

Análise da Associação entre indicadores do trabalho infantil e gravidez na adolescência na Amazônia Legal brasileira

Leonardo Lopes da Costa
Universidade Federal Rural da
Amazônia

Av. Pres. Tancredo Neves, Nº 2501
Cep 66.077-830, Belém-Pará-Brasil
leolopes671@gmail.com

Tayná de Oliveira Lopes
Universidade Federal Rural da
Amazônia

Av. Pres. Tancredo Neves, Nº 2501
Cep 66.077-830, Belém-Pará-Brasil
taylopes23@gmail.com

Aleksandra do Socorro da Silva
Universidade Federal Rural da
Amazônia

Av. Pres. Tancredo Neves, Nº 2501
Cep 66.077-830, Belém-Pará-Brasil
aleksandra.silva@ufra.edu.br

Silvana Rossy de Brito
Universidade Federal Rural da
Amazônia

Av. Pres. Tancredo Neves, Nº 2501
Cep 66.077-830, Belém-Pará-Brasil
silvana.rossy@ufra.edu.br

Maurílio de Abreu Monteiro
Universidade Federal do Sul e
Sudeste do Pará

Folha 31, Q7, s/n - Nova Marabá
Cep 66.077-830, Marabá-Pará-Brasil
maurilio@naea.ufpa.br

Carlos Renato Lisboa Francês
Universidade Federal Rural da
Amazônia

Av. Pres. Tancredo Neves, Nº 2501
Cep 66.077-830, Belém-Pará-Brasil
leolopes671@gmail.com

RESUMO

O objetivo deste trabalho é analisar a associação entre trabalho infantil e gravidez na adolescência, tomando por referência os municípios da Amazônia Legal brasileira em relação aos demais municípios do país. Para isso, utiliza dados obtidos do Censo Demográfico, realizado no Brasil no ano de 2010, pelo Instituto Brasileiro de Geografia e Estatística. A visualização da distribuição espacial dos fenômenos sociais investigados é apresentada em forma de mapas e, de forma interativa, são aplicados algoritmos de mineração para extrair regras de associação entre as variáveis alvo do problema. Gravidez na adolescência e trabalho infantil não são variáveis dependentes. Entretanto, assim como a gravidez na adolescência é um problema para os municípios da região Norte, o mesmo cenário foi encontrado para o fenômeno do trabalho infantil, ou seja, em municípios da Amazônia onde o trabalho infantil é alto ou muito alto, a probabilidade de também encontrar um percentual muito elevado de adolescentes com filhos é de 76%.

Palavras-chave

Saúde Pública; Gravidez na Adolescência; Tecnologia da Informação; Teorema de Bayes.

ABSTRACT

This study analyzes the association between child labor and teenage pregnancy, taking as reference the municipalities of the Brazilian Legal Amazon in comparison to the other municipalities of the country. For this, we used data obtained from the Demographic Census, conducted in Brazil in 2010, by the Brazilian Institute of Geography and Statistics. The visualization

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SBSI 2017, June 5th–8th, 2017, Lavras, Minas Gerais, Brazil.
Copyright SBC 2017.

of the spatial distribution of the investigated social phenomena is presented in the form of maps and we apply mining algorithms, in an interactive way, to extract rules of association between the target variables of the problem. Teenage pregnancy and child labor are not dependent variables. However, just as teenage pregnancy is a problem for municipalities in the North, the same scenario was found for the phenomenon of child labor, ie, in municipalities in the Amazon where child labor is high or very high, there is a probability of 76% that this municipality will present a very high percentage of teenage mothers.

CCS Concepts

• Information systems → Information systems applications → Data mining.

Keywords

Public health; Teenage pregnancy; Information Technology; Bayes Theorem.

1. INTRODUÇÃO

Crianças e adolescentes representam mais de 30% da população brasileira, isso equivale a cerca de 60 milhões de pessoas na faixa de 0 a 18 anos, segundo o Censo Demográfico Brasileiro de 2010, realizado pelo Instituto Brasileiro de Geografia e Estatística. No mundo, são cerca de 2,2 bilhões, que representam 31% da população mundial, quase 90% dessa população vive em países em desenvolvimento, 15% são forçadas ao trabalho infantil, tendo seu direito a educação violado, e são responsáveis por cerca de 11% dos nascimentos em todo mundo [10].

