occurs for individuals, which could lead to errors of ecological fallacy. Ecological fallacy can arise when considering that relationships for groups are maintained for individuals (Freedman, 1999). In our study, the inferences obtained are supported by the aggregated data and, therefore, cannot be extended to the level of the individual. In spite of this limitation, the association rules found here support the studies carried out whose arguments relate this phenomenon to the environment in which the individuals, afflicted with leprosy, are interposed. Knowledge of the vulnerability of people to transmittable diseases helps decision making with regard to aspects of health needs, due to social exclusion and fear. Identification of the variables and factors involved can lead to the development of strategies that may diminish the advance of diseases such as leprosy. Although the incidence rate of leprosy involves social factors related to vulnerability, for its prevention it is fundamental to provide information and adequate orientation, intensify actions to combat leprosy, including early diagnosis, prevention, treatment, physical and social rehabilitation, as well as promoting access to diagnosis and treatment in the municipalities with the higher levels of endemism. The scenario found after the analysis of the data reveals the peripheralization of Brazil, which represents a barrier to guarantee equality for the access to health services and information. Três Arrois (state of Rio Grande do Sul), for example, is a municipality that has 0.65% of a habitational density above 2; while at the other extreme, Uiramutã (state of Roraima) 88.64%, Santa Isabel do Rio Negro (state of Amazonas) 87.59%, Melgaço (state of Pará) 86.43% and Maraã (state of Amazonas) 82.75%, all in the Amazon region.

## REFERENCES

Agrawal, R., Imielinski, T. and Swami, A. 1993. Mining association rules between sets of items in large databases. IBM Almaden Research Center 650 Harry Road. San Jose, CA 95120.

- Augusto, DS. 2016. Neglected Diseases: Study of the knowledge and practices of teachers of primary education in the municipality of Jaboatão dos Guararapes/state of Pernambuco. Recife [Masters Dissertation] (*in portuguese*) FIOCRUZ
- Barreto, M L. 2013. Outline of a scenario of health conditions of the Brazilian population in 2022/2030. In Cruz, F.O. (Ed.). Health in Brazil in 2030: strategic prospection of the Brazilian health system – population and sanitary profile (*in portuguese*). Rio de Janeiro: Fiocruz / Ipea / Ministério da Saúde, p.97-120.
- Brito KKG, Araújo DAL, Uchôa REMN, Ferreira, JDL, Soares, MJGO, Lima, JO. 2014. Epidemiology of leprosy in a state of the northeast of Brazil (*in portuguese*). *Rev. Enferm, UFPE*. 8(8): 2686-93.
- Chen, Z. 2001. Data Mining and Uncertain Reasoning: An Integrated Approach. New York: John Wiley & Sons Inc,
- Cooper, GF. and Herskovits, E. 1992. A Bayesian method for the induction of probabilistic networks from data. *Machine Learning*, 9(4): 309-347.
- Freedman, DA. 1999 Ecological inference and the ecological fallacy. *International encyclopedia of the social & behavioral sciences*, (6): 4027-4030.
- Freitas LRS, Duarte, and EC Garcia, LP. 2017. Analysis of the epidemiological situation of leprosy in an endemic area of Brazil: spatial distribution for the periods 2001-2003 and 2010-2012 (*in portuguese*). *Rev. Bras. Epidemiol*, 20( 4): 702-713.
- Gracie, RP, JNB, Soares FBR, and Hacker MAVB. 2017. Analysis of the geographic distribution of leprosy cases in Rio de Janeiro 2001 to 2012 (*in portuguese*). *Ciência & Saúde Coletiva*, 22(5):1695-1704.
- Kerr-Pontes, LRS., Barreto, ML., Evangelista, CMN., Rodrigues, LC., Heukel bach, J. and Feldmeier H. 2006. Socio-economic, environmental and behavioural risk factors for leprosy in the northeast of Brazil: results of a case-control study (*in portuguese*). *Intern Jour of Epidemio*, 35 ( 4): 994–1000.
- Melinski, MC and Girardi, DR. 2014. Diabetes and Leprosy: stigma, cure, analysis using Bayesian Networks (*in portuguese*). FAE - Centro Universitário | Núcleo de Pesquisa Acadêmica. PUC-PR.
- Moura, LTR., Luna, ICF. and Vieira, MCA. 2013 Clinical-epidemiological profile of leprosy in children under 15 years of age in the municipality of Juazeiro, Bahia state (*in portuguese*). *Rev. Bras. Promoc Saude*, 26(2): 208-215.
- Oliveira, JCF., Leão, AMM. and Britto, FVS. 2014. Analysis of the epidemiological profile of leprosy in Maricá, Rio de Janeiro: a contribution of nursing (*in portuguese*). *Rev Enferm. UERJ*, 22(6):815-21.
- Pönnighaus, JM. *et al.* 1994. Rate of leprosy incidence in the Karonga district, north of Malauí: profiles by age, sex, status and BCG classification (*in portuguese*). *Int. J. of Leprosy and Other Mycobact. Diseases*. 62 (1): 10-23.
- Queiróz, ML. and Scatena, JHG. 2009. Spatial and temporal distribution of leprosy in Mato Grosso state, for the period 1996 to 2007 (*in portuguese*). *Cad. Saúde Colet*, 17 (1): 145 – 161.
- Sales, AM *et al.* 2011. Leprosy and patient contact: a multilevel study of risk factors (*in portuguese*). *PLOS Neglected Tropical Diseases*, 5 (8): 10.137.
- Suzuki, K., Takeshi, A., Kawashima, A., Yoshihara, A., Yotsu, RR. and Norihisa, ISHII. 2017. Present state of leprosy: epidemiology, basic science and clinical perspectives (*in portuguese*). *The Jourl of dermatol, Fev*, 39 ( 2): 121–9.
- UN. United Nations Millenium Declaration. 2000.