

## Ciência de dados na avaliação de amostras de cafés de qualidade extrema

Eric Batista Ferreira<sup>1\*</sup>, Luiz Otávio de Oliveira Pala<sup>2</sup>, Rosemary Gualberto Fonseca Alvarenga Pereira<sup>3</sup>, Alberto Frank Lázaro Aguirre<sup>4</sup>, Júnio César Rosa<sup>5</sup>

### Resumo

A busca por melhorias da qualidade de cafés especiais tem impulsionado o mercado e influenciado produtores quanto aos valores comerciais das sacas e da respectiva demanda. No mercado brasileiro, o estado de Minas Gerais contribui com um significativo percentual de produtividade, fato que vem sendo verificado por concursos de qualidade, que desencadeiam estudos acerca de atributos relacionados a qualidade sensorial. Este trabalho analisou as vinte amostras finalistas de café tipo cereja descascado do Concurso Mineiro de Qualidade do Café no ano de 2013, objetivando apontar atributos associados ao ranqueamento das amostras que resultaram em melhor qualidade e construir um modelo de regressão que possa prever as notas finais. Sob o enfoque quantitativo, foi realizada uma análise exploratória a partir da construção dos componentes principais com os 15 atributos. Posteriormente, foi avaliada a capacidade de predição de notas finais através de um modelo de árvore de regressão. Como resultado, os ácidos Linoleico e Palmítico foram os atributos que mais contribuíram para a construção dos primeiros componentes principais, e além disso, o ácido Linoleico foi atribuído como a raiz da árvore de regressão, ou seja, um atributo de grande importância para predição das notas finais.

**Palavras-chave:** Atributo sensorial. Análise multivariada. Árvore de regressão. Café cereja descascado.

## Data Science in the evaluation of extreme quality coffee samples

### Abstract

Demand for the quality of specialty coffees has driven the market and influenced the increased commercial value of coffee bags. In the Brazilian market, the state of Minas Gerais contributes a significant percentage of productivity, a fact that has been accompanied by quality coffee contests. This paper analyzed the first twenty samples of peeled cherry coffee ranked in the *Concurso Mineiro de Qualidade do Café* in 2013. Under the quantitative approach, an exploratory analysis was performed from the construction of the principal components with the 15 attributes. Subsequently, the prediction capacity of final grades was evaluated through a regression tree model. As a result, Linoleic and Palmitic acids were the attributes that most contributed to the construction of the first principal components. In addition, Linoleic acid was attributed as the root of the regression tree, which is an important attribute for the prediction of final scores.

**Keywords:** Sensory attribute. Multivariate analysis. Regression tree. Peeled cherry coffee.

<sup>1</sup>Universidade Federal de Alfenas. Alfenas, MG. Brasil.

<https://orcid.org/0000-0003-3361-0908>

<sup>2</sup>Universidade Federal de Alfenas. Alfenas, MG. Brasil.

<https://orcid.org/0000-0002-9941-7951>

<sup>3</sup>Universidade Federal de Lavras. Lavras, MG. Brasil.

<https://orcid.org/0000-0003-2111-3744>

<sup>4</sup>Universidade Federal de Alfenas. Alfenas, MG. Brasil.

<https://orcid.org/0000-0002-5783-2276>

<sup>5</sup>Universidade Federal de Alfenas. Alfenas, MG. Brasil.

<https://orcid.org/0000-0001-8451-8920>

\*Autor para correspondência: [eric.ferreira@unifal-mg.edu.br](mailto:eric.ferreira@unifal-mg.edu.br)

## Introdução

O cenário de mercados cada vez mais competitivos começou a atingir a atividade cafeeira mundial nas últimas décadas. Essa situação pode ser evidenciada na constante busca por melhorias na qualidade do grão produzido de modo a atender a demanda atual. Em relação ao mercado brasileiro, notou-se uma redução quanto ao nível de exportações a partir de 1920. Tal redução pode ser justificada pelo aumento da qualidade dos cafés produzidos pela Colômbia e por países da América Central (Ormond et al., 1999). Em consequência dessas questões, pode-se notar a expansão de grãos especiais e cafeterias *gourmet*, incentivando produtores mundiais (Ormond et al., 1999; Boaventura et al., 2018).

No mercado Brasileiro, o estado de Minas Gerais é um dos maiores produtores do país. Atualmente, os produtores do estado estão ganhando espaço quanto ao nível de qualidade da safra produzida. Tal fato une-se a criação de concursos de qualidade que incorporam valor ao café (Barbosa et al., 2009) e objetivam o aumento de emprego e de qualidade de vida para o cafeicultor, fortalecendo também a agricultura familiar (Emater, 2019).

No que diz respeito a qualidade de um produto, entende-se como um grão de qualidade aquele com características de sabor e aroma notáveis pelo consumidor. Esses produtos comumente passam a ser certificados no mercado, expandindo a demanda dos cafés produzidos pela propriedade produtora (Oliveira et al., 2012).

No processo de especificação da qualidade do grão, relaciona-se parâmetros considerados tradicionais como o clima, solo de cultivo e altitude, o que pode ser visto em Souza (2006) e Silva et al. (2015). Além destes, incorpora-se atributos sensoriais que “geralmente se traduzem em qualidade superior da bebida”. Desta forma, cafés considerados especiais admitem uma diversidade de fatores em sua qualidade (Souza, 2006).

