

TUTORIAL PARA APLICAÇÃO DIDÁTICA DE PLANEJAMENTOS DE MISTURAS UTILIZANDO LIBREOFFICE E RSTUDIO

Luiz Henrique Cantarino Adriano^a, João Marcos Beraldo Candido^{a,lb}, Olívia Brito de Oliveira Moreira^{a,lb}, Ieda Spacino Scarmínio^b, Roy Edward Bruns^c e Marcone Augusto Leal de Oliveira^{a,*,lb}

^aDepartamento de Química, Instituto de Ciências Exatas, Universidade Federal de Juiz de Fora, 36036-900 Juiz de Fora – MG, Brasil

^bDepartamento de Química, Centro de Ciências Exatas, Universidade Estadual de Londrina, 86051-990 Londrina – PR, Brasil

^cInstituto de Química, Universidade Estadual de Campinas, 13083-970 Campinas – SP, Brasil

Recebido em 21/07/2023; aceito em 10/11/2023; publicado na web 29/01/2024

TUTORIAL FOR DIDACTIC APPLICATION OF MIXTURE DESIGNS USING LIBREOFFICE AND RSTUDIO. Mixture designs are becoming frequently used as a powerful tool in modern chemical research. Their versatility opens the door to countless applications, making it a must-have tool in any experimental environment. Mixture designs combined with mathematical regressions allow us to understand how component proportions affect a measured response leading to system optimization at minimum cost and providing an accurate description of system behaviour. Dedicated software is often used to perform data analysis in mixture design experiments, but many of them are only commercially available hindering the use of mixture designs by scientific groups. The availability of free software would make these designs even more attractive, allowing users to benefit from the creative potential of this toolset without the prohibitive cost of proprietary tools. RStudio and LibreOffice are some of the free software with the capability to perform the necessary data analysis accompanying mixture designs. To exemplify how mixture design analysis can be executed the book “*Como Fazer Experimentos*” was used as reference material. Both software presented good results and the potential to be used to perform data analysis for mixture design experiments.

Keywords: mixture designs; RStudio; LibreOffice.

INTRODUÇÃO

Os planejamentos de misturas (PM) são baseados em princípios e fundamentos da estatística multivariada, sendo ferramentas muito poderosas para otimizar processos empíricos, pois tangenciam, simultaneamente, a maximização da geração de informações úteis e a minimização do esforço experimental ou laboral, impactando, substancialmente, o tempo e o custo do efetivo procedimental. O PM é um dos vários segmentos pertencentes a grande área da quimiometria, e tem como conceito e fundamento selecionar e modelar fatores ou variáveis, a serem considerados nos ensaios experimentais, cuja resposta estará sob a influência da proporção quantitativa de cada constituinte presente na mistura, ou seja, o somatório das proporções ou contribuições de cada variável será sempre igual a 1 ou 100%. Durante os procedimentos de investigação do modelo a ser construído, ajustado e otimizado, o PM é definido com o foco de considerar o maior número de parâmetros possíveis e imprescindíveis para descrever a realidade e a necessidade empírica, assim como corroborar a significância heurística do problema em questão com a significância estatística.^{1,2} O monitoramento da(s) resposta(s) será realizado através do ajuste de uma função polinomial, considerando níveis independentes das variáveis dependentes para os componentes da mistura. É importante destacar que as aplicações envolvendo PM extrapolam o ambiente acadêmico, permeando variados segmentos de interesse no mercado, evidenciando alguns exemplos, como o setor alimentício, combustível, biocombustível, polímeros, farmacêutico, novos materiais, dentre outros.³⁻⁵

A comunidade inexperiente, porém, interessada em usar o PM, pode ser induzida a pensar equivocadamente que a implementação dos cálculos necessários para a construção dos modelos exige a aquisição

de habilidades e competências em programação. No entanto, plataformas de programação com código livre (“*free software*”), ou seja, as plataformas que proporcionem aos usuários a liberdade de executar, copiar, distribuir, estudar, mudar e melhorar, tornam possível que as rotinas, previamente implementadas, sejam executadas de forma sistemática, didática e simples, facilitando, a compreensão dos conceitos e fundamentos imprescindíveis para o aprendizado e aplicação das ferramentas, assim como realizar a interpretação dos resultados de forma segura e confiável.

Neste contexto, os autores deste trabalho têm como objetivo apresentar um conjunto de programas, como o LibreOffice, um *software* de planilhas, e o RStudio, um *software* dedicado a cálculos estatísticos baseado na linguagem de programação R, como metodologias didáticas para o ensino e elaboração de planejamentos de misturas, democratizando seu uso no meio acadêmico e mercado de trabalho.

Programas

Neste trabalho, serão apresentados o RStudio versão 2022.07.1, e o LibreOffice versão 7.4.3, os quais são programas que podem ser aplicados para aquisição, organização, tratamento, análise e transferência de dados, construção e ajuste de modelos, assim como a construção e interpretação de gráficos e tabelas. Com o auxílio destes programas, será demonstrado, sob uma ótica didática, aplicações de PM descritos no livro *Como Fazer Experimentos*.⁶ Além disso, códigos e linhas de comando serão explicitados para fornecer um guia aos usuários que almejam utilizar estes programas computacionais no desenvolvimento de outros PM.

LibreOffice

O LibreOffice, o *software* de planilhas eletrônicas da cooperativa *The Document Foundation*, é um conjunto de aplicativos utilizados

*e-mail: marcone.oliveira@ufjf.br

para criação, edição, formatação e visualização de documentos, planilhas, apresentações, desenhos, bancos de dados, etc., sendo compatível com diferentes sistemas operacionais, incluindo Windows e Linux. É um *software* livre e de fonte aberta, o que permite a cópia, implementação e distribuição gratuita entre os usuários,⁷ cuja aquisição pode ser feita online diretamente na página oficial do LibreOffice (<https://pt-br.libreoffice.org/>). Além disso, é considerada uma plataforma com grande popularidade e familiaridade, possuindo uma interface simples, mas com grandes potencialidades, sendo possível implementar cálculos matriciais e estatísticos. Diante disso, o LibreOffice se torna uma opção muito atrativa para o estudo de PM. Vale destacar a existência de outro trabalho fazendo uso do *software* de planilhas pago Excel onde o tema de planejamento de também foi abordado.⁸

R e RStudio

O “R” é uma linguagem de programação criada em 1993 com o objetivo de ser um *software* livre, o que significa que o “R” possui uso e compartilhamento gratuito de todo seu código. Ele conta com um sistema de colaboração chamado *comprehensive R archive network* (CRAN) onde usuários criam e compartilham pacotes de códigos com funções para diferentes usos, entre eles estatística, ciência de dados e aprendizado de máquina (*machine learning*).