O trabalho infantil — período da vida que inicia aos 5 anos e termina aos 17 anos completos, atinge cerca de 7,5% da população brasileira, o que são mais de 3 milhões de crianças e adolescentes, segundo a Pesquisa Nacional por Amostra de Domicílios (PNAD) [6]. Na região Norte, o percentual do trabalho infantil (8,2%) é maior que a média brasileira (7,5%).

A gravidez na adolescência, fase da vida que começa aos 10 anos e termina aos 19 anos completos é responsável por 19% de nascidos no Brasil, gerando mais de 500 mil partos no país. A

região Norte é a que concentra a maior média percentual de adolescentes que tiveram filhos (26%), acima da média brasileira (19%). São mais de 180 mil casos de gravidez na adolescência apenas nesta região, de acordo com o Sistema de Informação sobre Nascidos Vivos [7]

Trabalho infantil e gravidez na adolescência são, portanto, fenômenos sociais importantes, com ocorrência na região da Amazônia legal brasileira e alvo de interesse de pesquisadores e profissionais das áreas socioeconômica, saúde coletiva e políticas públicas. Para esses pesquisadores, a área de Mineração de Dados pode contribuir por apresentar associações e cenários de investigação que suscitam discussões e análises especializadas com respeito às causas e relações entre esses indicadores. Há um grande volume de dados e informações acumulado, em distintos formatos e estrutura, que podem auxiliar essas análises e é o analista de dados, ou engenheiro do conhecimento, o profissional responsável pela extração de informação e de conhecimento sobre essas bases de dados.

A mineração de dados é uma possível técnica utilizada para investigar a associação entre variáveis de um estudo. Com respeito à gravidez na adolescência e o trabalho infantil, há um interesse em estudar as associações com outras variáveis, tais como o local de referência e educação. Nessa direção, aplica-se a mineração de dados para encontrar associação entre essas variáveis. O estudo, nesse cenário, é um tipo de extração de conhecimento da base de dados a partir de um aprendizado preditivo, onde se analisam os dados que representam eventos passados buscando relações entre estes e que permitam prever situações em novos dados futuros. Nessa direção, este trabalho explora a associação entre os fenômenos da gravidez na adolescência e o trabalho infantil, buscando indicar possíveis relações de causa-efeito, considerando o espaço de referência, ou seja, municípios da Amazônia legal brasileira em relação aos demais municípios do Brasil.

Assim, este trabalho apresenta uma análise da associação entre a gravidez na adolescência e o trabalho infantil na Amazônia Legal Brasileira e variáveis representativas das dimensões de local de referência e educação de alguns municípios do Maranhão, Pará, Amazonas e Boa vista. A análise realizada segue na linha de estudos apresentados em Brito et al. e Silva et al. [1, 9], utilizando a mesma abordagem de redes bayesianas, mas utilizando indicadores do trabalho infantil, a fim de contribuir para as análises dos especialistas para os fatores associados ao problema da gravidez na adolescência.

Este artigo apresenta uma análise da associação entre a gravidez na adolescência e o trabalho infantil e variáveis representativas das dimensões de local de referência e educação nesses municípios. O artigo está estruturado da seguinte forma: a seção 2 apresenta o referencial teórico, a seção 3 apresenta a metodologia, incluindo as fontes de dados, a seção 4 apresenta os resultados do estudo e finalmente, a seção 5 apresenta as conclusões do trabalho

2. MINERAÇÃO DE DADOS

A mineração de dados é uma técnica muito utilizada para investigar a associação entre variáveis de um estudo. Segundo Silva Netto [8], o processo mineração de dados é um conjunto de atividades contínuas composto, basicamente, por cinco etapas: Seleção dos Dados, Pré-Processamento e Limpeza, Formatação, Mineração de Dados e Interpretação, conforme ilustrado na Figura 1. Este processo, aplicado a este estudo é explanado na seção 3.