O aspecto sensorial é considerado o mais importante no processo de escolha do consumidor (Dutcosky, 2013). Estudos e aplicações da análise sensorial tem se mostrado como ferramenta útil para a avaliação de produtos (Minim, 2010). Este fato pode ser notado nos recentes estudos analisando a qualidade de produtos, como cultivares de café, em Sanchez e Chambers (2015) e Baqueta et al., (2019).

Na literatura, Chalfoun et al., (2013) analisaram os atributos sensoriais relacionados ao sabor, doçura, corpo e acidez em 21 cultivares do tipo cereja descascado. O estudo explorou as notas atribuídas a partir de metodologias da estatística experimental e multivariada, possibilitando o agrupamento de amostras com características semelhantes e a comparação dos cultivares avaliados.

A intensidade do sabor típico do café foi analisada por Kreuml et al., (2013) em amostras do tipo *Arábica* e *Robusta*. Como resultado, os componentes aromáticos foram importantes para a bebida do café e que mudanças sensoriais foram notadas com o tempo de armazenamento. Como exemplo deste fato, houve um aumento da intensidade de sabor azedo no café tipo *Arábica* e da acidez no *Robusta*.

Mori et al. (2018) descreveram a qualidade de cafés a partir de genótipos desenvolvidos para o estado do Espírito Santo, em que a componente de acidez foi uma das características de maior impacto na qualidade de um dos genótipos. Sanchez e Chambers (2015) discutiram impactos da forma de preparação de produtos como o café nas questões sensoriais, inferindo que o aroma e o sabor podem se alterar conforme o método de preparo; Baqueta et al., (2019) apontaram complexidades da avaliação sensorial por degustação e Borém et al. (2019) identificaram os principais atributos em cafés especiais de Minas Gerais.

Mais recentemente, métodos de aprendizado de máquina vem sendo utilizados na análise de qualidade de produtos, como o vinho e cafés, por meio e algoritmos de árvores, *boosting*, *bagging* e redes neurais (Goyal e Goyal, 2013; Lantz, 2015; Oliveira et al., 2019). Um método não paramétrico muito utilizado é a árvore de regressão, sendo construído a partir do particionamento no espaço dos dados (Loh, 2014). Cada particionamento é considerado um nó e cada desfecho final é denominado folha (Izbicki e Santos, 2018). Assim, com a criação de particionamentos distintos e disjuntos, a predição da variável resposta dado as covariáveis é dada por:

$$g(X) = \frac{1}{|\{i: X_i \in R_k\}|} \sum_{i: X_i \in R_k} y_i$$

Mais detalhes podem ser vistos em Loh (2014), James et al., (2013) e Izbicki e Santos (2018), bem como as etapas para a construção e poda de uma árvore, procedimento este que evita o super-ajuste.

Em conjunto a esses métodos de aprendizado, métodos multivariados como componentes principais também possibilitam analisar atributos de um produto, como visto em Chalfoun et al. (2013), Kreuml et al. (2013) e Kim et al. (2019). Conforme Jolliffe e Cadima (2016), a técnica possibilita o redimensionamento dos dados a partir da construção de variáveis ortogonais que maximizam a variância.

Conforme Mingoti (2005), para a construção dos componentes principais define-se o vetor aleatório  $\mathbf{X} = (X_1, \dots, X_p)$  com vetor de médias  $\mu$ , matriz de covariâncias  $\Sigma$ . Então, os componentes podem ser calculados via matriz de correlação ou covariâncias. Caso estes sejam

estimados pela matriz de covariâncias  $S$ , calcula-se os autovalores  $\hat{\lambda}_1, \dots, \hat{\lambda}_p$  e seus autovetores normalizados  $\hat{\epsilon}_1, \dots, \hat{\epsilon}_p$ . Desta forma, a  $i$ -ésima componente principal será dada por:

$$\hat{Y}_i = \hat{\epsilon}_i' X, \quad i = 1, \dots, p,$$

sob condições de que a variância de cada componente ( $\hat{Y}_i$ ) é correspondente a  $\hat{\lambda}_i$  e os componentes são ortogonais, ou seja, não correlacionados.

O nível de informação contido nos dados originais será retido nos componentes. Em geral, a literatura recomenda reter um número de componentes que expliquem em média 70% da variabilidade do conjunto original (Jolliffe e Cadima, 2016). Mais detalhes acerca da construção algébrica dos componentes principais estão em Johnson e Wichern (2015).

### Concursos de qualidade

As variações no processo de colheita da safra, como o tipo cereja descascado e natural, possibilitam estudos e concursos que explorem as particularidades de cada forma de colheita. Saraiva *et al.* (2010) obtiveram como resultado que os métodos por via úmida sem retirada da mucilagem apresentaram cafés de melhor qualidade e controle de processamento. Assim, os concursos de certificação do grão avaliam os cultivares em categorias conforme o método de colheita (Barbosa *et al.*, 2009).

