

# Árvores de decisão para a predição do estado nutricional de crianças em Centros Municipais de Educação Infantil de Lavras, MG

Paula Ribeiro Santos <sup>1</sup>, Luiz Felipe de Paiva Lourenção <sup>2</sup>, Míriam Monteiro de Castro Graciano <sup>3</sup>, Stela Márcia Pereira <sup>4</sup>, Izabela Regina Cardoso de Oliveira <sup>5</sup>

## 1 Introdução

Os Centros Municipais de Educação Infantil (CMEIs) têm um papel fundamental no crescimento e desenvolvimento de crianças pertencentes aos estratos sociais, que são pessoas classificadas em níveis comuns, menos favorecidos socialmente e financeiramente, tornando-se uma estratégia dos países subdesenvolvidos. Um papel importante desses espaços é fornecer alimentação adequada para as crianças (Barbosa, 2007). De acordo com Simon (2003), esse processo da criança receber uma dieta adequada envolve complexos fatores sociais, econômicos e culturais que interferem no estado nutricional infantil sendo que a segurança alimentar e nutricional são consideradas como um direito humano, ou seja, um bem público, conforme a Lei Orgânica de Segurança Alimentar e Nutricional no Brasil, aprovada em 2006.

Desta forma, é indiscutível a importância de se estudar os fatores relacionados ao estado nutricional infantil, que pode ser avaliado por meio de aplicação de questionários sobre frequência e segurança alimentar. Fatores socioeconômicos e clínicos, gestação e alimentação da criança são comumente considerados em estudos desse tipo. Para avaliar seus efeitos, métodos estatísticos clássicos como regressão ordinária são utilizados (Antunes et al., 2010; Villa, 2015).

De acordo com Ath (2000), a regressão é amplamente utilizada no contexto de um estudo equilibrado bem projetado, sendo eficaz para determinar quais fatores afetam uma resposta numérica, ou seja, variável de interesse em um experimento que é uma medida ou observação. No entanto, à medida que o número de variáveis que afetam a resposta (variáveis explicativas) e a complexidade dos dados aumentam os modelos lineares se tornam menos eficazes.

Logo, técnicas estatísticas mais modernas vêm surgindo para melhorar a eficiência dos métodos estudados. Um exemplo disso são técnicas de machine learning, como as árvores de decisão (Ath, 2000; Liaw, 2002). As árvores modelam uma variável resposta, que pode ser categórica (árvores de classificação) ou numérica (árvores de regressão), em função de uma ou mais variáveis explicativas, que podem ser tanto categóricas quanto numéricas. É construída subdividindo repetidamente os dados, que tem embutida a partição de cada nó, em dois caminhos por variável, isto é, ela faz apenas divisões dicotômicas, particionando em dois grupos mutuamente exclusivos, de maneira que os grupos sejam o mais homogêneo possível. O procedimento de divisão é então aplicado a cada grupo separadamente, de forma que eles sejam homogêneos e que, além disso, a árvore se mantenha razoavelmente

<sup>1</sup>Mestranda em Estatística e Experimentação Agropecuária, UFPA. e-mail: *paullasant.s@hotmail.com*

<sup>2</sup>Mestrando em Ciências da Saúde, UFPA. e-mail: *luizfelipecapaiva03@gmail.com*

<sup>3</sup>Departamento de Ciências da Saúde, DSA/UFPA. e-mail: *miriam.graciano2@dsa.ufpa.br*

<sup>4</sup>Departamento de Ciências da Saúde, DSA/UFPA. e-mail: *stelapereira@dsa.ufpa.br*

<sup>5</sup>Departamento de Estatística, DES/UFPA. e-mail: *izabela.oliveira@ufpa.br*

pequena, mas que também pode ser cultivada, e depois removida para o tamanho desejado (Breiman et al., 1984).

Nesse trabalho utilizamos dados coletados em crianças de 6 meses a 5 anos que frequentam CMEIs de Lavras, MG. Diversas variáveis foram avaliadas para cada uma das 581 crianças participantes do estudo e o objetivo é identificar aquelas que estão mais associadas à frequência e insegurança alimentar. A proposta desse trabalho é obter árvores de classificação (insegurança alimentar) e regressão (frequência alimentar) para esses dados e entender como fatores socioeconômicos e da saúde da criança estão relacionados com esses desfechos.

## 2 Materiais e Métodos

### 2.1 Material

Os dados utilizados nesse trabalho foram obtidos em parceria com o Departamento de Ciências da Saúde (DSA) da Universidade Federal de Lavras (UFLA) e são resultantes de um estudo transversal realizado entre abril e novembro de 2018. Selecionou-se, por amostragem probabilística, 650 crianças pré-escolares matriculadas nos CMEIs. Explicado o objetivo da pesquisa e o seu caráter voluntário, foi aplicado o Termo de Consentimento Livre e Esclarecido (TCLE), após os devidos esclarecimentos sobre a importância do estudo.

