

PLANEJAMENTO DE MISTURAS E VISUALIZAÇÃO DA REGIÃO ÓTIMA COM PLANILHAS NO EXCEL: UM TUTORIAL[†]

Felipe Ferri Hilário^a, Jeyne Pricylla Castro^a, Tomas Edson Barros^b e Edenir Rodrigues Pereira-Filho^{a,b,*} 

^aDepartamento de Química, Universidade Federal de São Carlos, 13565-905 São Carlos – SP, Brasil

^bDepartamento de Matemática, Universidade Federal de São Carlos, 13565-905 São Carlos – SP, Brasil

Recebido em 03/02/2021; aceito em 18/03/2021; publicado na web em 14/04/2021

MIXTURE DESIGN AND OPTIMUM REGION VISUALIZATION WITH SPREADSHEETS AT EXCEL: A TUTORIAL. This tutorial shows how to perform the data treatment for mixture design with two and multiple responses. Two data sets from the literature were used to exemplify the calculations. A template in Excel was proposed to prepare contour plot and response surface to visualize the optimum region of the mixture, which usually is made with commercial computational programs. The desirability function was used in example 2 (optimization of 5 responses simultaneously), and a code also available for download was proposed. Besides, 21 videos were prepared to show all the details for the readers and, it is available on YouTube. With this tutorial, it is possible to learn in a practical fashion how to handle the data from mixture design and, the authors hope to contribute with the researchers in this area.

Keywords: Octave; mixture design; analysis of variance; regression model; template mixtures; partial derivatives.

INTRODUÇÃO

O planejamento de misturas é bastante empregado em otimizações nas quais é fixa a relação entre as variáveis.¹ Um exemplo típico é a otimização da composição da fase móvel em um método de separação cromatográfico.² Nesse tipo de planejamento, os níveis das variáveis não podem ser alterados de forma independente e sempre é mantida uma relação entre as mesmas. A literatura científica apresenta diversos textos excelentes sobre este tipo de planejamento.³⁻⁷ Entretanto, a visualização da superfície de resposta e gráfico de contorno está restrita a alguns programas computacionais pagos e específicos.

Dessa forma, o objetivo deste tutorial é mostrar a aplicação de uma planilha no Microsoft Excel para a visualização do gráfico de contorno e superfície de resposta. Além disso, é empregado o Solver (Microsoft Excel) para o cálculo da melhor condição de trabalho. Os cálculos para obtenção do modelo de regressão podem ser efetuados em programa livre (Octave) e todos os procedimentos para execução dos cálculos já foram descritos no tutorial anterior publicado na Química Nova.⁸ Um dos objetivos deste tutorial é também apresentar uma análise criteriosa e com rigor matemático sobre a identificação da região ótima de um dado modelo.

Portanto, com este tutorial é possível apresentar ao público interessado uma alternativa viável e de fácil acesso para a utilização de planejamento fatorial no contexto do planejamento de misturas. Além da descrição dos conjuntos de dados e cálculos efetuados, foram preparados vídeos explicativos que estão disponíveis no Canal do prof. Edenir Pereira Filho no YouTube (<https://www.youtube.com/c/EdenirPereiraFilho>, ver Playlist 15).

PARTE EXPERIMENTAL

Descrição dos conjuntos de dados empregados

Para exemplificar as ferramentas deste tutorial, foram selecionados dois conjuntos de dados a partir de trabalhos publicados no periódico *Journal of the Brazilian Chemical Society* (JBCS). A utilização

desses estudos, bem como o conjunto de dados neles publicados, foi gentilmente cedida pela Editora da Sociedade Brasileira de Química (PubliSBQ). Tais artigos foram selecionados por utilizarem planejamento de misturas objetivando a resolução de diferentes problemas analíticos.

Exemplo 1

No primeiro exemplo foi utilizado o trabalho publicado por Damasceno *et al.*,⁹ no qual os autores utilizaram um planejamento de misturas do tipo *simplex centroide* para visualizar a relação do chumbo presente em descarte de material eletrônico (placas de circuito impresso) com diferentes tipos de solos. Nesse caso, cada pseudocomponente foi avaliado em 6 níveis. A Tabela 1 mostra em detalhes o planejamento experimental utilizado pelos autores com três variáveis, OM (v_1 - matéria orgânica), IO (v_2 - óxido de ferro) e KA (v_3 - caulinita), além de duas respostas que serão estudadas neste tutorial, F2 (concentração de Pb extraído na fração orgânica) e F3 (concentração de Pb extraído na fração residual). No experimento 1, por exemplo, da Tabela 1 temos uma mistura composta por 80 g (codificado em 1) de matéria orgânica (OM) e 10 g de cada um dos demais componentes: óxido de ferro (IO, codificado em 0) e caulinita (KA, codificado em 0).

Exemplo 2

Para o exemplo 2, foi utilizado o trabalho publicado por Estevão *et al.*¹⁰ e o objetivo foi abordar um estudo com múltiplas respostas. Nesse caso, os autores otimizaram um procedimento de extração utilizando a técnica de microextração líquido-líquido dispersivo em solvente binário para a determinação de pesticidas em amostras de água natural. Na parte experimental foram realizados dois planejamentos de misturas com três componentes. O primeiro planejamento abordou uma faixa muito ampla da composição da mistura dos componentes extratores, sendo necessário um segundo com uma faixa mais restrita. Nesse tutorial abordamos apenas o segundo planejamento.

A mistura foi realizada com dois extratores (diclorometano, DCM e clorofórmio, CLF) e um solvente dispersor (acetonitrila, ACN). O objetivo foi obter a melhor relação entre os três componentes que proporcionasse recuperações adequadas (da ordem de 100%) dos seguintes pesticidas nas amostras de água natural: *Diuron* (DIU),

*e-mail: erpf@ufscar.br

[†]Esse trabalho é dedicado à memória do Prof. Dr. Ronei Jesus Poppi

Tabela 1. Planejamento de misturas *simplex centroid* usado para verificar a influência das variáveis v_1 , v_2 e v_3 nas respostas F2 e F3 (Exemplo 1)

Experimentos	OM (v_1)		IO (v_2)		KA (v_3)		Concentração de Pb (mg kg ⁻¹)	
	Codificado	Real (g)	Codificado	Real (g)	Codificado	Real (g)	Fração Orgânica (F2)	Fração Residual (F3)
1	1	80	0	10	0	10	1931±259	2799±1114
2	0	10	1	80	0	10	2034±208	2818±428
3	0	10	0	10	1	80	138±78	692±154
4	1/2	45	1/2	45	0	10	1521±114	2146±371
5	0	10	1/2	45	1/2	45	1081±255	1691±85
6	1/2	45	0	10	1/2	45	1012±220	1007±233
7	1/3	33,3	1/3	33,3	1/3	33,3	1058±55	1309±98
8	2/3	56,67	1/6	21,66	1/6	21,66	1765±278	1592±157
9	1/6	21,66	2/3	56,67	1/6	21,66	1578±75	1750±88
10	1/6	21,66	1/6	21,66	1/6	54,67	777±189	840±401

teflubenzuron (TFB), *atrazine* (ATZ), *desisopropylatrazine* (DIA) e *desethylatrazine* (DEA). Como se trata de múltiplas respostas (nesse caso, 5), foi abordada neste trabalho a utilização do cálculo da desejabilidade,¹¹ diferente da abordagem original dos autores. A função de desejabilidade foi proposta por Derringer e Suich em 1980¹² e tem como objetivo encontrar uma condição de compromisso onde todas as respostas estejam dentro de uma região aceita como desejável. Dessa forma, as respostas originais são codificadas entre 0 (condição indesejável) e 1 (condição desejável). Para esse cálculo, foi criada uma rotina computacional chamada “mydes2” (ver detalhes no Material Suplementar).