O R possui um ambiente de desenvolvimento integrado chamado RStudio, também livre, que facilita o seu uso por possuir uma interface gráfica (*graphical user interface*, GUI) mais intuitiva, sendo possível carregar, trabalhar, analisar e visualizar dados complexos de forma simples e funcional. Para utilização, o usuário deve inicialmente fazer a instalação do R, cujo endereço para *download* é disponibilizado na plataforma CRAN (<https://cran.r-project.org/>). Após a execução do arquivo “.exe”, deve-se seguir as etapas indicadas pelo próprio arquivo recém adquirido ao abri-lo. Uma vez concluída essa etapa, o RStudio pode ser instalado, cuja aquisição também é feita diretamente na plataforma (<https://posit.co/download/rstudio-desktop/>). Novamente, ao seguir as indicações do arquivo de execução do instalador “.exe” recém adquirido, o *software* estará pronto para uso.

Entre os vários pacotes presentes no R temos o *mixexp* que permite trabalhar com planejamento de misturas dentro da interface do RStudio.⁹ Tal pacote permite a utilização de diferentes modelos de planejamento de mistura, incluindo análise com pseudocomponentes, visualização de dados e planejamentos mistos. O manual deste e outros pacotes auxiliares usando nos guias disponíveis neste artigo podem ser encontrados no CRAN.¹⁰

FUNDAMENTAÇÃO TEÓRICA

Quando a construção de modelos empíricos é o objetivo do estudo, a modelagem por mínimos quadrados ordinários (MQO) pode ser considerada a técnica mais comumente empregada. Entretanto, na avaliação da influência das variáveis do sistema, planejamentos fatoriais (PF) ou PM podem ser aplicadas equivocadamente.⁶ Existe uma diferença fundamental entre PF e PM. Em PM, a resposta depende apenas da mudança na proporção relativa entre seus componentes, ou seja, a soma das proporções deve permanecer constante. Dessa forma, neste tipo de planejamento experimental o objetivo é descobrir a proporção dos componentes de uma determinada mistura que geram a melhor resposta. Diferentemente do PF, cujo objetivo, além de otimização, é medir a influência na resposta quando ocorre a variação do nível entre os diversos fatores avaliados, seja quantidades intensivas como temperatura ou extensivas como concentrações.¹¹ Os níveis destes fatores podem ser mudados independente de um para outro. Além disso, em contrapartida aos PF, em problemas envolvendo misturas, as variáveis são matematicamente

dependentes, de modo que as proporções dos componentes somam uma unidade, conforme Equação 1.¹²

$$\sum_{i=1}^q x_i = 1 \quad 0 \leq x_i \leq 1 \quad (1)$$

Existem diversas abordagens em PM,¹³ como os planejamentos em rede simplex, centróide-simplex, propostos por Scheffé,¹⁴ e vértices extremos, proposto por McLean e Anderson,¹⁵ que possibilitam a exploração de diversificadas regiões experimentais,¹⁶ de modo que na seleção do modelo mais adequado se deve considerar o número de fatores e interações a serem estudadas, a complexidade do planejamento, a validade estatística e eficácia da modelagem sobre o sistema, bem como a capacidade da implementação, os custos e tempo associadas com cada modelo.¹⁷

Neste trabalho será apresentada uma estratégia simples de uso de planejamento e modelos de misturas que pode ser seguido pelo usuário. É claro que, condições de operacionalidade, tempo, custo, entre outros, podem necessitar de modificações para cada caso em particular. Como ponto de partida, considere a possibilidade de execução dos experimentos com até dez misturas representados pelos pontos no triângulo na Figura 1.

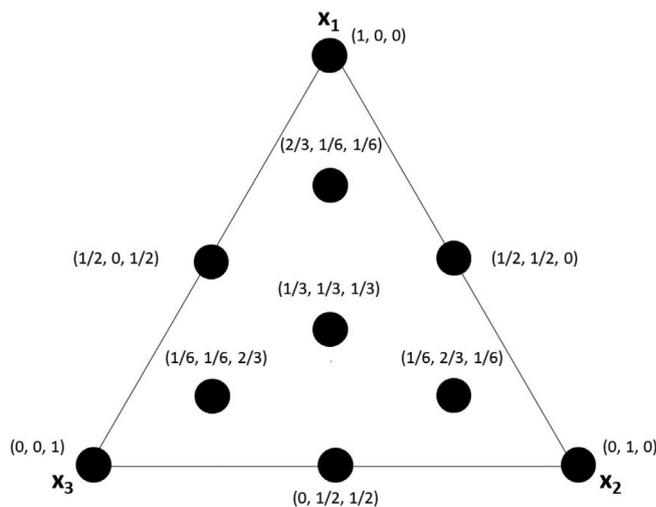


Figura 1. Planejamento centróide simplex para uma mistura de três componentes

O modelo linear pode ser usado para descrever misturas de componentes formando soluções ideais. Para uma mistura de três componentes a resposta é determinada pela Equação 2.

$$\hat{y} = b_1^* x_1 + b_2^* x_2 + b_3^* x_3 + \varepsilon \quad (2)$$

Nota-se que, diferentemente dos modelos para variáveis independentes, este modelo não tem um termo constante porque a Equação 1 foi usada na sua derivação. Os três coeficientes de regressão, ($i = 1, 2, 3$) são descritos com um sobrescrito em asterisco, para enfatizar que foram determinados através das variáveis vinculadas à Equação 1, e não de variáveis independentes.¹¹ Os coeficientes podem ser determinados sem a necessidade empírica de realizar qualquer mistura, pois os valores das respostas são simplesmente médias ponderadas das respostas dos componentes puros. Os componentes puros estão representados pelos pontos nos vértices do simplex triangular na Figura 1. Contudo, um maior número de experimentos com misturas deve ser executado para descrever possíveis interações de sinergismo ou antagonismo entre os componentes.

Interações binárias são contempladas no planejamento em rede

simplex, quando, além dos pontos que representam as componentes puras, localizados nos vértices do triângulo, também estão presentes as misturas binárias em proporções iguais, localizadas pelos pontos correspondentes a metade da distância das arestas dos lados do triângulo localizado na Figura 1. Sendo assim, é possível construir o modelo quadrático, conforme descrito na Equação 3, no qual os três últimos termos do modelo correspondem as três possíveis interações binárias no sistema de três componentes.

$$\hat{y} = b_1^*x_1 + b_2^*x_2 + b_3^*x_3 + b_{12}^*x_1x_2 + b_{13}^*x_1x_3 + b_{23}^*x_2x_3 + \varepsilon \quad (3)$$

Neste modelo é mandatório a realização de no mínimo seis experimentos para determinar os coeficientes. Para verificar se o modelo quadrático descreve adequadamente o sistema de misturas, é necessário realizar experimentos, com proporções de misturas, correspondentes a alguns pontos no interior do triângulo, com a finalidade de validar o modelo, ou seja, mostrar que o modelo está bem ajustado aos pontos experimentais, sem a ocorrência de falta de ajuste no intervalo de significância considerado.