No ano de 2013, a Empresa de Assistência Técnica e Extensão Rural do Estado de Minas Gerais (EMATER, 2019) coordenou o Concurso de Qualidades dos Cafés de Minas. No evento foram inscritas 1213 amostras de café tipo *Arábica*, dos quais 870 corresponderam ao tipo natural e 343 ao tipo cereja descascado. Destas, 198 foram selecionadas para a fase final do evento que ocorreu na cidade de Lavras-MG. As amostras premiadas foram direcionadas a leilão, ampliando e auxiliando o processo de venda das sacas (EMATER, 2013).

Desta forma, este trabalho analisou as vinte amostras finalistas de café tipo cereja descascado do Concurso Mineiro de Qualidade do Café no ano de 2013, com o objetivo de apontar atributos associados ao ranqueamento das amostras de melhor qualidade e construir um modelo de regressão que possa prever as notas finais. Sob o enfoque quantitativo, foi realizado uma análise exploratória com a construção dos componentes principais. Posteriormente, foi avaliado a capacidade de predição de notas finais com um modelo de árvore de regressão. Sendo possível, desta forma, discutir a importância destes atributos para o planejamento de melhora na qualidade do grão e de derivados.

### Material e métodos

Para este estudo foram analisadas as notas atribuídas às 20 amostras finalistas no Concurso Mineiro de Qualidade do Café, relativas ao tipo cereja descascado. As variáveis analisadas podem ser separadas em três grupos: **atributos sensoriais**: acidez, corpo, sabor e doçura da amostra; **ácidos graxos**: ácido palmítico, esteárico, oleico, linoleico, linolênico, araquídico e behênico; e **compostos bioativos**: cafeína, trigonelina, ácido clorogênico e sacarose.

A partir da avaliação destes 15 atributos, concedeu-se uma nota final a cada amostra por meio de uma média ponderada em escala de 0 a 100. Essa nota final possibilitou o ranqueamento das amostras, indicando aquelas que apresentaram uma melhor qualidade em relação as demais. Desta forma, as amostras ranqueadas foram identificadas conforme a colocação em escala crescente.

Visando analisar conjuntamente todos os atributos avaliados no concurso, utilizou-se de metodologias da análise multivariada, como componentes principais. Os componentes principais foram calculados com o objetivo do redimensionamento dos dados, possibilitando a escolha de variáveis latentes que contemplem o maior percentual da variabilidade. Assim, adotou-se um percentual mínimo de 70% de explicação da variabilidade total para a retenção dos componentes.

Analisaram-se os resultados fornecidos pelos componentes principais, relacionando estes com os atributos originais avaliados. Essa forma possibilitou a análise de quais atributos mais contribuíram para cada componente principal, permitindo a discussão de quais atributos estão mais relacionados com a qualidade do grão. Por fim, foi utilizado um modelo de aprendizado de máquina de modo a prever a nota final da amostra com base nas variáveis que mais contribuíram para a formação dos primeiros componentes principais. Ressalta-se que para o estudo foram utilizadas as bibliotecas disponíveis no software R (R Core Team, 2019).

### Resultados e discussão

As vinte amostras selecionadas para a fase final do concurso apresentaram notas finais similares, admitindo um coeficiente de variação de 1,252% em relação à média de 84,28 pontos. Este fato pode expressar certo nível de concordância entre as notas atribuídas pelos avaliadores, pelo fato de que estas amostras já eram finalistas e homogêneas. Conforme o diagrama de Boxplot, apresentado na Figura 1, pode-se observar a presença de uma nota extrema. Esta nota mínima corresponde a amostra de café 20, ranqueada na vigésima posição com nota final equivalente a 81,42.

Conforme a nota final recebida pela amostra 20, é necessário analisar quais atributos fizeram com que

esta fosse penalizada em relação a amostra 1, ou seja, identificar se a presença ou ausência de algum atributo reduziu a qualidade da amostra. Na figura 2, tem-se no eixo abcissas os atributos avaliados e no eixo das ordenadas as respectivas notas fornecidas. Observe que a amostra 1 obteve maiores notas de ácido palmítico, linoleico e sacarose quando comparado a amostra 20.

Resultado que vai de encontro aos estudos de [Figueiredo et al. \(2015\)](#) e [Borém et al. \(2016\)](#), em que ácidos graxos saturados como o palmítico contribuíram para a qualidade sensorial, e amostras com maiores notas também apresentaram maiores níveis de sacarose, contribuindo para a formação do sabor.

Figura 1 – Diagrama de Boxplot relativo as notas finais das amostras do tipo cereja descascado