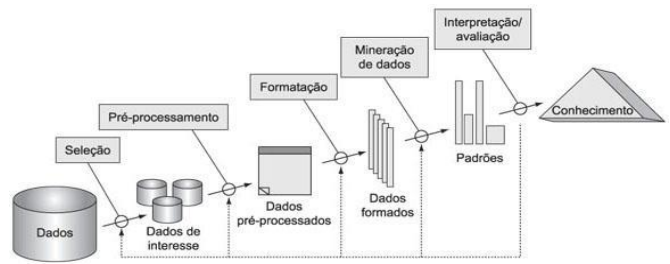


Figura 1. Processo de mineração de dados [4].

2.1 Redes Bayesianas

Redes bayesianas, também denominadas de redes causais ou modelos gráficos de dependência probabilística, podem ser vistas como modelos que codificam os relacionamentos probabilísticos entre variáveis que representam um determinado domínio. Segundo Chen [2], esses modelos são formados por uma estrutura qualitativa, representando as dependências entre os nós (variáveis) e uma estrutura quantitativa (tabelas de probabilidades condicionais dos nós), quantificando essas dependências em termos probabilísticos. Segundo Brito et al.[1], essas estruturas, oferecem uma representação eficiente da distribuição de probabilidades do conjunto de variáveis em estudo.

3. METODOLOGIA

Similar à abordagem utilizada nos estudos de Brito et al. [1] e Silva et al. [9], este é um estudo ecológico, baseado em dados obtidos do Censo Demográfico realizado pelo IBGE, no Brasil, em 2010 e informações constantes do portal Atlas Brasil (*Atlas do Desenvolvimento Humano no Brasil* - <http://www.atlasbrasil.org.br>) para o mesmo ano, afim de manter a consistência metodológica da pesquisa.

Assim, realizou-se o processo de integração das bases de dados a partir das planilhas extraídas dos Sistemas públicos do IBGE e Atlas Brasil. Para o presente estudo, as variáveis selecionadas envolvem fecundidade, renda, educação, trabalho infantil e espaço de referência (Brasil, Amazônia).

Nessa metodologia, são aplicados algoritmos de mineração para extrair regras de associação entre as variáveis alvo do problema, combinado com a visualização da distribuição espacial da gravidez na adolescência nos municípios brasileiros. Essa visualização é apresentada em forma de mapas e os percentuais médios de ocorrência dos fenômenos de gravidez na adolescência e trabalho infantil, nos municípios da Amazônia são apresentados para comparar as ocorrências no Brasil, de modo geral.

Na primeira etapa, de pré-processamento, as variáveis selecionadas foram local de referência — Amazônia ou outra região (variável: *Local*), educação (variável: *Idhm-Educação*), percentual de gravidez na adolescência (variável: *Gravidez_Na_Adolescência*) e trabalho infantil (variável: *Trabalho_infantil*). As variáveis contínuas passaram pelo processo de discretização, considerando intervalos de mesma frequência para distribuição dos valores de cada variável nas faixas de valores. Assim, as variáveis *Idhm-Educação*, *Gravidez_Na_Adolescência* e *Trabalho_infantil* apresentarem quatro faixas de valores, que representam aproximadamente 25% dos registros.

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

Nesta etapa, foi utilizado o software *bayesware discoverer* (<http://open.bu.edu/handle/2144/1288>), que aplica o algoritmo K2 [3]. Para efeitos de comparação, foi utilizado o software *weka* (<http://www.cs.waikato.ac.nz/ml/weka>), que aplica o algoritmo “*a priori*” na mineração das regras de associação. Os resultados são apresentados na seção 4. Por fim, foi utilizada a ferramenta QGIS (http://www.qgis.org/pt_PT/site) para demonstrar os mapas dos locais mais afetados pelo trabalho infantil, gravidez na adolescência e onde ambos os fenômenos ocorrem de forma simultânea.

4. RESULTADOS

Após a integração dos dados, foi utilizado o software *bayesware* que permitiu selecionar a rede bayesiana (Figura 2) que melhor representa a associação entre as variáveis do estudo.