Com apenas mais um experimento é possível construir o modelo cúbico especial, onde não somente as interações binárias são consideradas, mas também a interação ternária, ou seja, a interação dos três componentes. Tais considerações são preconizadas para descrição dos polinômios canônicos de Scheffé, de modo que o modelo cúbico especial é descrito pela Equação 4.¹⁸

$$\hat{y} = b_1^*x_1 + b_2^*x_2 + b_3^*x_3 + b_{12}^*x_1x_2 + b_{13}^*x_1x_3 + b_{23}^*x_2x_3 + b_{123}^*x_1x_2x_3 + \varepsilon \quad (4)$$

Este modelo tem sete coeficientes de regressão, e é por isso que é mandatório a execução de no mínimo sete ensaios experimentais, com diferentes proporções de misturas. Os estatísticos recomendam a execução da mistura ternária em partes iguais além das misturas da rede simplex, sendo este planejamento chamado de simplex-centroide. Para determinar se o modelo não tem falta de ajuste também se recomenda a inclusão dos três pontos axiais com as proporções (2/3, 1/6, 1/6), (1/6, 2/3, 1/6) e (1/6, 1/6, 2/3) que podem ser visualizados na Figura 1.

Uma estratégia frequentemente utilizada em experimentação com misturas para execução dos experimentos correspondentes as dez misturas, Figura 1, chamada de simplex-centroide com pontos axiais, consiste em progressivamente testar os modelos linear, quadrático e cúbico especial para falta de ajuste. O modelo escolhido para representar os dados experimentais deve ser aquele mais simples dentre os três, e que não apresente falta de ajuste no nível de confiança aceitável para a aplicação desejada.

Ainda considerando a Equação 1, um recorte pode ser analisado mediante limitações nos valores assumidos por cada constituinte na mistura, em função de restrições intrínsecas a cada sistema. Tais limitações, sejam elas superiores e/ou inferiores, dificultam a manipulação do espaço experimental cuja condição de contorno advém da aplicação do conceito de pseudocomponentes,¹⁹ que auxiliam na descrição de situações nas quais não é possível utilizar um componente com total pureza no modelo. É importante destacar que, de forma rigorosa, os pseudocomponentes são definidos como combinações dos componentes originais. O principal motivo para introduzir os pseudocomponentes em um ensaio experimental é simplificar tanto o planejamento experimental quanto o ajuste dos modelos, em comparação com quando são realizados com base no sistema de componentes original. Contudo, é muito importante destacar que os pseudocomponentes são “pseudo” e se alguém deseja fazer inferências sobre os componentes que realmente compõem o sistema de mistura, é mandatório ajustar um modelo com os componentes originais ou fazer a transformação inversa do pseudocomponente, de volta aos componentes originais, para produzir

um modelo ajustado nas componentes originais. Uma vez respeitado todo o formalismo inerente ao uso das pseudocomponentes, uma nova região pode ser visualizada como uma fração do espaço experimental original, preservando a geometria e propriedades.⁶

Tais considerações são preconizadas para descrição dos polinômios canônicos de Scheffé.¹⁸ Para determinar a equação de tais polinômios, as estimativas dos parâmetros do modelo podem ser encontradas através do método dos mínimos quadrados ordinários, bem como os erros padrão associados, segundo as Equações 5 e 6 sendo \mathbf{X} a matriz cujos elementos são determinados pelas proporções dos componentes da mistura, \mathbf{y} , um vetor das respostas dos experimentos, b^* , o vetor dos coeficientes do modelo, e s_p^2 a variância determinada a partir das réplicas das misturas.¹²

$$b^* = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (5)$$

$$\text{var}(b^*) = (\mathbf{X}^T \mathbf{X})^{-1} s_p^2 \quad (6)$$

Os conceitos relacionados ao formalismo matemático envolvendo o planejamento de misturas foram abordados superficialmente no escopo desse artigo. Para aprofundamento sobre o tema sugerimos como referências os seguintes livros: *Como Fazer Experimentos*⁶ e *Experiments with Mixtures*.¹¹

APLICAÇÕES

A partir das várias aplicações do uso de planejamentos e modelos de mistura apresentadas no livro *Como Fazer Experimentos*,⁶ a execução dos cálculos usando *software* livre são demonstradas neste trabalho, a partir do caso descrito a seguir. “Preparar blendas – isto é, misturas – é uma das formas que os pesquisadores mais usam para tentar otimizar as propriedades de produtos poliméricos. Entre as principais propriedades descritas nas patentes desses materiais estão a resistência ao impacto, a facilidade de processamento e a resistência à tração. Em um estudo⁶ visando maximizar a resistência à tração de blendas de polifluoreto de vinilideno (PVDF), polimetacrilato de metila (PMMA) e poliestireno (PS), foram obtidos os resultados apresentados na Tabela 1. Essa tabela representa a otimização de um processo de síntese de um polímero por meio de um PM, cujo objetivo é determinar a proporção ideal dos componentes para obter a máxima resistência mecânica no polímero sintetizado. PVDF, PMMA e PS são representados por x_1 , x_2 e x_3 , respectivamente, e a variável resposta é representada por y . Os 7 primeiros ensaios correspondem a um planejamento centroide-simplex, cujas réplicas experimentais estão representadas pelos ensaios de 8 a 16. Nos ensaios 17 a 20 a proporção PVDF:PS é mantida igual a 1, variando-se apenas a proporção de PMMA.⁶ Por exemplo, no ensaio 17, x_1 e x_3 foram mantidos na proporção de 49.0% cada, o que corresponde a proporção de PVDF:PS igual a 1, sendo a quantidade total de polímero completa por 2% de PMMA.” Os pontos experimentais são representados na Figura 2.

Construção dos modelos de misturas utilizando o LibreOffice

Existem duas formas de construir um modelo de regressão na interface do LibreOffice, sendo a primeira delas através de programação na própria linguagem do sistema, o que requer além do conhecimento técnico do programa, conhecimento em álgebra linear para a implementação dos cálculos matriciais. Para essa forma de trabalho é importante entender as relações operacionais dos comandos inerentes a plataforma para executar a regressão, conforme descrito na Equação 5. A implementação no LibreOffice para o cálculo da pseudo inversa, $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$, é dado pela seguinte linha de código:

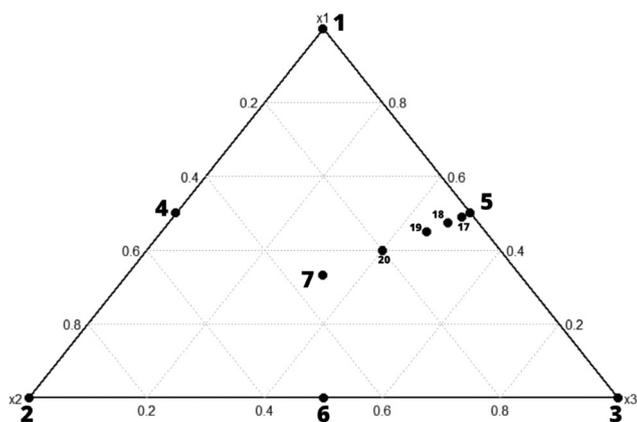


Figura 2. Representação gráfica dos ensaios correspondentes aos dados de planejamento de mistura das blendas poliméricas

Tabela 1. Planejamento para o estudo de misturas poliméricas ternárias¹

Ensaio	x_1	x_2	x_3	y
1	1,0	0	0	51,2
2	0,0	1,0	0,0	20,0
3	0,0	0,0	1,0	20,2
4	0,5	0,5	0,0	44,8
5	0,5	0,0	0,5	23,5
6	0,0	0,5	0,5	35,7
7	0,333	0,333	0,333	45,4
8	0,0	1,0	0,0	20,8
9	0,0	1,0	0,0	23,0
10	0,0	0,0	1,0	18,6
11	0,0	0,5	0,5	23,0
12	0,0	0,5	0,5	38,0
13	0,5	0,0	0,5	28,3
14	0,5	0,0	0,5	22,6
15	0,5	0,0	0,5	25,4
16	0,333	0,333	0,333	46,5
17	0,490	0,020	0,490	21,5
18	0,475	0,050	0,475	23,5
19	0,450	0,100	0,450	35,1
20	0,400	0,200	0,400	33,6

$=\text{MATRIZ.MULT}(\text{MATRIZ.MULT}(\text{MATRIZ.INVERSO}(\text{MATRIZ.MULT}(\text{TRANSPOR}(X);X));\text{TRANSPOR}(X));Y)$.

Em ambos os casos, a matriz X tem colunas com as proporções relativas dos componentes, e a matriz y é um vetor coluna com as respostas experimentais. O resultado dessa função é uma matriz contendo os coeficientes ajustados a partir do modelo de regressão construído. Um tutorial completo para o cálculo dos coeficientes do modelo de regressão, a partir da linha de código descrita acima, está disponível na seção 1 do Material Suplementar.

A segunda maneira de construir um modelo de regressão na interface do LibreOffice requer uma sequência de comandos, o que facilita a implementação do modelo, de maneira rápida, eficiente, precisa e correta, para os usuários com menos afinidade com a interface de programação. O passo a passo para a construção do modelo de regressão utilizando os recursos do comando “Análise de Dados”, no LibreOffice, é apresentado a seguir:

(i) Escrever a planilha com os dados. Como exemplo, os dados da

Tabela 1 podem ser importados para o LibreOffice conforme descrito também na seção 1 do Material Suplementar;

(ii) Na barra de tarefas principal do *software*, clicar em “dados”;

(iii) Selecionar a opção “estatística” > “regressão”;

(iv) Na seção “intervalo das variáveis independentes (X)” selecionar a matriz X . Destaca-se que se o usuário optar por selecionar os dados com os respectivos rótulos de cada coluna, essa informação deve ser assinalada na janela “regressão”, conforme Figura 3;

(v) Na seção “intervalo das variáveis dependentes (Y)” selecionar o vetor coluna y ;

(vi) Na seção “resultados para” selecionar um espaço vazio na planilha para a inserção dos resultados;

(vii) Selecionar a opção “regressão linear”;

(viii) Selecionar a opção “forçar intercessão a ser zero”, conforme Figura 3.

A opção “forçar intercessão igual a zero” garante a premissa que a soma das proporções dos constituintes seja igual a um, condição *sine qua non* para gerar os resultados corretamente em planejamento de misturas. A abordagem apresentada permite o cálculo dos parâmetros de regressão de maneira simples e rápida, mesmo para novos usuários da plataforma.

Modelo linear

A entrada dos dados para a construção do modelo linear é apresentada na Figura 3. A coluna A faz referência somente a nomeação dos ensaios experimentais, e não é utilizada para nenhum cálculo matricial. Neste caso, a tabela utilizada na construção dos modelos consiste somente de quatro colunas, sendo as colunas B, C, e D, as que apresentam as proporções relativas para cada um dos 3 solventes do problema, compondo a matriz X ; e a coluna E contém a resposta monitorada, ou seja, o vetor y . Conforme descrito anteriormente, há a opção de selecionar a entrada de dados com ou sem rótulos (primeira linha da planilha), mediante a seleção da opção “Ambos intervalos Y e X possuem rótulos”. Se for selecionada a opção, é mandatório que os rótulos sejam considerados na seleção de entrada dos dados. Do contrário, não é preciso.

Após a execução do passo a passo, de (i) a (viii), o resultado retorna somente três coeficientes. Vários valores são gerados no resultado da análise, de modo que iremos nos atentar somente aos coeficientes apresentados na Figura 4. Na coluna “coeficientes”, como o próprio nome sugere, são apresentados os coeficientes de cada uma das variáveis na equação de regressão, descrita pela Equação 7.

$$\hat{y} = 44,26(\pm 6,0)x_1 + 29,68(\pm 4,7)x_2 + 19,65(\pm 5,1)x_3 \quad (7)$$

A coluna rotulada como “erro padrão” apresenta os erros associados a cada um dos coeficientes, das variáveis. Esse parâmetro, bem como aqueles representados pelas colunas “estatística-t”, “valor p”, e as duas seguintes, equivalentes ao limite inferior e superior do intervalo de confiança, associada avaliação da significância estatística e outras fontes de variação dos coeficientes do modelo²⁰ cujos cálculos completos serão discutidos nas próximas seções.

Modelos quadrático e cúbico especial

Para a construção dos demais modelos, ou seja, modelo quadrático ou cúbico especial, os procedimentos são similares, mudando somente a matriz X inserida, a qual é específica para cada modelo, como mostrado nas Figuras 4S (modelo quadrático), e Figura 5S (modelo cúbico especial). Novas colunas precisam ser inseridas para considerar na modelagem o efeito de interação entre os componentes, no qual o valor encontrado não é nada mais que a multiplicação das proporções dos componentes considerados. Para o modelo quadrático são considerados somente a interações binárias, x_1x_2 , x_1x_3 e x_2x_3 , enquanto