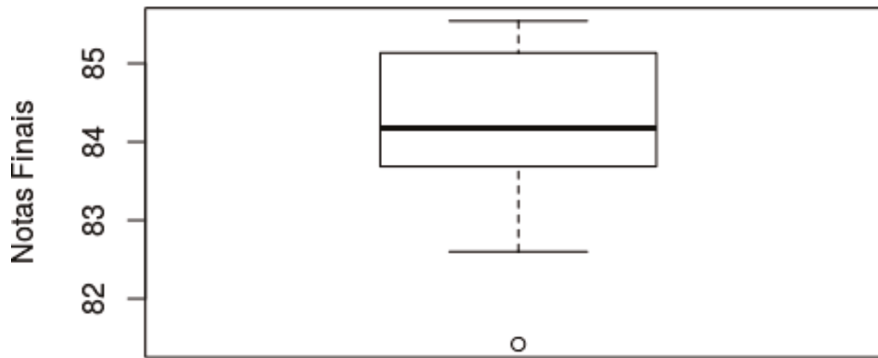
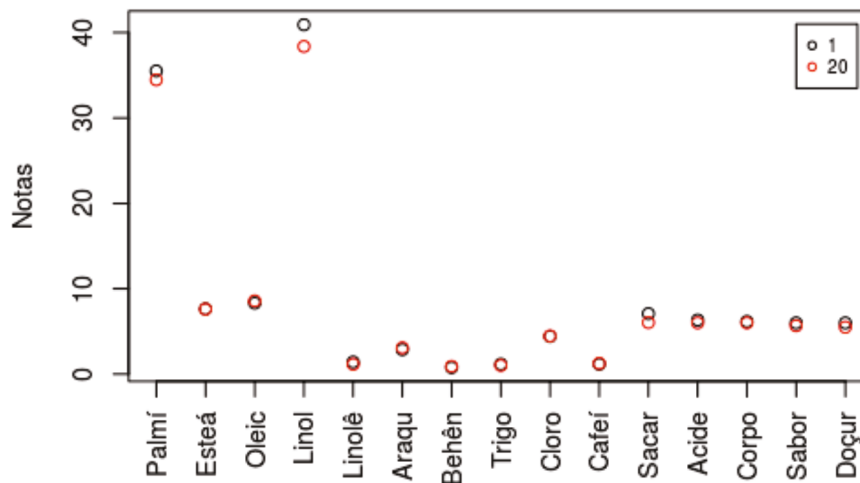


Figura 2 – Notas atribuídas às amostras 1 e 20, ou seja, a primeira e vigésima amostra ranqueada no concurso



Estendendo a análise ao contexto multivariado, os componentes principais foram construídos utilizando a matriz de covariâncias dos dados originais. Como resultado, 15 componentes foram obtidos, sendo os dois primeiros responsáveis por explicar 70,60% da variabilidade original dos dados. Conforme esse percentual de explicação, este trabalho concentrará as análises a partir dos dois primeiros componentes.

A Figura 3 apresenta as contribuições dos atributos para a construção dos dois primeiros componentes principais. Nota-se que os ácidos graxos linoleico e palmítico foram os atributos que mais contribuíram para a construção dos dois primeiros componentes.

A partir da Figura 4 pode-se notar que o ácido linoleico contribuiu com um percentual superior a 75%

para a criação do primeiro componente principal. Já para o segundo componente, o ácido palmítico foi o maior contribuidor, em um nível superior a 60%. Desta forma, pode-se afirmar que estes dois ácidos graxos foram os atributos mais significativos. Ressalta-se que a linha tracejada em vermelho representa a média esperada de contribuição de cada atributo para a criação do componente.

Entretanto, é possível analisar o relacionamento entre os atributos e as amostras. Conforme [Kassambara \(2017\)](#), no gráfico de relacionamento entre os indivíduos e os componentes, é necessário analisar a direção das variáveis sem considerar as posições absolutas. Assim, uma amostra localizada na mesma direção de uma dada variável, obteve a maior nota para tal variável.

Figura 3 – Contribuições das variáveis para a construção dos dois primeiros componentes principais

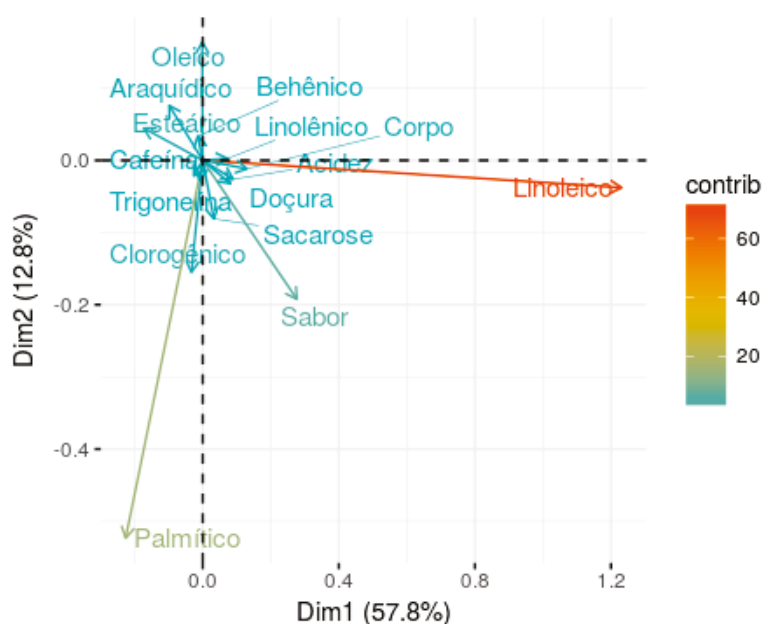
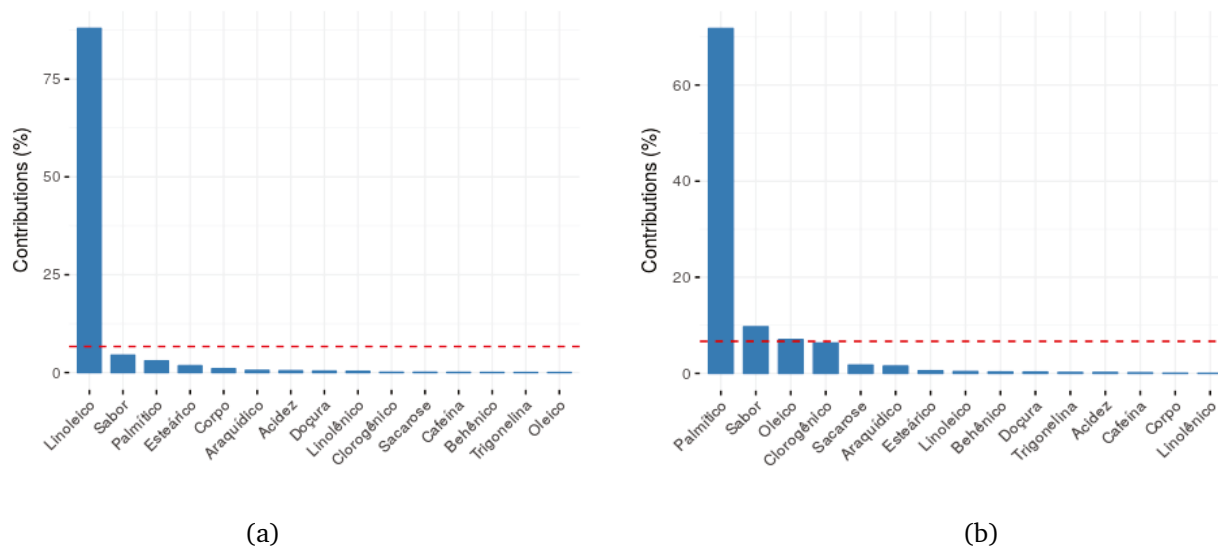