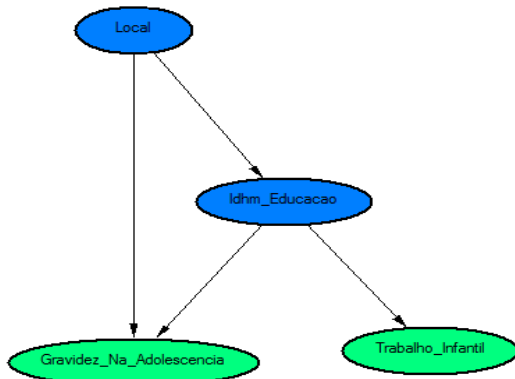


Figura 2. Rede bayesiana selecionada.

Na rede selecionada, *IDHM-Educação* é uma variável dependente de *Local* e *Gravidez_Na_Adolescência* é dependente de *Local* e *IDHM_Educação*, assim como *Trabalho_Infantil* é dependente de *IDHM-Educação*. As variáveis *Trabalho_Infantil* e *Gravidez_Na_Adolescência* são independentes. Assim, podemos inferir que de acordo com o local de referência e o índice de educação pode existir ou não gravidez na adolescência, assim como, em locais onde o *IDHM_Educação* é baixo, deve existir um alto percentual de gravidez na adolescência e trabalho infantil.

Utilizando o software *Bayesware Discoverer*, as variáveis contínuas foram discretizadas em faixas de valores, conforme apresentado na Figura 3.

Na sequência, após a aplicação do algoritmo K2 implementado no *Bayesware Discoverer* [3], também foi possível encontrar a distribuição de probabilidades para cada evidência do estudo. Por exemplo, dada a evidência de que o local de referência é a Amazônia, há uma probabilidade de 65,6% do percentual de Gravidez na Adolescência estar muito alto. Ainda, dada a evidência de que o *IDHM_Educação* esteja no nível muito baixo, há uma probabilidade de 40,2% do percentual de gravidez na adolescência e 38,4% do trabalho infantil estarem em níveis muito alto. Quando combinamos o local de Referência como sendo Amazônia e o *IDHM-Educação* no nível muito baixo, há uma probabilidade de 78% do percentual de gravidez na adolescência estar muito alto. Em outra análise, dada a evidência de que o *IDHM-Educação* esteja no nível muito baixo e que o percentual de gravidez na adolescência esteja muito alto, há uma probabilidade de 42,5% do local de referência ser a Amazônia.

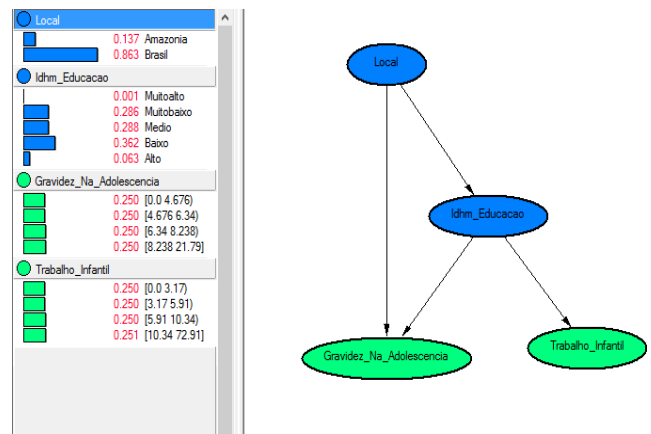


Figura 3. Rede bayesiana resultante do processo de mineração e faixas de valores para as variáveis

De acordo com a rede bayesiana, *Gravidez_Na_Adolescência* e o *Trabalho_Infantil* não são variáveis dependentes. Entretanto, assim como a gravidez na adolescência é um problema para os municípios da região Norte, o mesmo cenário foi encontrado para o fenômeno do trabalho infantil. Por exemplo, utilizando o software *weka*, foi encontrado que em municípios da Amazônia onde o trabalho infantil é alto ou muito alto, a probabilidade de também encontrar percentual de adolescentes com filhos muito alta é de 76% (Quadro 1). Esses resultados obtidos no *weka*, possuem suporte mínimo em 3% e confiança mínima de 60%.

Quadro 1. Regras de associação para *Trabalho_Infantil* alto ou muito alto e *Local = Amazônia*.