Rotinas computacionais utilizadas

Uma rotina computacional chamada “mydes2” foi preparada para a realização do cálculo de desejabilidade no programa livre Octave, independentemente do alvo de cada resposta (maximizar ou minimizar). Como *input* é necessário fornecer dois vetores/matrices: um deles é o conjunto de dados (respostas) que devem ser codificadas (matriz **X**) e o outro contém as condições (matriz **Cond**) para a realização do cálculo. Para **Cond**, é necessário fornecer o valor mínimo da resposta (posicionada na primeira linha), o valor máximo (segunda linha), o alvo (terceira linha) e o peso (quarta linha). O alvo é o objetivo de cada resposta. No caso do exemplo 2, a resposta monitorada foi a recuperação dos analitos, sendo assim o alvo foi igual a 100%. E o peso é quanto uma resposta influenciará no modelo final. Assim, pode-se colocar diferentes pesos para as respostas. Respostas que apresentaram modelos individuais melhor ajustados¹³ podem receber um peso maior que aquelas respostas que apresentaram modelos com falta de ajuste. Como *output*, a rotina irá fornecer dois resultados: a desejabilidade individual (*di*) de cada resposta e a desejabilidade global (*D*), ou seja, a média geométrica de todas as desejabilidades individuais. Com isso, pode-se calcular apenas um modelo de regressão que englobe todas as respostas utilizando a *D*. A rotina está disponível no Material Suplementar e, para os cálculos, utilizaram-se as equações mostradas na Figura 1.

RESULTADOS E DISCUSSÃO

Exemplo 1

No primeiro exemplo estudado os autores monitoraram como as diferentes proporções de matéria orgânica (v_1), óxido de ferro (v_2) e caulinita (v_3) influenciam na absorção de Pb por uma amostra de solo. As concentrações de Pb utilizadas são aquelas que podem ser encontradas devido ao descarte inadequado de material eletrônico.

$$di = \begin{cases} 0 & \text{se } y < L \\ \left(\frac{y-L}{T-L}\right)^s & \text{se } L \leq y \leq T \\ 1 & \text{se } y > T \end{cases} \quad di = \begin{cases} 1 & \text{se } y < T \\ \left(\frac{U-y}{U-T}\right)^t & \text{se } T \leq y \leq U \\ 0 & \text{se } y > U \end{cases}$$

a) *di* para maximizar

b) *di* para minimizar

y = resposta; *L* = mínimo; *T* = alvo; *U* = máximo; *s* e *t* = pesos

Figura 1. Equações para o cálculo da desejabilidade quando o objetivo é (a) maximizar e (b) minimizar

Segundo os autores, foram realizadas 3 réplicas em cada nível experimental, entretanto, são mostrados apenas o valor médio e desvio padrão obtido.

Para avaliar como os pseudocomponentes interferiram nas respostas avaliadas utilizou-se a rotina *regression2*, a qual foi proposta por Pereira e Pereira-Filho⁸ e tem a finalidade de calcular um modelo de regressão a partir de 4 dados de entrada (*inputs*): as matrices **X** e **Y**, a soma quadrática do erro puro (*pure error sum of squares* - SSPE) e seu respectivo grau de liberdade (*degree of freedom* - DF). A linha de comando para a utilização da rotina *regression2* é:

$$[\text{Mod_par,Coef,Pred}]=\text{regression2}(\text{X,Y,SSPE,DF});$$

Na parte direita da linha de comando pode-se observar os *inputs*. A matriz **X** dá origem aos coeficientes de regressão (Coef) a serem calculados e possui *m* linhas e *n* colunas, onde as linhas correspondem aos experimentos e as colunas aos coeficientes de regressão que serão calculados, essa matriz pode ser vista em detalhes na Tabela 2. Os coeficientes calculados para ambas as respostas avaliadas foram: $b_1, b_2, b_3, b_{12}, b_{13}, b_{23}, b_{123}$. Os três primeiros coeficientes, b_1, b_2 e b_3 , são obtidos a partir dos níveis codificados das variáveis v_1, v_2 e v_3 , mostrados na Tabela 1, o coeficiente b_{12} é obtido pela multiplicação da coluna correspondente ao coeficiente b_1 pela coluna correspondente ao coeficiente b_2 . Os demais coeficientes são obtidos de forma análoga. A matriz **Y** corresponde aos resultados obtidos em cada experimento realizado, dessa forma será composta por *m* linhas e *n* colunas, onde as linhas correspondem aos experimentos e as colunas, as respostas monitoradas (F2 e F3, Tabela 1). SSPE corresponde a soma quadrática do erro puro obtida a partir das réplicas, porém, nesse caso não foi possível o cálculo desse parâmetro devido à ausência desses valores, sendo considerado 0. Na utilização da rotina *regression2* é calculado um modelo para cada resposta. Dessa forma, recomenda-se o desmembramento da matriz **Y** em *n* vetores (as *n* respostas avaliadas).

A parte esquerda da linha de comando corresponde aos *outputs* (saídas) obtidos após o tratamento dos dados pela rotina *regression2*. *Mod_par* corresponde a tabela ANOVA (*Analysis of Variance*)

Tabela 2. Matriz **X** utilizada para os cálculos dos modelos de regressão para as respostas F2 e F3 (Exemplo 1)

Experimentos	Matriz X com os coeficientes utilizados na proposição do modelo						
	b_1	b_2	b_3	b_{12}	b_{13}	b_{23}	b_{123}
1	1,00	0,00	0,00	0,00	0,00	0,00	0,00
2	0,00	1,00	0,00	0,00	0,00	0,00	0,00
3	0,00	0,00	1,00	0,00	0,00	0,00	0,00
4	0,50	0,50	0,00	0,25	0,00	0,00	0,00
5	0,00	0,50	0,50	0,00	0,00	0,25	0,00
6	0,50	0,00	0,50	0,00	0,25	0,00	0,00
7	0,33	0,33	0,33	0,11	0,11	0,11	0,03
8	0,67	0,16	0,16	0,11	0,11	0,03	0,02
9	0,16	0,67	0,16	0,11	0,03	0,11	0,02
10	0,16	0,16	0,67	0,03	0,11	0,11	0,02

considerando 95% de confiança. Coef aos coeficientes obtidos com seus respectivos intervalos de confiança e Pred corresponde aos valores previstos para cada experimento. Pereira e Pereira-Filho⁸ fornecem uma explicação detalhada da rotina utilizada.