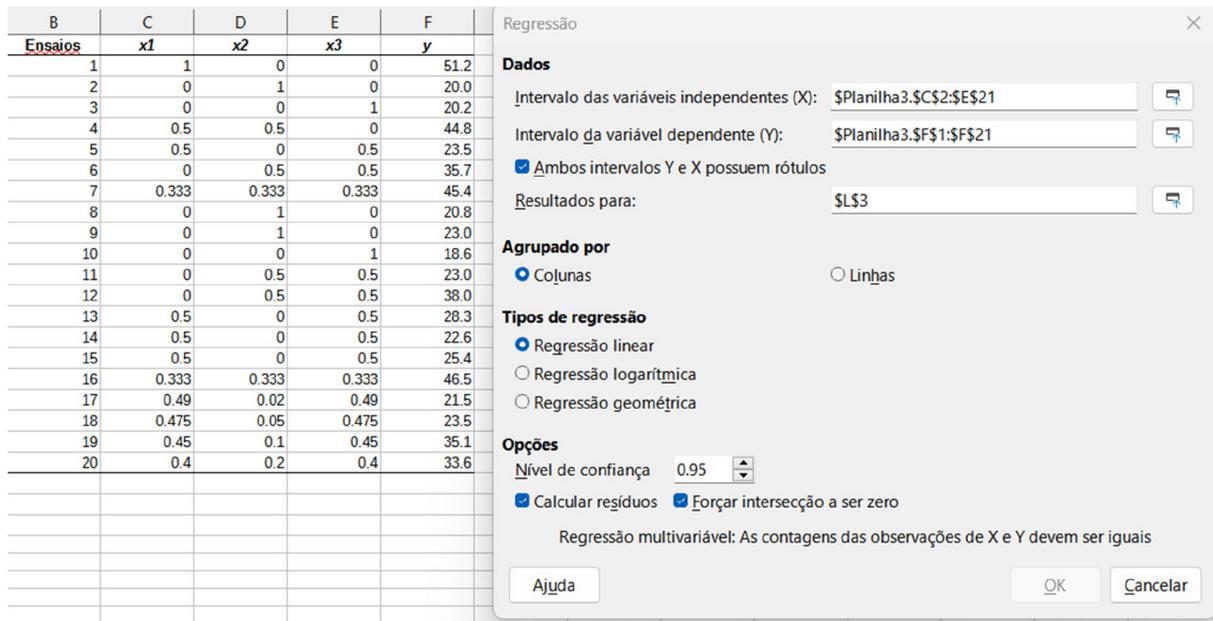


Figura 3. Matriz X para o modelo linear de planejamento de misturas consiste das colunas x1, x2 e x3. A coluna y forma o vetor y

	Coefficientes	Erro padrão	Estatística-t	Valor P	Inferior 95%	Superior 95%			
Interceptação	0	#N/DISP	#N/DISP	#N/DISP	#N/DISP	#N/DISP	MS	F	Significância F
x1	44,25706504	5,993478486	7,384203538	1,06716E-06	31,61193076	56,90219931	589,0911951	71,43485745	5,80459E-13
x2	29,68255089	4,661193878	6,368014647	7,00702E-06	19,84829143	39,51681034	86,8964664		
x3	19,64729304	5,059087049	3,883564929	0,001193502	8,973552372	30,3210337			

Figura 4. Resultado do modelo linear calculado no LibreOffice

para o modelo cúbico especial, além destas, também é considerada a interação ternária $x_1x_2x_3$. Para construção dos modelos quadráticos e cúbicos utilizando o LibreOffice, um tutorial é disponibilizado na seção 2 do Material Suplementar.

Construção dos modelos de regressão utilizando o RStudio

Para utilização do RStudio, também iniciamos com a matriz correspondente aos dados da Tabela 1. A primeira etapa, ainda a partir do LibreOffice, é salvar a planilha de dados como formato “Planilha ODF”. Em seguida, uma nova pasta deve ser criada, dentro de um diretório da preferência do usuário, e nesta, deve-se inserir o arquivo de dados recém-criado em formato “.ODS”. Essa será a pasta de trabalho que deve ser inserida no RStudio, considerando os comandos descritos a seguir.

Já utilizando o RStudio, a próxima etapa é a instalação de um pacote que permita ao RStudio se comunicar com arquivos no formato “.ODS” gerado pelo LibreOffice, através do pacote ReadODS. A instalação do pacote é muito simples sendo feita pela seguinte linha de comando: `install.packages("readODS")`.

Uma vez instalado, é necessário carregar o pacote, utilizando: `library(readODS)`.

Para importar a planilha, precisamos selecionar em qual pasta tal planilha está contida. Para isso, na barra de tarefas principal do RStudio clicar em: `session > set working directory > choose directory`, e selecionar a pasta.

Em seguida, os dados da planilha em formato “.ODS” devem ser lidos pelo *software*, a partir do comando: `dados > - read_ods("Tabela1.ods", col_names = T)`, na qual, conforme exemplo acima, “dados” será o nome da planilha de dados a ser registrada pelo RStudio e “Tabela1.ods” o nome do arquivo em formato “.ODS”. O comando `col_names = T` indica ao *software* que a planilha de dados contém rótulos em suas colunas.

Finalizado o procedimento, a planilha ficará carregada no ambiente do RStudio, como apresentado na Figura 5. É importante se atentar aos nomes registrados nas colunas, pois tais rótulos serão importantes para os procedimentos consecutivos.

Para o ajuste dos coeficientes é necessário construir a regressão pelo método dos mínimos quadrados usando a função `lm()`, que possui os seguintes parâmetros:

`lm(variável dependente ~ variável independente, data = matriz de dados)`.

A forma de representar o modelo selecionado para a regressão é através dos parâmetros. Para definir o intercepto com o valor zero, é necessário usar o comando `-1` nos parâmetros, como apresentado a seguir.

Modelo linear

O comando para construção de modelos lineares, segue o formato genérico:

`NOME DE SAÍDA DOS DADOS < -lm(formula = y~x1+x2+x3-1, data = "MATRIZ DE DADOS")`.

É importante destacar que os indicadores y, x1, x2, x3 e demais variáveis que forem utilizadas, quando for o caso, devem estar escritas da mesma forma que o rótulo da tabela de dados recém importada no RStudio. Do contrário, o *software* não reconhece quais dados serão utilizados para construção do modelo. Ajustados estes detalhes, basta inserir o comando no RStudio, conforme o exemplo: `modelo_linear<-lm(formula = y~x1+x2+x3-1, data =dados)`

Neste momento, uma nova linha de dados será criada no ambiente do RStudio contendo as respostas. Para visualização dos coeficientes, inserir o comando: `summary(NOME DE SAÍDA DOS DADOS)`.

De acordo com o exemplo, `summary(modelo_linear)`.

O resultado do RStudio retorna o valor dos coeficientes, assim como os respectivos erros padrões (Figura 6).