<i>Local=Amazônia, Trabalho_Infantil=Alto =></i> <i>Gravidez_Na_Adolescência=MuitoAlto</i>
<i>Local=Amazônia, Trabalho_Infantil=MuitoAlto =></i> <i>Gravidez_Na_Adolescência =MuitoAlto</i>

O trabalho infantil está presente em grande parte do território brasileiro (Figura 4), entretanto a região Amazônica chama a atenção pelos seus altos índices. Dos seus 756 municípios, 479 estão em situação de alto ou muito alto de uso do trabalho infantil. O município Amazônico com o maior índice (35,41) de trabalho infantil é Seringueira e está localizado no Estado de Rondônia. Contudo, o município brasileiro com maior índice (35,77) é Barra do Rio Azul (SC).

A gravidez na Adolescência é outro problema que se faz presente de forma muito séria na região Amazônica, pois se encontra com índices altos ou muito altos em praticamente toda região, e apenas 91 municípios não estão nessas faixas. Segundo o Sistema de Informações sobre Nascidos Vivos [7], somente na região norte (a que abrange maior área da Amazônia Legal Brasileira) foram 184.698 partos feitos. O município amazônico com maior índice (17,30%) de Gravidez na Adolescência é o Amajari, localizado no estado de Roraima. Todavia, o que possui o maior índice nacional (17,97%) é o município Antônio João situado no estado de Mato Grosso do Sul. Finalmente, é possível observar os municípios onde ambos, gravidez na adolescência e trabalho infantil, possuem índices muito altos (Figura 5).

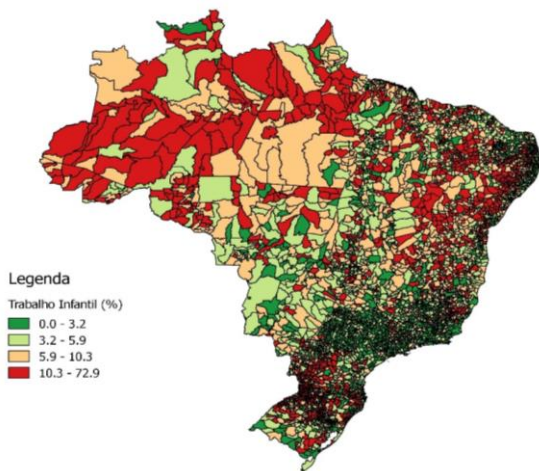


Figura 4. Mapa do trabalho infantil no Brasil.

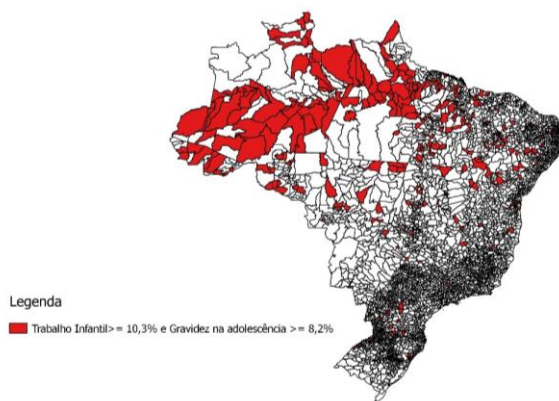


Figura 5. Municípios com alta ocorrência de trabalho infantil e gravidez na adolescência.

Dos 385 municípios que se encontram nessa situação, 181 pertencem à Amazônia Legal Brasileira. Dentre estes, existem municípios como Capixaba (Acre), Acará (Pará), Itapiranga (Amazonas), Nova Mamoré (Rondônia) e Alto Alegre (Roraima) que ficam próximos das capitais. Dos 181 municípios, 131 possuem o Índice de Desenvolvimento Humano Educacional muito baixo.

5. CONCLUSÕES

Este trabalho aplica a metodologia baseada em redes bayesianas para analisar a relação entre as variáveis consideradas no estudo. Com a utilização do software *Bayesware* como forma de gerar variáveis dependentes, este trabalho apresentou resultados satisfatórios mostrando a importância do local de referência e da educação para questões sociais como a gravidez na adolescência e o trabalho infantil.