O primeiro passo foi calcular um modelo de regressão para cada resposta a fim de verificar quais coeficientes de regressão são significativos, isso é, valores maiores que seus respectivos intervalos de confiança. O Material Suplementar mostra as tabelas Anova para as respostas F2 (Tabela 1S) e F3 (Tabela 2S). Além disso, as informações dos modelos são também mostradas na forma de figuras (Figuras 1S e 2S). Para a resposta F2 apenas os coeficientes b_1 e b_2 foram significativos (Figura 1S), já para a resposta F3, b_1 , b_2 , b_3 , b_{12} e b_{13} foram válidos (Figura 2S). Todas as avaliações foram feitas ao nível de confiança de 95%.

A rotina regression2 foi aplicada novamente para ambas as respostas utilizando somente os coeficientes válidos. Mesmo com a exclusão dos coeficientes não significativos para cada uma das respostas, não foram identificadas réplicas, isso é, mais de um

experimento com as mesmas condições, não sendo possível o cálculo de SSPE e DF.

A Figura 2, a Tabela 3 e a Equação 1 mostram os resultados obtidos após o refinamento do modelo para a resposta F2. E a Figura 3, Tabela 4 e Equação 2 descrevem os resultados para a resposta F3.

$$F2 = 1874_{\pm 328}v_1 + 1903_{\pm 328}v_2 \quad (1)$$

$$F3 = 2804_{\pm 396}v_1 + 2726_{\pm 535}v_2 + 695_{\pm 353}v_3 - 3133_{\pm 1831}v_1v_2 - 3340_{\pm 1831}v_1v_3 \quad (2)$$

Nas Equação 1 e 2, v_1 , v_2 e v_3 são relativos à matéria orgânica, óxido de ferro e caulinita, respectivamente. Ambas as equações descrevem uma superfície de resposta com as três variáveis interrelacionadas. O ponto “ótimo” da superfície descrita pela equação representa as proporções de v_1 , v_2 e v_3 em que as respostas encontram seus valores “ótimos”, ou seja, região onde se observa maior fração de massa (concentração) de Pb nas frações orgânica (F2) e residual

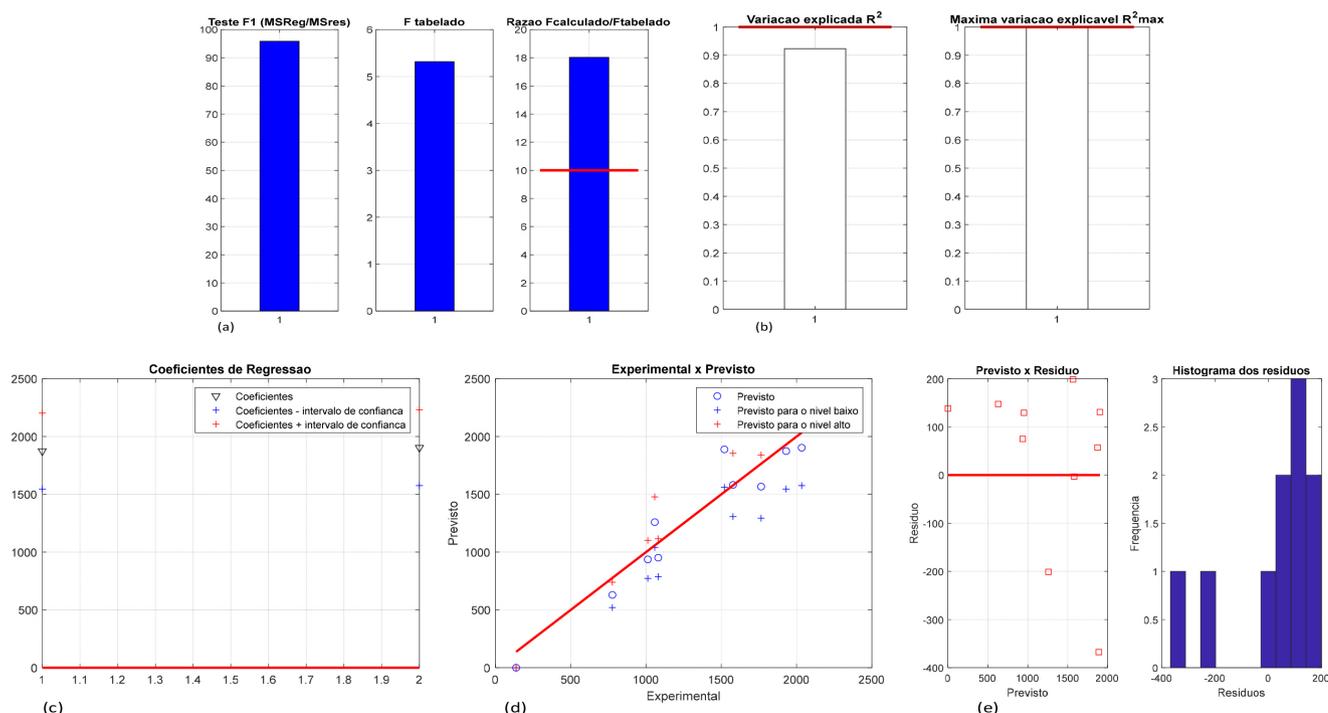


Figura 2. Teste F para a média quadrática de regressão (MQR) e média quadrática de resíduo (MQres) (a), variação explicada (R^2) e explicável (R^2_{max}) (b), coeficientes de regressão (c), gráfico dos valores experimental versus previsto (d), previsto versus o resíduo e histograma dos resíduos (e), para os cálculos da resposta F2 após o refinamento do modelo

Tabela 3. Tabela Anova para o modelo recalculado da resposta F2 com os coeficientes válidos (Exemplo 1)

	Soma quadrática (SQ)	Graus de liberdade	Média quadrática (MQ)	F calculado	F tabelado
Regressão (Reg)	3,58.10 ⁶	1	3,58.10 ⁶	97,87	5,32
Resíduo (res)	2,98.10 ⁵	8	3,73.10 ⁴		
Total (T)	3,88.10 ⁶	9	4,31.10 ⁵		
Erro puro (EP)	0	0	0		
Falta de ajuste (Faj)	2,98.10 ⁵	8	3,73.10 ⁴		