De modo similar, com as devidas adaptações são construídos os modelos quadráticos e cúbico especial, apresentados a seguir. Os

Ensaios	x1	x2	x3	y	
1	1	1.000	0.000	0.000	51.2
2	2	0.000	1.000	0.000	20.0
3	3	0.000	0.000	1.000	20.2
4	4	0.500	0.500	0.000	44.8
5	5	0.500	0.000	0.500	23.5
6	6	0.000	0.500	0.500	35.7
7	7	0.333	0.333	0.333	45.4
8	8	0.000	1.000	0.000	20.8
9	9	0.000	1.000	0.000	23.0
10	10	0.000	0.000	1.000	18.6
11	11	0.000	0.500	0.500	23.0
12	12	0.000	0.500	0.500	38.0
13	13	0.500	0.000	0.500	28.3
14	14	0.500	0.000	0.500	22.6
15	15	0.500	0.000	0.500	25.4
16	16	0.333	0.333	0.333	46.5
17	17	0.490	0.020	0.490	21.5
18	18	0.475	0.050	0.475	23.5
19	19	0.450	0.100	0.450	35.1
20	20	0.400	0.200	0.400	33.6

Figura 5. Planilha carregada no ambiente do RStudio

```
> modelo_linear<-lm(formula = y~x1+x2+x3-1, data = dados)
> summary(modelo_linear)

Call:
lm(formula = y ~ x1 + x2 + x3 - 1, data = dados)

Residuals:
    Min       1Q   Median       3Q      Max
-10.407  -8.367  -1.356   7.165  15.336

Coefficients:
      Estimate Std. Error t value Pr(>|t|)
x1    44.257     5.993   7.384 1.07e-06 ***
x2    29.683     4.661   6.368 7.01e-06 ***
x3    19.647     5.059   3.884 0.00119 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.322 on 17 degrees of freedom
Multiple R-squared:  0.9265,    Adjusted R-squared:  0.9135
F-statistic: 71.43 on 3 and 17 DF,  p-value: 7.657e-10
```

Figura 6. Output da função lm() para o modelo linear

resultados referentes as construções destes modelos são apresentados nas Figuras 7 e 8.

Modelo quadrático

```
modelo_quadratico<-lm(formula = y~x1+x2+x3+x1:x2+x1:
x3+x2:x3-1, data = dados)
summary(modelo_quadratico)
```

Modelo cúbico especial

```
modelo_cubico<-lm(formula = y~x1+x2+x3+x1:x2+x1:x3+x2:
x3+x1:x2:x3-1, data = dados)
summary(modelo_cubico)
```

Fontes de variação do modelo construído

A falta de ajuste é um parâmetro estatístico que envolve o estudo da ANOVA, a qual permite a avaliação do quão bem ajustado de fato está o modelo construído, ou seja, se o modelo permite predições significativas no intervalo de confiança considerado. A falta de ajuste é uma razão entre a média quadrática da falta de ajuste (MQ_{faj}) e a média quadrática do erro puro (MQ_{ep}) que deve ser comparada a um valor F tabelado de acordo com os graus de liberdade de cada um desses parâmetros, onde um valor de F calculado menor do que o F tabelado, para os graus de liberdade estabelecidos, significa que o modelo ajustado não apresenta falta de

```

> modelo_quadratico<-lm(formula = y~x1+x2+x3+x1:x2+x1:x3+x2:x3-1, data =dados)
> summary(modelo_quadratico)

Call:
lm(formula = y ~ x1 + x2 + x3 + x1:x2 + x1:x3 + x2:x3 - 1, data = dados)

Residuals:
    Min       1Q   Median       3Q      Max
-11.2060  -1.7875   0.2169   2.4870   7.6140

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
x1          49.725     5.106   9.739 1.30e-07 ***
x2          20.793     2.962   7.021 6.04e-06 ***
x3          18.663     3.626   5.147 0.000148 ***
x1:x2       61.835    20.266   3.051 0.008629 **
x1:x3      -36.883    14.418  -2.558 0.022757 *
x2:x3       57.913    14.412   4.018 0.001269 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.149 on 14 degrees of freedom
Multiple R-squared:  0.9815,    Adjusted R-squared:  0.9736
F-statistic: 124 on 6 and 14 DF,  p-value: 2.551e-11

```

Figura 7. Output da função `lm()` para o modelo quadrático

```

> modelo_cubico<-lm(formula = y~x1+x2+x3+x1:x2+x1:x3+x2:x3+x1:x2:x3-1, data =dados)
> summary(modelo_cubico)

Call:
lm(formula = y ~ x1 + x2 + x3 + x1:x2 + x1:x3 + x2:x3 + x1:x2:x3 -
  1, data = dados)

Residuals:
    Min       1Q   Median       3Q      Max
-9.3867  -0.8872   0.2242   2.2904   5.6133

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
x1          51.087     4.444  11.495 3.49e-08 ***
x2          21.319     2.566   8.308 1.48e-06 ***
x3          19.343     3.143   6.155 3.46e-05 ***
x1:x2       36.229    20.481   1.769 0.10035
x1:x3      -48.201    13.307  -3.622 0.00310 **
x2:x3       48.222    13.078   3.687 0.00274 **
x1:x2:x3   235.986    98.124   2.405 0.03179 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.445 on 13 degrees of freedom
Multiple R-squared:  0.9872,    Adjusted R-squared:  0.9803
F-statistic: 143.5 on 7 and 13 DF,  p-value: 2.795e-11

```

Figura 8. Output da função `lm()` para o modelo cúbico

ajuste. Vale ressaltar que, o cálculo da falta de ajuste só é possível quando o número de níveis experimentais é maior que o número de parâmetros no modelo, motivo pelo qual no exemplo mostrado foram realizados experimentos além dos pontos mínimos necessários para a construção do modelo.

A significância da regressão, por sua vez, está associada a dependência da resposta nas proporções dos componentes, e é calculada comparando a razão entre a média quadrática da regressão (MQ_{reg}) e a média quadrática dos resíduos (MQ_r), onde um valor de F calculado maior do que o F tabelado, para os graus de liberdade estabelecidos, caracteriza que a regressão ajustada é estatisticamente significativa. Do contrário, ou seja, se o valor de F calculado for menor do que o valor de F tabelado, caracteriza que a regressão ajustada não é estatisticamente significativa.

A falta de ajuste, significância da regressão e demais fontes de variação do modelo são baseadas nas equações descritas na Tabela 1S e podem ser calculadas a partir de cálculos matriciais. Um tutorial contendo todos os cálculos matriciais para determinação das

fontes de variação do modelo está disponível na seção 3 do Material Suplementar. O RStudio também é uma possibilidade para cálculo de tais valores, a partir do uso do pacote “`alr3`”. Uma vez que o “`alr3`” não está mais disponível na plataforma CRAN, alguns passos adicionais são necessários para instalação de pacotes disponibilizados por outras plataformas, feito a partir dos comandos abaixo:

```

install.packages("remotes")
library(remotes)
remotes::install_github("cran/alr3")

```

Uma vez instalado, o pacote “`alr3`” deve ser carregado no RStudio:

```
library(alr3).
```

Em seguida, esse pacote será utilizado para calcular, a partir do ANOVA, os fatores de variação do modelo construído. Assim, ao inserir a linha de comando, deve-se escrever o nome de saída dado pelo usuário a cada um de seus modelos, conforme exemplos:

```

pureErrorAnova(modelo_linear)
pureErrorAnova(modelo_quadratico)
pureErrorAnova(modelo_cubico)

```

Os dados resultantes dos modelos linear, quadrático e cúbico estão descritos nas Figura 9, 10 e 11, respectivamente.