As redes bayesianas obtidas, apontando as associações diretas e indiretas são importantes instrumentos para serem apresentados para especialistas das áreas de saúde pública, socioeconômica, dentre outros. Entretanto, a partir da identificação das associações, são necessárias análises mais profundas desses especialistas. Assim, este estudo, procura contribuir para as análises desses especialistas através da

identificação de associações que não são facilmente observadas a partir de grandes bases de dados.

Uma limitação deste estudo diz respeito ao fenômeno conhecido como “falácia ecológica”, explanado por Freedman [5], um fenômeno que consiste em pensar que a relação observada para grupos necessariamente se mantém para os indivíduos. E isso é um erro. Ou seja, uma associação encontrada no estudo de sua ocorrência coletiva, ou seja, para municípios, não pode ser reproduzida para nível individual.

Como trabalhos futuros pretende-se avançar no estudo da relação entre as variáveis estudadas procurando identificar as principais características de municípios com alta ocorrência de gravidez na adolescência e alta ocorrência de trabalho infantil.

6. REFERÊNCIAS

- [1] Brito, S. R., Silva, A. S., Cruz, A. G., Barroso, R. F. F., Monteiro, M. A., Costa, J. C. W. A., Francês, C. R. L. Gravidez na adolescência e o acesso às Tecnologias de Informação e Comunicação na Amazônia. *Mundo Amazônico*, 6, 2 (2015), p. 87. DOI= <https://doi.org/10.15446/ma.v6n2.53742>.
- [2] Chen Z. *Data mining and uncertain reasoning: an integrated approach*. John Wiley & Sons, New York, 2001.
- [3] Cooper, G. F. and Herskovits, E. A Bayesian method for the induction of probabilistic networks from data. *Mach Learning*, 9, 4 (1992), 309-47. DOI= [10.1007/BF00994110](https://doi.org/10.1007/BF00994110).
- [4] Fayyad, U. M., Piatetsky-Shapiro, G., and Smyth, P. From Data Mining to Knowledge Discovery: An Overview. *AI Magazine*, 17, 3 (1996), 37-54.
- [5] Freedman, D. A. *Ecological inference and the ecological fallacy*. Technical Report. University of California, 1999.
- [6] IBGE. Instituto Brasileiro de Geografia e Estatística. Pesquisa Nacional por Amostra de Domicílios. Retrieved December 09, 2016, from IBGE: <http://www.sidra.ibge.gov.br>.
- [7] IBGE. Instituto Brasileiro de Geografia e Estatística. Sistema de Informações de Nascidos Vivos – SINASC. Retrieved December 09, 2016, from IBGE: <http://ces.ibge.gov.br/base-de-dados/metadados/ministerio-da-saude/sistema-de-informacoes-de-nascidos-vivos-sinasc.html>.
- [8] Silva Neto, M. A., Villwock, R., Scheer, S., Steiner, M.T.A., Dyminski, A. S. Técnicas de mineração visual de dados aplicadas aos dados de instrumentação da barragem de Itaipu. *Gest. Prod.*, 17, 4 (Dec 2010), 721-734. DOI= [http://dx.doi.org/10.1590/S0104-530X2010000400007](https://doi.org/10.1590/S0104-530X2010000400007).
- [9] Silva, A. S., Brito, S. R., Barroso, R. F. F., Cruz, A. G., Monteiro, M. A., Costa, J. C. W. A., Francês, C. R. L. Gravidez na adolescência e associação com indicadores de renda, educação e acesso às Tecnologias de Informação e Comunicação no Brasil e na Amazônia Legal Brasileira. *Saúde e Desenvolvimento Humano*, 4, 2 (2016), 21-33. DOI= [http://dx.doi.org/10.18316/2317-8582.16.27](https://doi.org/10.18316/2317-8582.16.27).
- [10] UNICEF. United Nations Children's Fund. The State of the World's Children 2014 in numbers. Technical Report. Retrieved January, 05, 2017, from UNICEF: https://www.unicef.org/sowc2014/numbers/documents/english/SOWC2014_In%20Numbers_28%20Jan.pdf