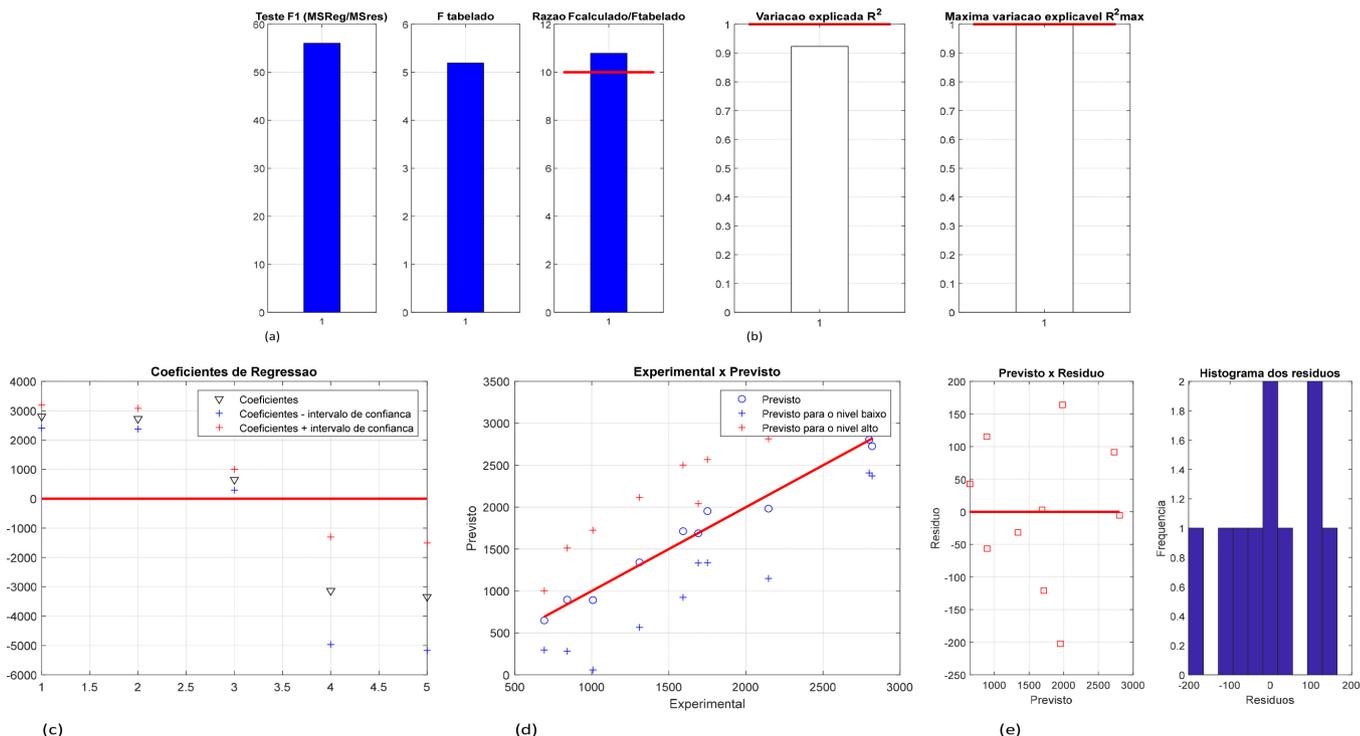


Figura 3. Teste F para MQR e MQres (a), variação explicada (R^2) e explicável (R^2_{max}) (b), coeficientes de regressão (c), gráfico dos valores experimental versus previsto (d), previsto versus o resíduo e histograma dos resíduos (e), para os cálculos da resposta F3 após o refinamento do modelo

Tabela 4. Tabela Anova para o modelo recalculado da resposta F3 com os coeficientes válidos (Exemplo 1)

	Soma quadrática (SQ)	Graus de liberdade	Média quadrática (MQ)	F calculado	F tabelado
Regressão (Reg)	4,94.10 ⁶	4	1,23.10 ⁶	56,03	5,19
Resíduo (res)	1,10.10 ⁵	5	2,20.10 ⁴		
Total (T)	5,05.10 ⁶	9	5,61.10 ⁵		
Erro puro (EP)	0	0	0		
Falta de ajuste (Faj)	1,10.10 ⁵	5	2,20.10 ⁴		

(F3), respectivamente. Os modelos obtidos foram analisados em um template preparado no Excel, que tem como finalidade a obtenção de um gráfico de contorno a partir dos valores dos coeficientes obtidos em um modelo de regressão. As Figuras 4a e 4b mostram os gráficos de contorno obtidos para F2 e F3, respectivamente. Analisando os gráficos obtidos é possível observar as proporções das pseudocomponentes que resultam em maior teor de Pb, em vermelho, e as proporções que resultam em menor teor de Pb, em verde.

Além do template descrito também foi utilizado a ferramenta do Solver no Excel para avaliar os modelos obtidos. Uma descrição detalhada deste recurso e sua utilização podem ser encontrados no tutorial anterior⁸ e na Playlist de número 7 que descreve o uso do Solver (https://www.youtube.com/playlist?list=PL4CuftF4I_fb7LRMzFBmw59FrigmAb0hy). As proporções que resultaram em maior teor de Pb na F2 do solo são: 10% de matéria orgânica (v_1),

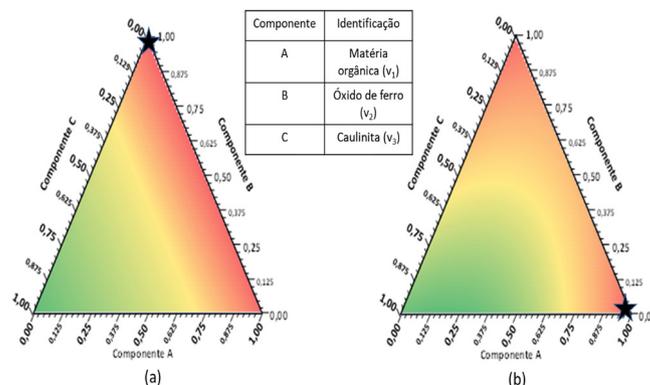


Figura 4. Gráficos de contorno obtidos para o modelo final da resposta F2 (a), e para o modelo final da resposta F3 (b)

80% de óxido de ferro (v_2) e 10% de caulinita (v_3). No caso da F3, a melhor condição exige que a v_1 esteja no nível máximo.

Para melhor entendimento do processo de tratamento de dados descrito neste trabalho foi elaborada uma lista de vídeos explicativos de livre acesso publicados no YouTube (https://www.youtube.com/playlist?list=PL4CutfF4l_fDBW0qIJ1gXZneN1Krgw5z_).

Exemplo 2

Na Tabela 5, tem-se o planejamento de misturas utilizado pelos autores com três variáveis (CLF, DCM e ACN), onde as proporções das condições escolhidas variaram de 0 a 1 quando convertidas em pseudocomponentes. As codificações de cada uma das variáveis são mostradas, sendo que a soma entre elas é 1. As respostas monitoradas são apresentadas na forma de recuperação dos cinco pesticidas determinados.

