



Modelagem em Extração de Cafeína de sementes de *Paullinia cupana* (guaraná)

Everton da Costa¹, Robson Marcelo Rossi², Vanderly Janeiro³, Patricia Stüp⁴, Jean Carlos Cardoso⁵, João Pedro Serenini⁶.

¹ Universidade Estadual de Maringá, Discente do PBE, Departamento de Estatística. Maringá, PR, Brasil. E-mail: everto.cost@gmail.com

² Universidade Estadual de Maringá, Docente do PBE, Departamento de Estatística. Maringá, PR, Brasil. E-mail: rmrossi@uem.br

³ Universidade Estadual de Maringá, Docente do PBE, Departamento de Estatística. Maringá, PR, Brasil. E-mail: vjaneiro@uem.br

⁴ Universidade Estadual de Maringá, Discente do PBE, Departamento de Estatística. Maringá, PR, Brasil. E-mail: patriciastulp2@gmail.com

⁵ Universidade Estadual de Maringá, Discente do PBE, Departamento de Estatística. Maringá, PR, Brasil. E-mail: jeancarlos.card@gmail.com

⁶ Universidade Estadual de Maringá, Discente do PBE, Departamento de Estatística. Maringá, PR, Brasil. E-mail: jp_serenini@hotmail.com

RESUMO

Visando estabelecer uma relação entre a quantidade de cafeína extraída das sementes de *Paullinia cupana* e as variáveis regressoras (independentes), um modelo de regressão linear múltipla se mostrou apropriado. Em um experimento, foram observados o percentual de cafeína extraído e as variáveis tempo (20, 40 e 60 min), temperatura (40, 50 e 60°C), pressão (100, 200 e 300 bar) e modificador (etanol, metanol e etanol com metanol). Utilizando uma modelagem Bayesiana, foram consideradas distribuições normais a priori não informativas para o vetor de parâmetros e utilizadas estimativas da média a posteriori da distribuição marginal condicional. Fazendo uso do método de seleção de variáveis *Stepwise* com enfoque bayesiano comparamos os valores do DIC (*Deviance Information Criterion*) para uma maior parcimônia do modelo, foi selecionando aquele que possui o menor DIC. Observamos que a cada mudança de nível do modificador a quantidade de cafeína sofre um decréscimo de 1,75%, e a cada mudança de nível da temperatura temos um acréscimo de 2,59% de cafeína.

Palavras chave: Análise Bayesiana; Modelo de Regressão Linear Múltipla; *Stepwise*.

1 INTRODUÇÃO

O guaraná (*Paullinia cupana* Kunth, Sapindaceae) é uma planta originária da região amazônica brasileira. São várias as propriedades estimulantes da semente torrada do guaraná tais como aumentar a resistência do organismo e diminuir a fadiga.[1]

O Brasil é o maior produtor mundial de guaraná. Segundo [2] a produção brasileira tem praticamente sua totalidade consumida no mercado interno. Estimando-se que pelo menos 70% da produção seja absorvida pelos fabricantes de refrigerantes, enquanto o restante é comercializado na forma de xarope, pó, extrato e subprodutos.

Dentre os principais constituintes químicos do guaraná esta a cafeína, sendo muito utilizado na indústria alimentícia na forma de xaropes, extratos e destilados, empregados principalmente como agentes flavorizantes e como uma fonte de cafeína pela indústria de refrigerantes [1] *apud* [3], [4]

Este estudo tem como objetivo verificar quais são as variáveis significativas na extração de cafeína. Para isso foram analisadas as variáveis tempo, temperatura, pressão e modificador que nesse caso é composto de três níveis, etanol, metanol e uma mistura de etanol com metanol. Para cumprir tal objetivo foram utilizados os dados coletados no trabalho desenvolvido por [1].

2 METODOLOGIA

Como relatado no estudo [1], as sementes torradas de *P. cupana* foram obtidas em Alta Floresta, estado do Mato Grosso, Brasil. Uma exsicata (HUEM N. 9065) foi depositada no Herbário, Departamento de Biologia, Universidade Estadual de Maringá.

As partículas foram medidas e o cálculo do diâmetro foi realizado. O teor de umidade foi determinado por gravimetria usando uma balança analítica (Ohaus, MB35) equipada com um sistema de secagem por infravermelho. Maiores detalhes sobre o procedimento de extração e controle de qualidade podem ser encontrados em [1].

Neste estudo três modificadores foram adotados etanol, metanol e a mistura de etanol com metanol, cada modificador com percentual utilizado de 40%. Para a variável tempo de extração foi considerado 20, 40 e 60 minutos, temperatura 40, 50 e 60°C e pressão 100, 200 e 300 bar. Todos os experimentos foram realizados em triplicata e estão apresentados na Tabela 1. A variável modificador, tempo, temperatura e pressão foram consideradas como fatores.