Figura 4 – Percentuais de contribuição de cada variável para a construção do primeiro e segundo componente em (a) e (b), respectivamente



Conforme a Figura 5, a amostra 1 manteve-se direcionada para os atributos sabor e sacarose, recebendo maiores notas para estes atributos. Na literatura, Alvarenga (2017) concluiu que a nota final de cafés do tipo cereja descascado está diretamente relacionada com atributos sensoriais como o Sabor. Por outro lado, a amostra 20 foi alocada no sentido inverso aos atributos de sabor, doçura, acidez, recebendo maiores notas para os atributos ácido esteárico, araquídico e cafeína.

Dada a homogeneidade entre as classificações das amostras, um padrão que pode ser analisado é que entre as amostras 1 a 5, 80% estiveram inversamente direcionadas aos ácidos araquídico, esteárico e behênico

e direcionadas no sentido do ácido linoleico. A presença deste último ácido também foi discutida por Alvarenga (2017), apontando que as melhores amostras foram avaliadas com elevados teores deste.

Tendo em vista que os ácidos linoleico e palmítico foram importantes para a construção dos componentes principais e discutidos nos estudos de [Figueiredo et al. \(2015\)](#) e [Borém et al. \(2016\)](#), foi considerado o método de árvore de regressão para previsão das notas finais das amostras. Considerando essas variáveis, tem-se a matriz de correlações apresentada na Tabela 1. Note que a correlação entre as variáveis independentes (palmítico e linoleico) pode ser considerada fraca.



Figura 5 – Relacionamento entre as amostras de café analisadas com os dois componentes principais