O primeiro passo foi calcular a desejabilidade individual para cada uma das respostas utilizando a rotina “mydes2” e construir a matriz **X** com os coeficientes cúbicos (b_1 , b_2 , b_3 , b_{12} , b_{13} , b_{23} e b_{123}) que se deseja calcular. Neste caso, o alvo para o cálculo da desejabilidade foi 100% por se tratar de recuperações de cada um dos pesticidas. Fez-se um modelo de regressão utilizando a rotina regression2 para cada uma das desejabilidades individuais com o objetivo de atribuir diferentes pesos a cada uma das respostas

e observar quais são mais influenciadas pelas variáveis testadas.⁸ Como no artigo original não foram fornecidos os valores das réplicas, não foi possível calcular o valor da soma quadrática do erro puro. Os pesos foram atribuídos de acordo com o resultado da razão entre a média quadrática de regressão (MQR) e a média quadrática do resíduo (MQres). Quanto maior essa razão, melhor é a regressão do modelo. De acordo com os resultados, DIA e ATZ obtiveram razões na faixa de 2,2; DEA e DIU na faixa de 1,7 e TFB na faixa de 0,9. Portanto, três diferentes pesos (s , Figura 1) foram atribuídos: 3 (DIA e ATZ), 2 (DEA e DIU) e 1 (TFB).

O segundo passo foi recalculando as desejabilidades individuais com os devidos pesos para cada uma das respostas e assim obter a desejabilidade global (D), a qual foi utilizada para calcular um modelo de regressão englobando todas as respostas simultaneamente. A matriz **X** (coeficientes cúbicos) e o vetor **y** (desejabilidade global, D) são mostrados na Tabela 6.

O modelo de regressão é calculado pela rotina “regression2” disponibilizada em Pereira e Pereira-Filho,⁸ a partir do qual é possível calcular e avaliar a Tabela Anova. Após o cálculo, apenas dois coeficientes foram significativos: b_{12} e b_{13} . Assim, o modelo foi recalculado apenas com os coeficientes válidos, onde quatro réplicas foram identificadas permitindo obter a soma quadrática do erro puro. Na Tabela 3S, tem-se a tabela Anova para o modelo recalculado e na Figura 3S (ver ambos no Material Suplementar) tem-se os valores

Tabela 5. Planejamento de misturas com três variáveis e cinco respostas monitoradas (Exemplo 2)

Experimentos	CLF		DCM		ACN		Soma	Respostas (Recuperação, %)				
	Real	Codificada	Real	Cod	Real	Cod		DIA	DEA	ATZ	DIU	TFB
1	70	1,000	0	0,000	30	0,000	1	44,6	69,6	81,1	85,0	56,7
2	50	0,500	20	0,500	30	0,000	1	46,6	71,2	87,6	90,7	60,6
3	30	0,000	40	1,000	30	0,000	1	43,5	65,6	78,5	80,8	53,4
4	30	0,000	20	0,500	50	0,500	1	46,1	69,3	86,8	89,8	58,1
5	30	0,000	0	0,000	70	1,000	1	39,8	62,5	86,1	90,7	57,3
6	50	0,500	0	0,000	50	0,500	1	46,3	70,4	91,1	93,4	61,3
7	60	0,750	5	0,125	35	0,125	1	47,1	72,3	88,3	90,3	63,8
8	35	0,125	30	0,750	35	0,125	1	44,9	68,4	83,1	85,2	61,6
9	35	0,125	5	0,125	60	0,750	1	42,7	65,9	86,2	88,4	58,2
10	50	0,500	10	0,250	40	0,250	1	48,5	74,3	91,9	95,8	61,1
11	40	0,250	20	0,500	40	0,250	1	41,9	62,8	77,7	80,3	54,7
12	40	0,250	10	0,250	50	0,500	1	48,8	74,0	92,7	97,2	61,1
13	35	0,125	17,5	0,438	47,5	0,438	1	45,4	69,4	87,1	90,6	58,8
14	47,5	0,438	17,5	0,438	35	0,125	1	47,2	72,0	89,2	91,2	73,9
15	47,5	0,438	5	0,125	47,5	0,438	1	48,1	73,4	92,3	96,1	74,1

Tabela 6. Matriz **X** e vetor **y** para a realização do modelo de regressão global (Exemplo 2)

Experimentos	Matriz X (coeficientes cúbicos)								Vetor y D global
	b_1	b_2	b_3	b_{12}	b_{13}	b_{23}	b_{123}		
1	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	
2	0,50	0,50	0,00	0,25	0,00	0,00	0,00	0,05	
3	0,00	1,00	0,00	0,00	0,00	0,00	0,00	0,00	
4	0,00	0,50	0,50	0,00	0,00	0,25	0,00	0,04	
5	0,00	0,00	1,00	0,00	0,00	0,00	0,00	0,00	
6	0,50	0,00	0,50	0,00	0,25	0,00	0,00	0,06	
7	0,75	0,13	0,13	0,09	0,09	0,02	0,01	0,06	
8	0,13	0,75	0,13	0,09	0,02	0,09	0,01	0,02	
9	0,13	0,13	0,75	0,02	0,09	0,09	0,01	0,02	
10	0,50	0,25	0,25	0,13	0,13	0,06	0,03	0,10	
11	0,25	0,50	0,25	0,13	0,06	0,13	0,03	0,00	
12	0,25	0,25	0,50	0,06	0,13	0,13	0,03	0,10	
13	0,13	0,44	0,44	0,05	0,05	0,19	0,02	0,04	
14	0,44	0,44	0,13	0,19	0,05	0,05	0,02	0,07	
15	0,44	0,13	0,44	0,05	0,19	0,05	0,02	0,11	

da razão do teste $F_{\text{calculado}}$ e F_{tabelado} para a falta de ajuste (a) e para a regressão (b), além da variância explicada do modelo (c).

Ao analisar o resultado da razão do primeiro teste $F_{\text{calculado}}$ ($MQ_{\text{Faj}}/MQ_{\text{Ep}}$) e F_{tabelado} , tem-se que o $F_{\text{calculado}}$ é menor que o F_{tabelado} , em que os dois se confundem, portanto, o modelo não apresenta falta de ajuste (razão menor que 1, Figura 3Sa). Além da falta de ajuste, é necessário verificar a regressão do modelo, sendo que, nesse caso, a razão entre o $F_{\text{calculado}}$ ($MQ_{\text{Reg}}/MQ_{\text{res}}$) e o F_{tabelado} está na faixa de 4,5 (Figura 3Sb), ou seja, a regressão é maior que o resíduo. Esse aspecto reflete que o modelo está adequado, porém, a situação ideal é que a razão dos valores de F fosse acima de dez.¹⁴ Os parâmetros utilizados para os cálculos do teste F são encontrados na Tabela Anova (Tabela 3S). A variância explicada do modelo está na faixa de 0,6 (Figura 3Sc), o que pode afirmar que o modelo não é excelente, mas por outro lado é possível utiliza-lo para identificar uma condição de compromisso que poderá apresentar valores de recuperação adequados para todos os analitos. Assim, o modelo final foi:

$$D = 0,22v_1v_2 + 0,38v_1v_3 \quad (3)$$

em que v_1 , v_2 e v_3 são CLF, DCM e ACN, respectivamente. D é a resposta (no caso, a deseabilidade global). Nota-se que apenas os coeficientes quadráticos foram significativos.

