

## Modelos Estatísticos Fatoriais: Conceitos e Aplicações Factorial Statistical Models: Concepts and Applications

Pedro Henrique Mainardi<sup>1</sup>; Ederio Dino Bidoia<sup>2</sup>

<sup>1</sup>Universidade Estadual Paulista Júlio de Mesquita Filho - UNESP, Instituto de Biociências, *Departamento de Biologia Geral e Aplicada*, Rio Claro, SP, Brasil. E-mail: [pedro.h.mainardi@gmail.com](mailto:pedro.h.mainardi@gmail.com); ORCID: <https://orcid.org/0000-0003-0234-0052>.

<sup>2</sup>Universidade Estadual Paulista Júlio de Mesquita Filho - UNESP, Instituto de Biociências, *Departamento de Biologia Geral e Aplicada*, Rio Claro, SP, Brasil. E-mail: [ederio.bidoia@unesp.br](mailto:ederio.bidoia@unesp.br); ORCID: <https://orcid.org/0000-0001-7040-1983>.

### Resumo

Modelos estatísticos são fundamentais em investigações científicas, amplamente utilizados na caracterização da relação entre variáveis de interesse, quantificação dos efeitos de fatores, estimação da variabilidade experimental, otimização de processos e também na derivação de funções matemáticas que descrevem e preveem o comportamento dos resultados. A partir da utilização de matrizes com todas as combinações possíveis, modelos estatísticos fatoriais foram vistos por serem capazes de gerar grande quantidade de informação em um número reduzido de ensaios, além de considerar simultaneamente as interações de efeito de experimentos com múltiplas variáveis, ajustar modelos com curvatura e supor com precisão a magnitude de erros experimentais. Modelos fatoriais têm sido fracionalizados com sucesso, reduzindo consideravelmente o número de experimentos sem a perda de informação relevante. As características de ortogonalidade e rotacionalidade (rotatabilidade), adquiridas com a adição de pontos estrelas e centrais aos arranjos dos planejamentos fatoriais, têm possibilitado análises robustas dos efeitos individuais e combinados das variáveis, além de propiciar a construção de equações matemáticas quadráticas de forma eficaz. Relatos publicados na literatura indicaram que modelos estatísticos fatoriais, também chamados de designs, delineamentos e planejamentos fatoriais, foram aplicados em extensivos e compreensivos estudos com múltiplos fatores e utilizados com sucesso na elaboração de equações polinomiais simples ou complexas relacionadas com os resultados experimentais. Foram também vastamente utilizados no desenho de gráficos de superfície de resposta, plotes de contorno e na assunção de importantes características de experimentos científicos e processos industriais, como a influência dos efeitos individuais e das interações entre os fatores, a variação residual, e ótimas condições operacionais. Nesta revisão, foram conceitualizados os aspectos fundamentais dos modelos estatísticos fatoriais e descritas aplicações bem sucedidas dessa poderosa ferramenta.

**Palavras-chave:** Planejamento experimental, delineamento experimental, desenho experimental, design experimental, modelagem estatística, otimização de processos.

### Abstract

Statistical models are fundamental in scientific investigations, widely used to characterize the relationship between variables of interest, quantify the effects of factors, estimate the experimental variability, optimize processes and also to derive mathematical functions that describe and predict the behavior of the results. From the use of a matrix with all the possible combinations, factorial statistical models were seen to be able to generate a large amount of information in a reduced number of tests, as well to simultaneously consider the interaction effects of multiple variables experiments, fit models with curvature and accurately presume the magnitude of experimental errors. Factorial models have been successfully fractionalized, considerably reducing the number of experiments without the loss of relevant information. The orthogonality and rotationality (rotatability) characteristics, acquired with the addition of star and central points to the factorial arrays, have allowed robust analyzes of the individual

and combined effects of the variables, and effectively enabled the construction of quadratic mathematical equations. Reports published in the literature indicated that factorial statistical models, also called factorial designs and plans, were applied in extensive and comprehensive multiple factors studies and successfully used to elaborate simple or complex polynomial equations related to the experimental results. They were also widely used to draw response surface graphs, contour plots and to assume important characteristics of scientific experiments and industrial processes, such as the influence of individual and the interaction effects of the factors, the residual variation, and the optimal operational conditions. In this review, the fundamental aspects of the factorial statistical models were conceptualized and successful applications of this powerful tool were described.