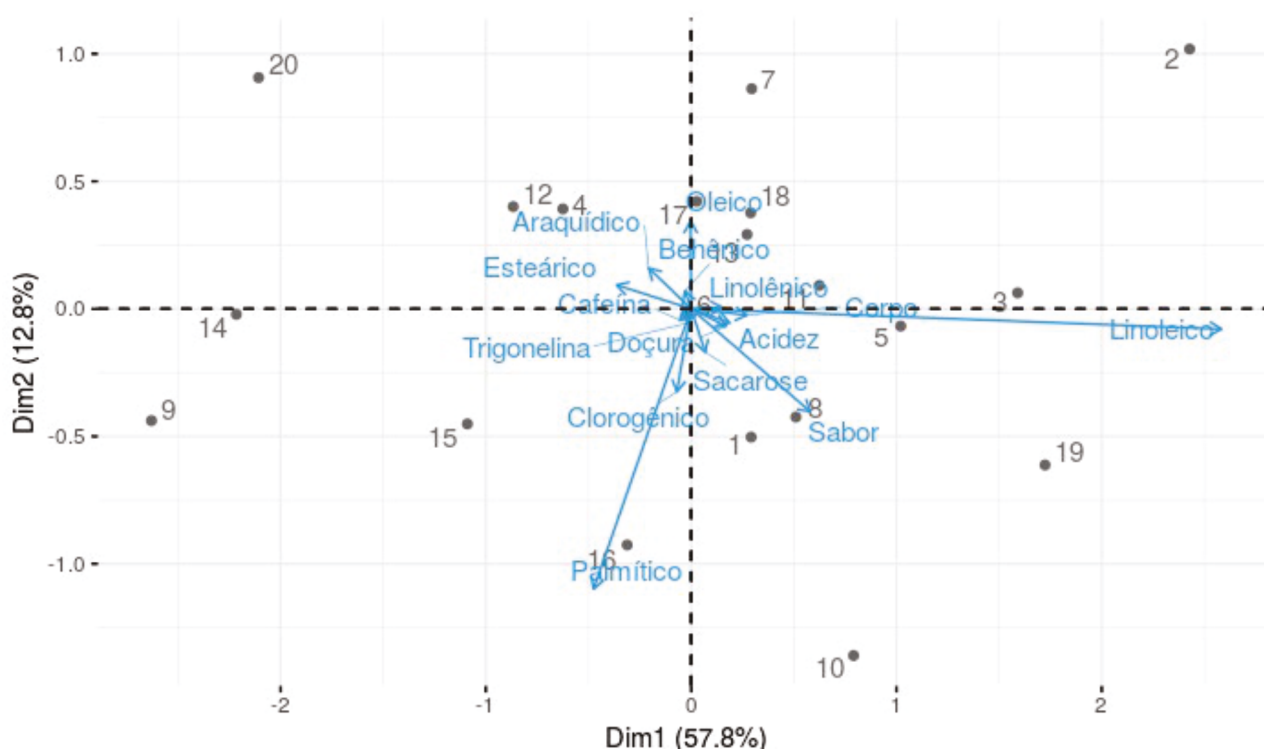


Tabela 1 – Correlações entre as variáveis utilizadas no modelo de regressão

	Notas	Palmítico	Limoneico
Notas	1,00	-0,09	0,35
Palmítico	-0,09	1,00	-0,34
Limoneico	0,35	-0,34	1,00

Para o ajuste da árvore de regressão foram consideradas aleatoriamente 90% das amostras do conjunto original, isto é, 90% para treino e 10% para teste. As amostras selecionadas para teste corresponderam a

amostra 8 e 14. Após o ajuste, foram obtidas as seguintes métricas de qualidade, apresentadas na Tabela 2. Note que o MAPE (erro percentual médio absoluto) indica uma boa acurácia.

Tabela 2 – Métricas de qualidade de ajuste para o modelo\*

Modelo	MAE	RMSE	RAE	MAPE
Árvore	0,8061	0,9541	1,7810	0,0095

\*MAE: Média absoluta de erro; RMSE: Raiz do erro quadrático médio; RAE: Erro relativo absoluto; MAPE: Erro percentual absoluto médio.

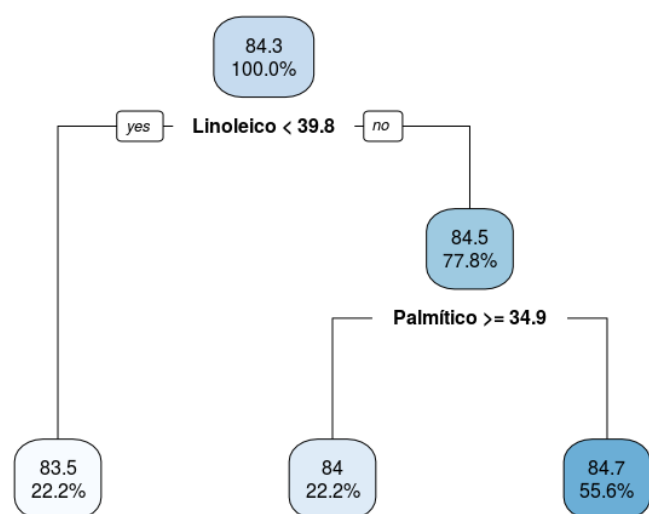
Dado o modelo considerado, tem-se que a covariável linoleico é a de maior importância no modelo (2,91%), seguido por palmítico (1,29%). Conforme o modelo, tem-se o seguinte resultado apresentado na Figura 6. Note que há uma nota média comum entre as amostras de 84,3, que é amplificada, em média, se a nota atribuída ao ácido linoleico for superior a 39,8.

Caso a nota atribuída ao ácido linoleico seja superior a 39,8, há um novo ramo que desagrega as notas em relação ao ácido palmítico. Neste caso, a maior nota final esperada é relacionada a amostra de linoleico superior a 39,8 e palmítico inferior a 34,9. No entanto, notas de

linoleico superiores a 39,8 e de palmítico inferiores a 34,9 conduziram a nota final de 84 em 22,2% das amostras avaliadas.

Note ainda que se fosse considerado apenas o ácido linoleico, uma nota atribuída a este superior a 39,8 já resultaria em uma média final de 84,5. Resultado similar ao de Alvarenga (2017) em que conforme o método de regressão por meio de componentes principais, altos teores de linoleico estiveram relacionados com melhores cafés, mas oposto ao discutido em [Figueiredo et al., \(2015\)](#), em que ácidos como linoleico relacionaram-se com cafés de menor acidez, sabor e corpo.

Figura 6 – Modelo de árvore de regressão ajustado as notas finais



Este fato indica que concentrações de ácido linoleico no grão podem ser utilizados para prever a nota final atribuída a amostra de café. De modo a avaliar a capacidade de previsão das notas finais das amostras, foi considerado os 10% restantes da amostra para treino, os resultados obtidos estão apresentados na Tabela 3, com erro percentual absoluto médio de 0,3811, indicando boa acurácia. Ainda conforme a Tabela 3, observe que o modelo conseguiu manter a ordenação das amostras, ou seja, a amostra 8 apresenta qualidade superior em relação a amostra 14.