Tabela 1: Informações amostrais

	Cafeína	Modificador	Tempo (min)	Temperatura (°C)	Pressão (bar)
1	50,94	Etanol	20	40	100
2	56,08	Etanol	20	40	100
3	53,51	Etanol	20	40	100
4	52,41	Etanol	40	50	200
5	51,52	Etanol	40	50	200
6	51,97	Etanol	40	50	200
7	58,60	Etanol	60	60	300
8	54,33	Etanol	60	60	300
9	56,46	Etanol	60	60	300
10	49,64	Metanol	20	50	300
11	50,44	Metanol	20	50	300
12	50,04	Metanol	20	50	300
13	49,87	Metanol	40	60	100
14	48,96	Metanol	40	60	100
15	49,41	Metanol	40	60	100
16	47,32	Metanol	60	40	200
17	45,45	Metanol	60	40	200
18	46,38	Metanol	60	40	200
19	59,86	Etanol/Metanol	20	60	200
20	61,89	Etanol/Metanol	20	60	200
21	60,87	Etanol/Metanol	20	60	200
22	53,43	Etanol/Metanol	40	40	300
23	49,20	Etanol/Metanol	40	40	300
24	51,32	Etanol/Metanol	40	40	300
25	50,70	Etanol/Metanol	60	50	100
26	51,35	Etanol/Metanol	60	50	100
27	51,02	Etanol/Metanol	60	50	100

Nesse estudo foi adotado o Modelo de Regressão Linear Múltipla (MRLM) para modelar a variável resposta y_i , cafeína, em função das variáveis (X) regressoras sendo elas: modificador, tempo, temperatura e pressão. O modelo,

$$y_i = \beta_0 + \sum_{j=1}^p \beta_j X_{ij} + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad \varepsilon_i \sim N(0, \sigma_e^2), \quad \text{onde } \beta \text{ é o vetor de parâmetros associado a}$$

matriz de incidência X . Analogamente, tem-se $y_i \sim N(\mu_x, \sigma_e^2)$ onde $\mu_x = \beta_0 + \sum_{j=1}^p \beta_j X_{ij}$, é o preditor linear.

Para a modelagem Bayesiana foram consideradas distribuições normais *a priori* não informativas para o vetor de parâmetros β , tal que: $\beta \sim N(0, 10^{-6})$ e $\tau \sim \Gamma(0, 10^{-6})$ (parametrização OpenBUGS, [5]).

A obtenção das distribuições marginais *a posteriori* para todos os parâmetros foi por meio do pacote *BRugs* do programa *R* [6]. Foram gerados 1.000.000 de valores em um processo MCMC (*Monte Carlo*

Markov Chain), considerando um período de descarte amostral de 10.000 valores iniciais, assim a amostra final, tomada em saltos de tamanho 10, contém 100.000 valores gerados. A convergência das cadeias foi verificada por meio do pacote *coda* do programa *R*, pelo critério de Heidelberger e Welch [7].

Foram utilizadas as estimativas da média *a posteriori* da distribuição marginal condicional para os parâmetros. A significância desses no modelo foi verificada caso o valor zero pertence ao seu respectivo intervalo de 95% de credibilidade. Com o objetivo de encontrar um modelo mais simples que explique os dados sem que ocorra perda de informações, ou seja, um modelo mais parcimonioso, utilizamos do critério de seleção de variáveis *Stepwise* que (a cada passo) retira ou acrescenta variáveis que são consideradas redundantes no modelo [8]. O procedimento chega ao final após a realização de todas as combinação possíveis entre as variáveis.

O Critério de comparação entre os modelos foi o valor do *Deviance Information Criterion*, DIC proposto por [9], que faz uso da esperança *a posteriori* do logaritmo da distribuição condicional dos dados, para estabelecer uma medida de qualidade do ajuste. Segundo os autores, os modelos com menores valores para o DIC podem ser considerados mais adequados e diferenças superiores a 5 são consideradas significativas.

3 RESULTADOS E DISCUSSÕES

A Tabela 2 apresenta os resultados para o MRLM utilizando todas as variáveis, nela verificamos que apenas os parâmetros $\beta_2 = -1,754$ ($-3,510; -0,004$) e $\beta_3 = 2,591$ ($0,828; 4,347$), foram significativos.

Tabela 2: Estimativas *a posteriori* dos parâmetros, MRLM, com seus respectivos intervalos de credibilidade ($P_{2,5\%} - P_{97,5\%}$) em nível de 95%.

	<i>Média</i>	<i>Desvio-Padrão</i>	$P_{2,5\%}$	<i>Mediana</i>	$P_{97,5\%}$
β_0	48,930	3,615	41,790	48,940	56,110
β_1	0,218	0,887	-1,531	0,218	1,966
β_2	-1,754	0,886	-3,510	-1,755	-0,004
β_3	2,591	0,889	0,828	2,592	4,347
β_4	0,644	0,886	-1,108	0,646	2,403
σ	3,709	0,588	2,770	3,636	5,052
τ	0,078	0,024	0,039	0,076	0,130

Fazendo uso do Método de seleção de variáveis *Stepwise* todas as combinações possíveis entre as variáveis Modificador, Tempo, Temperatura e Pressão foram testadas. Os resultados são apresentados na Tabela 3.