Na Figura 5, tem-se o gráfico de contorno com os coeficientes significativos, no qual observa-se uma região de ótimo na coloração vermelha da figura, resultando em 0,5 das variáveis 1 (CLF) e 3 (ACN) e 0 da variável 2 (DCM), ou seja, uma mistura “ideal” seria 50% de CLF e 50% de ACN, e 0% de DCM. A condição identificada pelos autores do trabalho foi: 47,5% de CLF e 47,5% de ACN, e 5% de DCM, ou seja, mesmo com diferentes abordagens para o tratamento de dados, os resultados foram praticamente os mesmos. O gráfico foi feito com o template de misturas no Excel disponibilizado neste tutorial.

Na Tabela 7, tem-se o resultado das respostas previstas para cada um dos pesticidas, onde empregou-se o modelo global com os coeficientes b_{12} e b_{13} . Assim, utilizou-se os valores desses coeficientes (b_{12} e b_{13}) dos modelos individuais calculados no início do tratamento dos dados para encontrar a d_i (desejabilidade individual) de cada um dos pesticidas em todos os níveis (coeficiente \pm intervalo de confiança) e consequentemente, a resposta prevista. Para a d_i prevista, utilizou-se

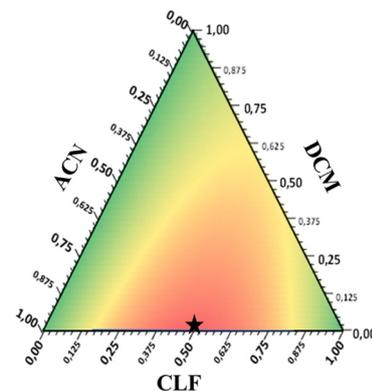


Figura 5. Gráfico de contorno do modelo cúbico com dois coeficientes significativos (Equação 3)

a Equação 3 com os devidos coeficientes significativos para cada nível e cada resposta, como pode ser observado no exemplo (Equação 4) para o pesticida DIA:

$$d_{\text{DIA}} = 0,15 \times 0,5 \times 0 + 0,35 \times 0,5 \times 0,5 = 0,09 \quad (4)$$

Lembrando que $v_1 = 0,5$; $v_2 = 0$ e $v_3 = 0,5$. A recuperação prevista (y) foi calculada de acordo com a equação da Figura 1a. Nesse caso, tem-se a d_i prevista e o novo y será a resposta prevista em termos de recuperação. Ao observar a resposta do artigo (Tabela 7), tem-se valores muito semelhantes com os resultados obtidos no maior nível (coeficiente mais o intervalo de confiança). Vale ressaltar que ambos os trabalhos utilizaram abordagens diferentes e com isso os resultados podem ser considerados equivalentes.

Encontrando a região de ótimo usando ferramentas de cálculo diferencial

A identificação da região ótima de trabalho de determinado modelo não é uma tarefa fácil e pode trazer dificuldades consideráveis de acordo com os coeficientes válidos do modelo. Para o caso da Equação 1, por exemplo, a região em que há a máxima retenção de Pb está na aresta em que a variável 2 (óxido de ferro) é máxima (1,0). Nesse ponto, a fração de massa prevista para o Pb é de 1903 mg kg⁻¹.

Tabela 7. Respostas previstas para cada um dos pesticidas, tendo os coeficientes \pm intervalo de confiança e a resposta obtida pelos autores do Exemplo 2

Coefficientes	DIA	DEA	ATZ	DIU	TFB
Respostas para: coeficientes menos o intervalo de confiança					
b_{12}	-0,01	-0,07	0,56	0,38	0,14
b_{13}	0,19	0,25	0,85	0,53	0,10
d_i prevista	0,05	0,06	0,21	0,13	0,03
Recuperação prevista	42,65	64,87	82,46	82,91	54,60
Respostas para: coeficientes					
b_{12}	0,15	0,34	1,36	1,41	0,72
b_{13}	0,35	0,66	1,65	1,56	0,69
d_i prevista	0,09	0,16	0,41	0,39	0,17
Recuperação prevista	45,03	68,64	86,92	88,00	61,40
Respostas para: coeficientes mais o intervalo de confiança					
b_{12}	0,31	0,74	2,16	2,44	1,31
b_{13}	0,51	1,06	2,45	2,60	1,27
d_i prevista	0,13	0,26	0,61	0,65	0,32
Recuperação prevista	47,41	72,42	91,38	93,09	68,21
Resposta obtida pelos autores do artigo	49,40	73,20	95,10	94,70	74,90

Essa constatação corrobora com a aplicação do Solver do Microsoft Excel®.

Para o caso do modelo obtido para a resposta F3 (Equação 2), podemos calcular as derivadas parciais em função das variáveis v_1 e v_2 e substituir a v_3 por $1-v_1-v_2$:

$$f_3(v_1, v_2) = F_3(v_1, v_2, 1-v_1-v_2) = 2804v_1 + 2726v_2 + 695(1-v_1-v_2) - 3133v_1v_2 - 3340(1-v_1-v_2) \quad (5)$$

As derivadas parciais são:

$$\partial f_3 / \partial v_1(v_1, v_2) = 6680v_1 + 207v_2 - 1231 \quad (6)$$

$$\partial f_3 / \partial v_2(v_1, v_2) = 207v_1 + 2031 \quad (7)$$

(v_1, v_2) é o ponto crítico de $f_3 \Leftrightarrow \partial f_3 / \partial v_1(v_1, v_2) = 0$ e $\partial f_3 / \partial v_2(v_1, v_2) = 0$.

$$\Leftrightarrow v_2 = \frac{1231 - 6680v_1}{207} \text{ e } v_1 = \frac{-2031}{207} \sim -9,81 \quad (8)$$

$$\Leftrightarrow v_1 = -9,81 \text{ e } v_2 = 322,57 \Leftrightarrow (v_1, v_2) = (-9,81; 322,57)$$

Entretanto, esse ponto está fora do domínio $T = \{(v_1, v_2): v_1 \geq 0, v_2 \geq 0 \text{ e } v_1 + v_2 \leq 1\}$ da função f_3 . Portanto, novamente aqui, f_3 não possui pontos críticos no interior de T.

O domínio de F_3 é a região plana delimitada pelo triângulo equilátero com vértices (1,0,0), (0,1,0) e (0,0,1), mostrado na Figura 4S (Material Suplementar).

O domínio T de f_3 é a região plana delimitada pelo triângulo retângulo com vértices (0,0), (1,0) e (0,1), como mostra a Figura 5S (Material Suplementar).