Este resultado indica a possibilidade de utilização de métodos de aprendizado de máquina à contextos da estatística sensorial, permitindo a predição de notas de produtos conforme seus atributos, principalmente em casos de alta dimensão, onde um modelo de regressão usual pode apresentar problemas.

Tabela 3 – Valores observados e preditos para os 10% restantes da amostra, ou seja, amostras 8 e 14

Amostra	8	14
Nota Observada	84,92	83,90
Nota Predita	84,67	83,51

### Considerações finais

A identificação dos atributos que possibilitaram a escolha de uma amostra vencedora pode não ser trivial. Fato que pode ser justificado pela homogeneidade das notas e das características das amostras utilizadas neste estudo. De fato, todas apresentam boa qualidade.

Entretanto, é possível discutir acerca de atributos associados com a melhor qualidade do produto, mais especificamente, do café. Neste caso, a partir da análise exploratória via componentes principais, foi notado que os ácidos Linoleico e Palmítico foram atributos de grande importância para a construção dos primeiros componentes principais. Estes atributos foram usados como covariáveis em um modelo de regressão em árvore, de modo a prever as notas finais.

O modelo construído apresentou um baixo erro de predição para as notas finais das amostras, indicando que maiores concentrações de ácido linoleico possibilitaram maiores notas finais. Este atributo foi considerado como raiz da árvore pelo algoritmo, isto é, uma covariável de alta importância. Por fim, outros algoritmos podem ser utilizados de modo a avaliar a capacidade de predição de notas e atributos sensoriais em novos trabalhos, como por exemplo, considerando s árvores de regressão.

### Agradecimentos

O trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

### Referências

- Alvarenga, S. T. 2017. Caracterização química e sensorial de cafés especiais do sul de Minas Gerais. 120 f. Tese (Doutorado em Ciência dos Alimentos) - Universidade Federal de Lavras, Lavras.
- Barbosa, J.; Borém, F.; Alves, H.; Volpato, M.; Souza, V.; Santos, W. ; Andrade, L. 2009. Distribuição espacial de cafés do Estado de Minas Gerais e sua relação com a qualidade. In: Simpósio de Pesquisa dos Cafés do Brasil, 6, 2009, Vitória. Anais. Vitória. p. 1-7.
- Baqueta, M; Coqueiro, A; Valderrama, P. 2019. Brazilian coffee blends: a simple and fast method by near-infrared spectroscopy for the determination of the sensory attributes elicited in professional coffee cupping. Journal of food science, 84:1247-1255. Doi: <http://dx.doi.org/10.1111/1750-3841.14617>.

- Boaventura, P; Abdalla, C; Araújo, C; Arakelian, J. 2018. Cocriação de valor na cadeia do café especial: o movimento da terceira onda do café. Revista de Administração de Empresas, São Paulo, 58: 254-266. Doi: <http://dx.doi.org/10.1590/s0034-759020180306>.
- Borém, F; Cirillo, M; Alves, A; Santos, C; Liska, G; Ramos, M; Lima, R. 2019. Coffee sensory quality study based on spatial distribution in the Mantiqueira mountain region of Brazil. Journal of Sensory Studies, e12552. Doi: <https://doi.org/10.1111/joss.12552>.
- Borém, F; Figueiredo, L; Ribeiro, F; Taveira, J; Giomo, G; Salva, T. 2016. The relationship between organic acids, sucrose and the quality of specialty coffees. African Journal of Agricultural Research, 709-717 709-717. Doi: <http://dx.doi.org/10.5897/AJAR2015.10569>.