Tabela 3: Modelos testados e respectivos valores para o DIC

Modelo	Variáveis regressoras do modelo	DIC
M1	Modificador	159,8
M2	Tempo	156,4
M3	Temperatura	151,8
M4	Pressão	149,4
M5	Temperatura, Modificador	153,8
M6	Temperatura, Tempo	149,2
M7	Temperatura, Pressão	153,4
M8	Temperatura, Tempo, Modificador	151,4
M9	Temperatura, Tempo, Pressão	150,8
M10	Tempo, Modificador	150,3
M11	Tempo, Pressão	158,1
M12	Temperatura, Tempo, Pressão, Modificador	153,0
M13	Tempo, Pressão, Modificador	160,0
M14	Temperatura, Pressão, Modificador	155,6

Percebemos que o Modelo 6 possui o menor valor para do DIC, levando em consideração as variáveis Temperatura e Tempo. Este modelo é apresentado na Tabela 4.

Tabela 4: Estimativas a posteriori dos parâmetros, MRLM somente para os parâmetros selecionados via *Stepwise*, com seus respectivos intervalos de credibilidade ($P_{2.5\%} - P_{97.5\%}$) em nível de 95%.

	<i>Média</i>	<i>Desvio-Padrão</i>	$P_{2.5\%}$	<i>Mediana</i>	$P_{97.5\%}$
β_0	50.66000	2.52200	45.67000	50.67000	55.66000
β_1	-1.75200	0.85590	-3.43600	-1.75700	-0.05795
β_2	2.58900	0.85360	0.91080	2.58800	4.27800
σ	3.59300	0.54790	2.71400	3.52700	4.85000
τ	0.08265	0.02394	0.04252	0.08038	0.13580

A cada mudança de nível da variável explanatória tempo temos que a variável resposta cafeína sofre um decréscimo de aproximadamente 1,75%, a mesma análise pode ser feita para a temperatura, ou seja, a cada mudança de nível da variável temperatura temos um acréscimo de 2,59%.

4 CONCLUSÃO

Tendo em vista os resultados apresentados, percebemos que a metodologia Bayesiana se mostrou de fácil utilização e eficiente para modelar a extração de cafeína em função da temperatura, tempo, pressão e modificador. Dessa forma, concluímos que no processo de extração, as variáveis tempo e temperatura são significativas, para um experimento com 40% de cada modificador.

Como proposta de trabalhos futuros, podemos citar a introdução dos percentuais de 20% e 60% para cada modificador, tendo em vista a vastidão do campo de pesquisa na extração de compostos das sementes de *Paullinia cupana*. Além disso, outros compostos importantes podem ser analisados como, por exemplo, a catequina e a epicatequina. A possibilidade realizar uma nova análise bayesiana para além de quantificar a diferença entre cada modificador verificar qual seria o valor ótimo para a extração de cafeína.

Referências

- [1] MARQUES, L. L. M. Investigações fitoquímica e biológicas de extratos obtidos por tecnologia supercrítica de sementes de paullinia cupana (guaraná). 2016. Tese (Doutorado em Ciências Farmacêuticas) - Programa de Pós-Graduação em Ciências Farmacêuticas, Universidade Estadual de Maringá, Maringá, 2016.
- [2] PEREIRA, J. C. R. Cultura do guaranzeiro no Amazonas. Embrapa Amazônia Ocidental-Sistema de Produção (INFOTECA-E), 2005.
- [3] HENMAN, A. R. Guaraná (paullinia cupana var. sorbilis): ecological and social perspectives on an economic plant of the central amazon basin. *Journal of ethnopharmacology*, v. 6, n. 3, p. 311338, 1982.
- [4] DUKE JA. 1987. Handbook of medicinal herbs. 2 ed. Florida: CRC Press LLC.
- [5] SPIEGELHALTER, D.; THOMAS, A.; BEST, N.; GILKS, W. Bugs 0.5: Bayesian inference using gibbs sampling manual (version ii). MRC Biostatistics Unit, Institute of Public Health, Cambridge, UK, p. 159, 1996.
- [6] TEAM, R. C. R language de nition. Vienna, Austria: R foundation for statistical computing, 2000.
- [7] HEIDELBERGER, P.; WELCH, P. D. Simulation run length control in the presence of an initial transient. *Operations Research*, v. 31, n. 6, p. 11091144, 1983.
- [8] HOCKING, R. R. A Biometrics Invited Paper. The Analysis and Selection of Variables in Linear Regression. International Biometrics Society, 1976. 32, 1-49.
- [9] SPIEGELHALTER, D. J.; BEST, N. G.; CARLIN, B. P.; VAN DER LINDE, A. BUGS – A Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society*, v.64, Série B, p.583-639, 2002.