Notamos que o bordo de T é dado pela união dos segmentos a , b e c (dados em vermelho, azul e verde, respectivamente, ver Figura 5S).

Precisamos analisar os candidatos a ponto de máximo de f_3 no bordo de T.

1) Em $a = \{(v_1, v_2) \in T: v_1 \geq 0, v_2 \geq 0 \text{ e } v_1 + v_2 = 1\}$

Nesse caso, analisamos os pontos de máximo da função $g_1: [0,1] \rightarrow [0,1]$, definida por $g_1(v_1) = f_3(v_1, 1-v_1) = 3133v_1^2 - 3055v_1 + 2726$, a qual tem um único ponto crítico $v_1 = 235/482 \sim 0,49 \in (0,1)$. Assim, nesse caso, os candidatos a ponto de máximo de F_3 são, (0,49, 0,51, 0), (0,1,0) e (1,0,0).

2) Em $b = \{(0, v_2) \in T: 0 \leq v_2 \leq 1\}$

Nesse caso, analisamos os pontos de máximo da função $g_2: [0,1] \rightarrow [0,1]$, definida por $g_2(v_2) = f_3(0, v_2) = 2031v_2 + 695$, a qual não possui pontos críticos. Logo, nesse caso, os candidatos a ponto de máximo de F_3 são (0,0,1) e (0,1,0).

3) Em $c = \{(v_1, 0) \in T: 0 \leq v_1 \leq 1\}$

Nesse caso, analisamos os pontos de máximo da função $g_3: [0,1] \rightarrow [0,1]$, definida por $g_3(v_1) = f_3(v_1, 0) = 3340v_1^2 - 1231v_1 + 695$, a qual tem um único ponto crítico $x = 1231/6680 \sim 0,18 \in (0,1)$. Assim, nesse caso, os candidatos a ponto de máximo de F_3 são, (0,18, 0,82), (0,0,1) e (1,0,0).

Agora testamos todos os candidatos:

$$F_3(0,49, 0,51, 0) = 1981$$

$$F_3(0,18, 0, 0,82) = 582$$

$$F_3(1, 0, 0) = 2804$$

$$F_3(0, 1, 0) = 2726$$

$$F_3(0, 0, 1) = 695$$

Logo, o ponto de máximo procurado é (1,0,0), ou seja, a variável 1 ($v_1 =$ Matéria Orgânica) no nível alto (1).

Para o caso do exemplo 2, no qual o modelo obtido é aquele

mostrado na Equação 3, precisamos determinar o ponto de máximo absoluto da função

$$D(v_1, v_2, v_3) = 0,22v_1v_2 + 0,38v_1v_3 \text{ sendo que } v_1 + v_2 + v_3 = 1.$$

Chamando $v_1 = x$, $v_2 = y$ e $v_3 = z$ temos,

$$D(x, y, z) = 0,22xy + 0,38xz \text{ e } x+y+z = 1$$

Isso é equivalente à:

$$d(x, y) = D(x, y, 1-x-y) = -0,38x^2 - 0,16xy + 0,38x$$

com $x \geq 0, y \geq 0$ e $x+y \leq 1$

O ponto de máximo da função d ocorre entre os pontos críticos de d no interior de $T = \{(x, y): x \geq 0, y \geq 0 \text{ e } x+y \leq 1\}$ e os pontos críticos do bordo de T, ou seja, $b(T) = \{(x, y): x = 0 \text{ ou } y = 0 \text{ ou } x+y = 1\}$.

• Pontos críticos no interior de T

(x, y) é ponto crítico de $d \Leftrightarrow \partial d / \partial x(x, y) = -0,76x - 0,16y + 0,38 = 0$ e

$$\partial d / \partial y(x, y) = -0,16x = 0$$

$$\Leftrightarrow (x, y) = (0, 2,38)$$

Como o ponto (0, 2,38) não pertence ao conjunto T (e, portanto, não pertence ao interior desse conjunto), temos que d não possui candidatos a ponto de máximo absoluto no interior de T.

• Pontos críticos no $b(T)$

1) $x = 0$

Nesse caso, analisamos os pontos de máximo da função $f_1: [0,1] \rightarrow [0,1]$, definida por $f_1(y) = d(0, y) = 0$. Como essa é uma função constante, todos os pontos (0, y) são candidatos a ponto de máximo de d , ou seja, todos os pontos (0, y, 1-y) com $0 \leq y \leq 1$ são candidatos a ponto de máximo de D.

2) $y = 0$

Nesse caso, analisamos os pontos de máximo da função $f_2: [0,1] \rightarrow [0,1]$, definida por $f_2(x) = d(x, 0) = 0,38(x-x^2)$, a qual tem um único ponto crítico $x = 1/2 \in (0,1)$. Assim, nesse caso, os candidatos a ponto de máximo de D são, (1/2, 0, 1/2), (0,0,1) e (1,0,0).

3) $x+y = 1$

Nesse caso, analisamos os pontos de máximo da função $f_3: [0,1] \rightarrow [0,1]$, definida por $f_3(x) = d(x, 1-x) = 0,22(x-x^2)$, a qual tem um único ponto crítico $x = 1/2 \in (0,1)$. Assim, nesse caso, os candidatos a ponto de máximo de D são (1/2, 1/2, 0), (0,1,0) e (1,0,0).

Agora testamos todos os candidatos:

$$D(0, y, 1-y) = D(0, 0, 1) = D(0, 1, 0) = 0$$

$$D(1/2, 0, 1/2) = 0,095$$

$$D(1/2, 1/2, 0) = 0,055$$

$$D(1,0,0) = 0$$

Logo, o ponto de máximo procurado é (1/2, 0, 1/2). Assim, as variáveis v_1 (CLF) e v_3 (ACN) devem ser 50% de cada.

CONCLUSÕES

Com este tutorial é possível aprender de forma didática como realizar o tratamento de dados de um planejamento de misturas para duas e múltiplas respostas. Também foi elaborado um *template* no Excel para a construção de gráfico de contorno e superfície de resposta com os resultados dos modelos de regressão de misturas, os quais são disponíveis apenas em programas computacionais pagos e restritos. O uso do Solver também foi exemplificado neste tutorial, sendo uma ferramenta gratuita que pode ser utilizada no

Excel. Além disso, uma rotina computacional foi proposta para o cálculo das desejabilidades.