- Chalfoun, S.; Pereira, M.; Carvalho, G.; Pereira, A.; Savian, T.; Botelho, D. 2013. Sensorial characteristics of coffee (*coffea arabica* L.) varieties in the alto paranaíba region. *Coffe Science*, 8: 43-52. Doi: <http://dx.doi.org/10.25186/cs.v8i1.330>.
- Dutcosky, S. D. 2013. *Análise sensorial de alimentos*. 4. ed. Curitiba: Editora Universitária Champagnat, 2013.
- EMATER, 2019. Vencedores do 10º Concurso de Qualidade dos Cafés de Minas são premiados na UFLA. Portal do Agronegócio, 17 dez. 2013. Disponível em: <https://www.portaldogronegocio.com.br/noticia/vencedores-do-10-concurso-de-qualidade-dos-cafes-de-minas-sao-premiados-na-ufla-102238>.
- EMATER, 2013. XVI Concurso de Qualidade dos Cafés de Minas Gerais–2019. Disponível em: [http://www.emater.mg.gov.br/doc/intranet/upload/cafe2018/regulamento\\_2019.pdf](http://www.emater.mg.gov.br/doc/intranet/upload/cafe2018/regulamento_2019.pdf).
- Figueiredo, L; Borém, F; Ribeiro, F; Giomo, G; Taveira, J; Malta, M. 2015. Fatty acid profiles and parameters of quality of specialty coffees produced in different Brazilian regions. *African Journal of Agricultural Research*, 10: 3484-3493. Doi: [10.5897/AJAR2015.9697](https://doi.org/10.5897/AJAR2015.9697).
- Goyal, S; Goyal, G. 2013. Machine learning ann models for predicting sensory quality of roasted coffee flavoured sterilized drink. 2013. *Advances in Distributed Computing and Artificial Intelligence Journal*, 2: 9-13. Doi: <http://dx.doi.org/10.14201/ADCAIJ201426913>.
- Izbicki, R; Santos, T. 2018. Machine learning sob a ótica estatística: uma abordagem preditivista para estatística com exemplos em R. Disponível em: <http://www.rizbicki.ufscar.br/sml.pdf>.
- James, G; Witten, D; Hastie, T; Tibshirani, R. 2013. *An Introduction to Statistical Learning*. Reino Unido: Springer.
- Johnson, R; Wichern, D. 2015. *Applied Multivariate Statistical Analysis*. 6. ed. Pearson.
- Jolliffe, I; Cadima, J. 2016. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374 (2065). Doi: <http://dx.doi.org/10.1098/rsta.2015.0202>.
- Kassambara, A. 2017. Principal Component Methods. In: *R: practical guide*. Disponível em: <https://www.datanovia.com/en/wp-content/uploads/dn-tutorials/book-preview/principal-component-methods-in-r-preview.pdf>.
- Kim, H; Hong, D; Yu, J; Lee, S; Lee, Y. 2019. Identification of Headspace Volatile Compounds of Blended Coffee and Application to Principal Component Analysis. *Preventive Nutrition and Food Science*, 24, 217-223. Doi: [10.3746/pnf.2019.24.2.217](https://doi.org/10.3746/pnf.2019.24.2.217).
- Kreuml, M.; Majchrzak, D.; Ploederl, B.; Koenig, J. 2013. Changes in sensory quality characteristics of coffee during storage. *Food Science & Nutrition*, 4:267–272. Doi: [10.1002/fsn3.35](https://doi.org/10.1002/fsn3.35).
- Lantz, B. 2015 *Machine Learning with R*. 2. ed. Reino Unido: Packt Publishing.
- Loh, W. 2014. Classification and regression tree methods. *Wiley StatsRef: Statistics Reference Online*.
- Mingoti, S. 2005. *Análise de dados através de métodos de estatística multivariada*. Belo Horizonte: Editora UFMG.
- Minim, V. *Análise sensorial: estudo com consumidores*. 2. ed. Viçosa: Editora da Universidade Federal de Viçosa, 2010.
- Mori, A; Garcia, A; Ferrão, M; Fonseca, A; Ferrão, R; Benassi, M. 2018. Sensory profile of conilon coffee brews from the state of Espírito Santo, Brazil. *Pesquisa Agropecuária Brasileira*, 53: 1061-1069. Doi: <http://dx.doi.org/10.1590/s0100-204x2018000900010>.
- Oliveira, J; Elias, T; Lessa, M. 2012. Café Especial: Agregação de Valor ao Tradicional Café. *Revista Eletrônica de Comunicação*, 3: 1-8. Disponível em: <http://periodicos.unifacel.com.br/index.php/rec/article/view/445>.
- Oliveira, L; Menezes, F; Cirillo, M; Saúde, A; Borém, F; Liska, G. 2019. Machine Learning techniques in multiclass problems with application in sensorial analysis. *Concurrency and Computation: Practice and Experience*. Doi: [10.1002/cpe.5579](https://doi.org/10.1002/cpe.5579).
- Ormond, J; Paula, S; Farevet Filho, P. 1999. Café: (re)conquista dos mercados. Disponível em: [https://www.bndes.gov.br/SiteBNDES/export/sites/default/bndes\\_pt/Galerias/Arquivos/conhecimento/bnset/set1001.pdf](https://www.bndes.gov.br/SiteBNDES/export/sites/default/bndes_pt/Galerias/Arquivos/conhecimento/bnset/set1001.pdf).
- R CORE TEAM. 2019. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Sanchez, K.; Chambers, E. 2015. How Does Product Preparation Affect Sensory Properties? An Example with Coffee. *Journal of Sensory Studies*, 30. Doi: [10.1111/joss.12184](https://doi.org/10.1111/joss.12184).
- Saraiva, S; Araújo, M; Lucia, S, Silva, L; Zeferino, L. 2010. Comparação dos tipos de processamento pós-colheita de café. In: *Simpósio de Pesquisa dos Cafés Do Brasil*. Vitória. Anais. Brasília, DF: Embrapa.
- Silva, S; Queiroz, D; Ferreira, W; Corrêa, P; Rufino, J. 2015. Mapping the potential beverage quality of coffee produced in the Zona da Mata, Minas Gerais, Brazil. *Journal of the science of food and agriculture*, 96:3098-30108. Doi: [10.1002/jsfa.7485](https://doi.org/10.1002/jsfa.7485).
- Souza, M. 2006. Cafés sustentáveis e denominação de origem: a certificação de qualidade na diferenciação de cafés orgânicos, sombreados e solidários. Tese (Doutorado em Ciência Ambiental) - Ciência Ambiental, Universidade de São Paulo, São Paulo.