A amostra final foi composta por 581 pré-escolares de seis meses a 5 anos de idade, matriculados em 15 CMEIs de Lavras-MG, classificadas e agrupadas, em tercís de nível socioeconômico, considerando: a) condições socioeconômicas do bairro para obter um número homogêneo de crianças; b) cálculo do estrato por faixas etárias similares em cada grupo; c) utilizada a mesma proporção da faixa etária por cada CMEI inserido na pesquisa. Os CMEIs foram alocados em três grupos, sendo o mesmo número de escolas de piores condições socioeconômico, de condições intermediárias de melhores condições, iguais em cada grupo de intervenção.

Aplicou-se um questionário composto por 47 variáveis explicativas, sendo elas numéricas, ordinais e categóricas, divididas em três grupos: fatores socioeconômicos e clínicos, como por exemplo, situação econômica da família, grau de instrução dos pais; sobre a gestação da mãe e condições de nascimento da criança (como peso, tamanho, apgar); e dados gerais de alimentação e saúde da criança, se estava em aleitamento materno, se fez algum uso de tipo de leite, quando foram introduzidos outros alimentos, número de refeições, dentre outras. Foram consideradas duas variáveis respostas, frequência e insegurança alimentar. Na primeira, cada criança recebeu um índice num intervalo -10 a 17 de acordo com a frequência que ela ingeria leite e derivados, óleos, gorduras, cereais, leguminosas, doces, frutas, verduras e legumes, em que maiores índices correspondem a uma frequência alimentar adequada. No caso da insegurança alimentar, cada criança era classificada em em segurança ou não de acordo com a escala brasileira de insegurança alimentar.

### 2.2 Métodos e programas estatísticos

Para a variável resposta numérica frequência alimentar foi criada uma árvore de regressão e para a variável resposta categórica insegurança alimentar foi criada uma árvore de classificação utilizando o pacote *rpart* (Therneau et al., 2018) do programa R (R CORE

TEAM, 2018). Para a obtenção dos gráficos, utilizamos o pacote *rpart.plot* (Milborrow et al., 2018).

Dessa forma, para construí-las, são considerados dois grupos, os de treinamento, que correspondiam a 75% dos dados e os de teste, referente aos 25% restantes. Considerando o grupo de treinamento a árvore é gerada da seguinte forma: a variável que está causando mais variabilidade nos dados é selecionada para o nó raiz, e como as árvores fazem apenas divisões dicotômicas, dois novos ramos são criados no sentido inferior da árvore. O processo de divisão é continuado resultando em árvores totalmente crescidas até que um critério de parada definido seja alcançado. Por fim, pelo grupo teste podemos avaliar a capacidade preditiva do modelo.

### 3 Resultados e Discussão

Na árvore criada para a frequência alimentar apresentada na Figura 1, a variável mais importante para discriminar os indivíduos, que é o nó raiz foi escolaridade materna. Logo se a mãe possui primário incompleto ou ginásial incompleto ou 2º grau incompleto e a água de beber na residência é fervida ou não é tratada a frequência alimentar da criança é avaliada em 1,2, ou seja, uma frequência alimentar muito ruim. Outras interpretações interessantes podem ser feitas como em relação à situação econômica da família, idade e profissão. Nessas partições à direita da árvore, temos os maiores valores preditos para a variável resposta, isto é, melhores condições de frequência alimentar.

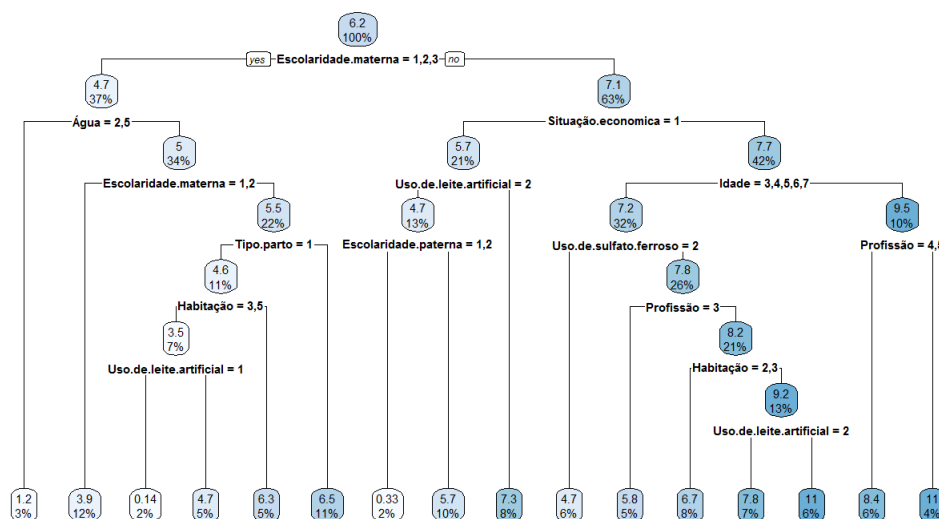


Figura 1: Árvore de regressão para a frequência alimentar.