Tabela 8. Lista dos vídeos preparados com os dois exemplos ilustrados neste tutorial

ID do vídeo	Descrição
Vídeo 1	Introdução + Planejamento de misturas com um exemplo didático (preparando um suco)
Vídeo 2	Exemplo didático – Preparo da matriz de dados
Vídeo 3	Exemplo didático – Obtendo um modelo empírico
Vídeo 4	Exemplo didático – Refinando o modelo
Vídeo 5	Exemplo didático – Visualizando a melhor condição de trabalho
Vídeo 6	Exemplo 1 (duas respostas, F2 e F3) – Preparo da matriz de dados
Vídeo 7	Exemplo 1 – Cálculo do modelo de regressão para F2
Vídeo 8	Exemplo 1 – Refinando o modelo de regressão para F2
Vídeo 9	Exemplo 1 – Obtenção do gráfico de contorno para F2
Vídeo 10	Exemplo 1 – Uso do solver para obter a condição “ótima” para F2
Vídeo 11	Exemplo 1 – Cálculo do modelo de regressão para F3
Vídeo 12	Exemplo 1 – Refinando o modelo de regressão para F3
Vídeo 13	Exemplo 1 – Obtenção do gráfico de contorno para F3
Vídeo 14	Exemplo 1 – Obtenção do gráfico de contorno para F3 e uso do Solver
Vídeo 15	Exemplo 2 (múltiplas respostas) – Preparo da matriz de dados
Vídeo 16	Exemplo 2 – Codificando as respostas utilizando desejabilidade e construção da matriz X
Vídeo 17	Exemplo 2 – Cálculo do modelo de regressão para a primeira resposta codificada
Vídeo 18	Exemplo 2 – Cálculo dos modelos de regressão para as demais respostas codificadas
Vídeo 19	Exemplo 2 – Obtenção do peso das respostas codificadas, cálculo da desejabilidade com os pesos e obtenção do modelo de regressão global (D global)
Vídeo 20	Exemplo 2 – Refinando o modelo de regressão global e obtenção do gráfico de contorno com o template de misturas no Excel
Vídeo 21	Exemplo 2 – Cálculos da d_i e da recuperação prevista para cada resposta. Comparação dos resultados com o artigo de referência

A Tabela 8 mostra uma lista com os vídeos elaborados para a construção deste tutorial. Os vídeos são curtos e mostram o passo a passo até a resposta final. No link a seguir é possível acessar a Playlist 15, criada no canal do professor Edenir Pereira Filho no YouTube: https://www.youtube.com/playlist?list=PL4CufF4l_fDBW0qJJ1gXZneN1Krgw5z_

As rotinas (“regression2”, “template_misturas” e “mydes2”) utilizadas neste trabalho estão disponíveis para download em: https://www.dropbox.com/sh/aiuvutcknbch5te/AAC-6Bef2WUsJZ-LJP2imh_Ea?dl=0

MATERIAL SUPLEMENTAR

O Material Suplementar apresenta os seguintes itens:

- Rotina computacional “mydes2”
- Tabela 1S. Tabela Anova para o modelo calculado da resposta F2 (Exemplo 1).
- Tabela 2S. Tabela Anova para o modelo calculado da resposta F3 (Exemplo 1).
- Tabela 3S. Tabela Anova do modelo calculado para a resposta D (Exemplo 2).
- Figura 1S. Teste F para MQR e MQR (a), teste F para a MQfaj e MQEP (b) variação explicada (R^2) e explicável (R^2_{max}) (c),

coeficientes de regressão (d), gráfico dos valores experimental *versus* previsto (e), previsto *versus* o resíduo e histograma dos resíduos (f) para os cálculos da resposta F2 (Exemplo 1).

- Figura 2S. Teste F para MQR e MQR (a), teste F para a MQfaj e MQEP (b) variação explicada (R^2) e explicável (R^2_{max}) (c), coeficientes de regressão (d), gráfico dos valores experimental *versus* previsto (e), previsto *versus* o resíduo e histograma dos resíduos (f) para os cálculos da resposta F3 (Exemplo 1).
- Figura 3S. Teste F para MQR e MQR (b), teste F para a MQfaj e MQEP (a) variação explicada (R^2) (c) para os cálculos da resposta D (Exemplo 2).
- Figura 4S. Região plana delimitada pelo triângulo equilátero com vértices (1,0,0), (0,1,0) e (0,0,1) que contém o domínio de F3.
- Figura 5S. Domínio T de f3 delimitado pelo triângulo retângulo com vértices (0,0), (1,0) e (0,1).

Todo o conteúdo se encontra em <http://quimicanova.s bq.org.br>, em formato PDF, com acesso livre.

AGRADECIMENTOS

Os autores são gratos à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, processos 2016/17221-8 e 2019/24223-5) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, processos 302719/2020-2). Os autores expressam gratidão à PubliSBQ por permitir o uso dos exemplos em caráter ilustrativo neste tutorial. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

REFERÊNCIAS

1. Leardi, R.; *Anal. Chim. Acta* **2009**, *652*, 161.
2. Sahu, P. K.; Ramiseti, N. R.; Cecchi, T.; Swain, S.; Patro, C. S.; Panda, J.; *J. Pharm. Biomed. Anal.* **2018**, *147*, 590
3. Coscione, A. R.; De Andrade, J. C.; May, G. M.; *Quim. Nova* **2005**, *28*, 1116.
4. Novaes, C. G.; Yamaki, R. T.; de Paula, V. F.; do Nascimento Júnior, B. B.; Barreto, J. A.; Valasques, G. S.; Bezerra, M. A.; *Rev. Virtual Quim.* **2017**, *9*, 1184.
5. Novaes, C. G.; Yamaki, R. T.; de Paula, V. F.; do Nascimento Júnior, B. B.; Barreto, J. A.; Valasques, G. S.; Bezerra, M. A.; *Rev. Virtual Quim.* **2018**, *10*.
6. Bezerra, M. A.; Ferreira, S. L. C.; Novaes, C. G.; dos Santos, A. M. P.; Valasques, G. S.; Cerqueira, U. M. F.; Alves, J. P. S.; *Talanta* **2019**, *194*, 941.
7. Candiotti, L. V.; Zan, M. M.; Cámara, M. S.; Goicoechea, H. C.; *Talanta*, **2014**, *124*, 123.
8. Pereira, F. M. V.; Pereira-Filho, E. R.; *Quim. Nova* **2018**, *41*, 1061.
9. Damasceno, A. I. C.; Reis, C.; Reis, E. L.; Bellato, C. R.; Oliveira, A. F.; *J. Braz. Chem. Soc.* **2016**, *27*, 877.
10. Estevão, P. L. S.; Peralta-Zamora, P.; Nagata, N.; *J. Braz. Chem. Soc.* **2018**, *29*, 2104.
11. Costa, N. R.; Lourenço, J.; Pereira, Z. L.; *Chemometr. Intell. Lab. Syst.* **2011**, *107*, 234.
12. Derringer, G.; Suich, R.; *J. Qual. Technol.* **1980**, *12*, 214.
13. Myers, R. H.; Montgomery, D. C.; Anderson-Cook, C. M.; *Response surface methodology process and product optimization using designed experiments*, Wiley: Hoboken, 2009.
14. Barros Neto, B.; Scarminio, I. S.; Bruns, R. E.; *Como fazer experimentos*, Bookman: Porto Alegre, 2010.