Construção da superfície de resposta

A partir do LibreOffice, não é possível gerar uma superfície de resposta que possa ser interpretada de forma simples, de modo que, a elaboração de uma superfície útil demanda demasiado tempo e, ainda assim, permeada por uma análise complexa dos resultados. É importante destacar que por não ser um programa dedicado a esta finalidade, é razoável que existam limitações, posto que um gráfico ternário não segue a mesma lógica de construção de gráficos comumente utilizados na estatística descritiva. Enquanto os gráficos tradicionais representam 2 variáveis em um espaço de duas dimensões, o gráfico ternário necessita representar 4 variáveis (x_1 , x_2 , x_3 e y) no mesmo espaço de duas dimensões. O RStudio possui um pacote dedicado a este propósito, chamado *mixexp*, que pode ser instalado e carregado pelos comandos a seguir:

```
instal.packages("mixexp")
library(mixexp)
```

A função usada para criar o gráfico ternário é chamada *MixturePlot()* e possui os seguintes parâmetros: *MixturePlot(des=MATRIZ_DE DADOS, despts = TRUE, cols = TRUE, mod = n, corner.labs = c("x3",*

"x2", "x1"), x1lab = " ", x2lab = " ", x3lab = " ", color.palette = *heat.colors)*, onde:

des = objeto onde se encontra a planilha dos dados;

despts = TRUE ou FALSE, que diz se os pontos experimentais devem ou não ser mostrados na figura a ser gerada;

cols = TRUE ou FALSE, que diz se as cores para diferenciar os níveis devem ou não ser mostradas na figura a ser gerada;

mod = diz respeito ao modelo para gerar a figura, sendo: 1- modelo linear, 2- modelo quadrático, 4- modelo cúbico especial;

corner.labs = nomes das arestas do triângulo;

xnlab = o nome nas laterais do triângulo, nesse caso definido para não ser mostrado;

color.palette = paleta de cores a ser utilizada na escala de cores.

As informações sublinhadas na linha de comando genérica descrita acima devem ser substituídas de acordo com os seus dados e o modelo que se deseja construir. Como mostrado no exemplo abaixo, para a construção do modelo linear: *MixturePlot(des=dados, despts = TRUE, cols = TRUE, mod = 1, corner.labs = c("x3", "x2", "x1"), x1lab = " ", x2lab = " ", x3lab = " ", color.palette = heat.colors)*.

As superfícies de resposta resultantes serão apresentadas na seção "Plots", ou seja, na seção inferior direita do *layout* do *software*. Superfícies de resposta construídas para os modelos linear, quadrático e cúbico especial são apresentadas na Figura 12.

```
> pureErrorAnova(modelo_linear)
Analysis of Variance Table

Response: y
      Df Sum Sq Mean Sq F value    Pr(>F)
x1      1 12739.2 12739.2  733.5986 6.175e-10 ***
x2      1  4572.5  4572.5  263.3126 5.689e-08 ***
x3      1  1310.6  1310.6   75.4708 1.139e-05 ***
Residuals 17  1477.2    86.9
Lack of fit  8  1321.0   165.1    9.5085 0.001389 **
Pure Error   9   156.3    17.4
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figura 9. Output da função "alr3" para o modelo linear