Já para a árvore criada para a insegurança alimentar apresentada Figura 2, se a renda familiar é maior que um salário mínimo e se possuem um ou mais automóveis a criança está em segurança alimentar (0). Esses resultados eram esperados já que a escala de insegurança alimentar está muito relacionada a aspectos socioeconômicos da família. Por outro lado, se a renda é até um salário mínimo e a criança faz menos que 4 refeições no

```

graph TD
    Root((0  
0.43  
100%)) -- yes --> Node1((0.29  
56%))
    Root -- no --> Node2((0.61  
46%))
    Node2 --> Node3((0.55  
37%))
    Node2 --> Node4((0.60  
34%))
    Node3 --> Node5((0.57  
29%))
    Node3 --> Node6((0.54  
27%))
    Node4 --> Node7((0.62  
24%))
    Node5 --> Node8((0.57  
18%))
    Node6 --> Node9((0.59  
23%))
    Node7 --> Node10((0.64  
14%))
    Node8 --> Node11((0.44  
9%))
    Node8 --> Node12((0.64  
14%))
    Node9 --> Node13((0.25  
30%))
    Node9 --> Node14((0.25  
30%))
    Node10 --> Node15((0.29  
20%))
    Node10 --> Node16((0.71  
90%))
    Node11 --> Node17((0.23  
42%))
    Node11 --> Node18((0.30  
70%))
    Node11 --> Node19((0.71  
60%))
    Node12 --> Node20((0.35  
70%))
    Node13 --> Node21((0.23  
30%))
    Node13 --> Node22((0.25  
20%))
    Node14 --> Node23((0.60  
40%))
    Node15 --> Node24((0.29  
20%))
    Node15 --> Node25((0.71  
90%))
    Node16 --> Node26((0.87  
40%))
    Node16 --> Node27((0.89  
20%))
    Node16 --> Node28((1.00  
20%))
    Node16 --> Node29((0.87  
80%))
  
```

## 4 Considerações finais

## 5 Agradecimientos

## Referências Bibliográficas

4

domiciliar. *Cad. Saúde Pública*. Rio de Janeiro, 26(8):1642-1650, ago, 2010.

ATH, G. de; FABRICIUS, K.. Classification and Regression trees: a powerful yet simple technique for ecological data analysis. *Ecology*, 81(11), 2000, pp. 3178-3192.

BARBOSA, R. M. S.; SOARES, E. A.; LANZILLOTTI, H. S.. Avaliação da ingestão de nutrientes de crianças de uma creche filantrópica: aplicação do Consumo Dietético de Referência. *Rev. Bras. Saúde Matern. Infantil*, Recife, 7 (2): 159-166, abr. / jun., 2007.

Brasil. Lei nº. 11.346, de 15 de setembro de 2006. Cria o Sistema Nacional de Segurança Alimentar e Nutricional SISAN com vistas em assegurar o direito humano à alimentação adequada e dá outras providências. Diário Oficial da União 2006; 18 set.

BREIMAN, L.; FRIEDMAN, J. H.; OLSHEN, R. A.; STONE, C. J.. Classification and Regression Trees. Michigan: Wadsworth International Group, 1984.

LIAW, A.; WIENER, M.. Classification and Regression by randomForest. *R News*, Vol. 2/3, December 2002.

MILBORROW, Stephen. rpart.plot: Plot 'rpart' Models: An Enhanced Version of 'plot.rpart'. 2018. R package version 3.0.6. URL <https://CRAN.R-project.org/package=rpart.plot>.

SIMON, V. G. N.; SOUZA, J. M. P. de; SOUZA, S. B. de. Introdução de alimentos complementares e sua relação com variáveis demográficas e socioeconômicas, em crianças no primeiro ano de vida, nascidas em Hospital Universitário no município de São Paulo. *Rev. Bras. Epidemiol.*, vol. 6, nº 1, 2003.

VILLA, J. K. D.; SILVA, A. R. e; SANTOS, T. S. S.; RIBEIRO, A. Q.; P., M. C.; SANT'ANA, L. F. da R.. Padrões alimentares de crianças e determinantes socioeconômicos, comportamentais e maternos. *Revista Paulista de Pediatria*, 2015;33(3):302-309.

THERNEAU, T.; ATKINSON, B.. rpart: Recursive Partitioning and Regression Trees. 2018. R package version 4.1-13. URL <https://CRAN.R-project.org/package=rpart>

TOSCHKE, A. M.; BEYERLEIN, A.; KRIES, R.. Children at High Risk for Overweight: A Classification and Regression Trees Analysis Approach. *OBESEITY RESEARCH*, vol. 13, no. 7, July 2005.

R CORE TEAM. *R*: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2018. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.