```
> pureErrorAnova(modelo_quadratico)
Analysis of Variance Table

Response: y
      Df Sum Sq Mean Sq F value    Pr(>F)
x1      1 12739.2 12739.2  733.5986 6.175e-10 ***
x2      1  4572.5  4572.5  263.3126 5.689e-08 ***
x3      1  1310.6  1310.6   75.4708 1.139e-05 ***
x1:x2   1   413.9   413.9   23.8368 0.0008686 ***
x1:x3   1   264.1   264.1   15.2064 0.0036219 **
x2:x3   1   428.1   428.1   24.6522 0.0007751 ***
Residuals 14   371.1    26.5
Lack of fit  5   214.9    43.0    2.4745 0.1124341
Pure Error   9   156.3    17.4
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figura 10. Output da função "alr3" para o modelo quadrático

```
> pureErrorAnova(modelo_cubico)
Analysis of Variance Table

Response: y
      Df Sum Sq Mean Sq F value    Pr(>F)
x1      1 12739.2 12739.2  733.5986 6.175e-10 ***
x2      1  4572.5  4572.5  263.3126 5.689e-08 ***
x3      1  1310.6  1310.6   75.4708 1.139e-05 ***
x1:x2   1   413.9   413.9   23.8368 0.0008686 ***
x1:x3   1   264.1   264.1   15.2064 0.0036219 **
x2:x3   1   428.1   428.1   24.6522 0.0007751 ***
x1:x2:x3 1   114.3   114.3    6.5810 0.0304187 *
Residuals 13   256.9    19.8
Lack of fit  4   100.6    25.1    1.4479 0.2952795
Pure Error   9   156.3    17.4
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figura 11. Output da função "alr3" para o modelo cúbico

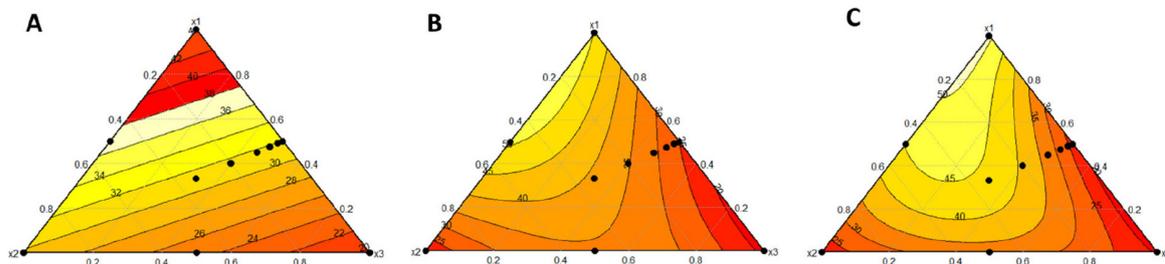


Figura 12. Superfícies de resposta para os modelos (A) linear; (B) quadrático; (C) cúbico especial

Considerando os diferentes elementos presentes na Figura 12, podemos elucidar que a superfície de resposta para os modelos linear, quadrático e cúbico especial é representada por um triângulo equilátero, tendo em vista que qualquer ponto neste sistema pode ser representado pelas coordenadas baricêntricas por meio das variáveis (x_1 , x_2 , x_3), denotadas pela proporção dos três componentes. Cada aresta corresponde a um sistema binário, logo, os pontos demarcados sob os vértices do triângulo representam uma proporção pura de um dos componentes. É interessante salientar que as cores presentes no interior do triângulo são um artifício utilizado para mapear a superfície de resposta, e geralmente, uma escala de cores é usada para indicar o nível de resposta em cada ponto no interior do triângulo. Tomando como referência o modelo cúbico especial apresentado na Figura 12C, como possibilidade de interpretação e análise da superfície, as regiões são delimitadas com cores quentes, na qual o vermelho representa uma resposta no nível baixo, o laranja uma resposta no nível intermediário, e o amarelo uma resposta em um nível alto, tomando como referência os valores da resposta predita pelo modelo estatístico. Este recurso gráfico facilita a compreensão dessa resposta, sendo possível identificar padrões e tendências ao longo do espaço ternário, auxiliando os usuários na tomada de decisão para otimizar a proporção que atenda adequadamente aos objetivos do experimento. Logo, é possível inferir que a região amarela apresenta a proporção de componentes que detém a máxima resistência mecânica para o polímero sintetizado, cujos valores são maiores ou iguais a 50 MPa, tomando como referência a curva de nível mais próxima a esta região, na direção do vértice x_1 , o que corresponde a blendas com predomínio de PVDF e com pouco ou nenhum poliestireno em sua composição. Infelizmente este ajuste ainda está longe de ser satisfatório. Em primeiro lugar, as previsões para três dos quatro ensaios que não pertencem ao planejamento simplex (os pontos pretos na Figura 12C) estão bastante superestimadas, indicando que o modelo não está representando muito bem essa região. Mas o mais preocupante é que os pontos 11 e 12 são ensaios repetidos, e a diferença entre suas respostas é quase a metade da diferença entre os valores extremos de toda a tabela. Como o erro puro é calculado a partir dos ensaios repetidos, essa enorme variação inflaciona MQ_{ep} e termina fazendo com que um modelo problemático pareça bem ajustado.

CONCLUSÕES

A importância e utilidade do planejamento de experimentos em especial, em sistemas de misturas, está cada vez mais evidenciado na literatura, e vem ganhando espaço em outros ambientes, principalmente nas indústrias, tendo em vista os benefícios diante da redução de custos e de tempo em processos envolvendo melhorias. A modelagem de misturas pode ser feita através de pacotes estatísticos pagos. Contudo, um caminho para a popularização e acessibilidade dessa ferramenta pode ser mediado pelo uso de programas livres, como o LibreOffice e RStudio, pois eles apresentam desempenho satisfatório, como demonstrado neste trabalho, além de abrir o leque de habilidades e competências para ampla diversidade de aplicações.

Os tutoriais envolvendo comandos para execução de tarefas em ambos os programas, LibreOffice e RStudio, foram cuidadosamente testados e validados, para que os usuários interessados possam aplicar, no cotidiano, de maneira ampla, amigável e bem-sucedida, os planejamentos de misturas.

MATERIAL SUPLEMENTAR

O conteúdo suplementar referenciado no texto está disponível em <http://quimicanova.s bq.org.br>, na forma de arquivo PDF, com acesso livre.

AGRADECIMENTOS

Os autores agradecem à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), ao Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq (Projetos 303355/2017-4 e 424032/2018-0, 142502/2020-0), ao Instituto Nacional de Ciência e Tecnologia de Bioanalítica - INCTBio (Projetos FAPESP 2014/50867-3 e CNPq 465389/2014-7), à Financiadora de Inovação e Pesquisa - FINEP (Project CT-INFRA 01/2013-Ref 0633/13), à Fundação de Amparo à Pesquisa do Estado de Minas Gerais - FAPEMIG (Processo APQ-00585-21), e à Rede Mineira de Química - RQ-MG (Projeto CEX.RED-00010-14).

REFERÊNCIAS

- Bezerra, M. A.; Lemos, V. A.; Novaes, C. G.; de Jesus, R. M.; Souza Filho, H. R.; Araújo, S. A.; Alves, J. P. S.; *Microchem. J.* **2020**, *152*, 104336. [Crossref]
- Azcarate, S. M.; Pinto, L.; Goicoechea, H. C.; *J. Chemom.* **2020**, *34*, e3246. [Crossref]
- Li, Z.; Lu, D.; Gao, X.; *Journal of Building Engineering* **2021**, *36*, 102101. [Crossref]
- Liu, Z.; Li, M.; Weng, Y.; Wong, T. N.; Tan, M. J.; *Constr. Build. Mater.* **2019**, *198*, 245. [Crossref]
- do Nascimento, M. P.; Marinho, M. V.; de Sousa, R. A.; de Oliveira, M. A. L.; *Anal. Methods* **2023**, *15*, 311. [Crossref]
- Barros Neto, B.; Scarminio, I. S.; Bruns, R. E.; *Como Fazer Experimentos Pesquisa e Desenvolvimento na Ciência e na Indústria*, 2ª ed.; Editora Unicamp: Campinas, 2001.
- LibreOffice, <https://www.libreoffice.org/discover/calc/>, acessado em Novembro 2023.
- Hilário, F.; Castro, J.; Barros, T.; Pereira-Filho, E.; *Quim. Nova* **2021**, *44*, 874. [Crossref]
- Lawson, J.; Willden, C.; *Journal of Statistical Software* **2016**, *72*, 1. [Crossref]
- RStudio, <https://www.rstudio.com/>, acessado em Novembro 2023.
- Cornell, J. A.; *Experiments with Mixtures*, 3rd ed.; John Wiley & Sons: New York, 2002.
- Reis, C.; de Andrade, J. C.; *Quim. Nova* **1996**, *19*, 313. [Link] acessado em Dezembro 2023

13. Piepel, G. F.; Cornell, J. A.; *Journal of Quality Technology* **1994**, *26*, 177. [Crossref]
14. Scheffé, H.; *Journal of the Royal Statistical Society: Series B (Methodological)* **1958**, *20*, 344. [Crossref]
15. McLean, R. A.; Anderson, V. L.; *Technometrics* **1966**, *8*, 447. [Crossref]
16. Steinberg, D. M.; Hunter, W. G.; *Technometrics* **1984**, *26*, 71. [Crossref]
17. Sahin, Y. B.; Demirtaş, E. A.; Burnak, N.; *Pamukkale University Journal of Engineering Sciences* **2016**, *22*, 297. [Crossref]
18. Cornell, J. A.; *A Primer on Experiments with Mixtures*, Illustrated ed.; John Wiley & Sons: New Jersey, 2011.
19. Crosier, R. B.; *Technometrics* **1984**, *26*, 209. [Crossref]
20. Marquardt, D. W.; Snee, R. D.; *Technometrics* **1974**, *16*, 533. [Crossref]