**Keywords:** Experimental planning, experimental design, DOE, design of experiments, statistical modeling, process optimization.

## Introdução

A experimentação tem sido parte essencial da metodologia científica, fundamental para a produção de conhecimento, progresso da ciência e desenvolvimento tecnológico [1]. Através de modelos estatísticos, constituídos a partir de um conjunto de ferramentas de descrições, interpretações e suposições de dados, os investigadores têm conseguido gerar o máximo de informação relevante em um número mínimo de experimentações possíveis. Os modelos, derivados de observações empíricas e cálculos de vários parâmetros estatísticos, têm sido amplamente utilizados para caracterizar as relações entre as respostas de diferentes tipos de variáveis e estimar com precisão a posição e a influência dos fatores estudados [2,3,4]. Os modelos, chamados de delineamentos e designs estatísticos, também têm propiciado supor combinações ideais de tratamento, diferenciar as variações assinaláveis das incontroláveis e calcular a magnitude de erros de medida [5,6].

Através de modelos estatísticos, também tem sido possível a construção de equações matemáticas que descrevem o comportamento das respostas experimentais, expressas, em geral, como demonstrado na equação 1 [7]. Na equação, a parte suave ou regular  $M$ , refere-se a parte previsível do modelo estatístico. Já a parte  $R$ , considerada tão importante quanto a parte  $M$ , refere-se a “erros de ajustamento”, também chamado de resíduo. O resíduo, calculado a partir de fontes incontroláveis de variação, tem sido interpretado como uma medida de discrepância entre os valores observados e propostos pelos modelos estatísticos [3]. As análises de resíduos, geralmente feitas através de medidas de intervalo e de ajuste, ou de forma subjetiva utilizando-se de gráficos e diagramas, têm sido cruciais para validar a adequação dos modelos teóricos [2,7].

$$D = M + R \quad (1)$$

onde  $D$  refere-se a representação ou descrição dos dados observados,  $M$  ao modelo estatístico, e  $R$  ao resíduo do modelo estatístico.

Os dados utilizados na derivação dos modelos estatísticos têm sido sujeitos a vários tipos de especificações, observações, medidas e estimativas de erros experimentais [3]. Valores de cálculos estatísticos descritivos, que sumarizam conjuntos de dados, como a média aritmética simples, moda, mediana, têm tido papel fundamental na caracterização da relação entre as variáveis de interesse e as respostas experimentais. Medidas de dispersões, como a amplitude, desvios e variâncias, também têm sido cruciais na interpretação e construção de modelos estatísticos (Equações 2, 3, 4, 5, 6 e 7) [8,9,10,11,12,13]. A adequação das assunções dos modelos estatísticos, que podem ter muitas formas e configurações, tem sido confirmada de forma visual ou a partir de procedimentos de inferências e medidas de ajuste [2,3,14].

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1 + x_2 + \dots + x_n}{n} \quad (2)$$

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (3)$$

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2} \quad (4)$$

$$EP = \frac{\sigma}{\sqrt{N}} \quad \text{ou} \quad \frac{S}{\sqrt{n}} \quad (5)$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (6)$$

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 \quad (7)$$

onde  $\bar{x}$  refere-se a média aritmética,  $x_i$  a resposta de uma variável randômica,  $n$  ao tamanho amostral,  $S$  ao desvio padrão amostral,  $\sigma$  ao desvio padrão populacional,  $\mu$  a média da população,  $N$  ao tamanho populacional,  $EP$  ao erro padrão da média amostral ou populacional,  $S^2$  a variância amostral, e  $\sigma^2$  a variância populacional.

### Inferências estatísticas

Modelos estatísticos têm propiciado a dedução de afirmações sobre determinadas características populacionais e processos de interesse através de medidas ou observações feitas em um conjunto representativo, chamado de grupo amostral ou amostragem [2,7]. As afirmações, também chamadas de inferências, têm envolvido um grau de incerteza devido a variação experimental incontrolável [3]. As inferências têm sido feitas a partir da utilização de modelos estatísticos que representam a probabilidade da ocorrência de eventos, chamadas de distribuições de frequências probabilísticas. Em grande parte dos modelos estatísticos, têm se atribuído o “teorema do limite central” (TLC), com distribuições independentes e frequências próximas da normalidade [15]. Em casos assim, têm se assumido que os dados se concentravam simetricamente em torno de um valor central. Em amostragens pequenas ( $n < 30$ ), devido à ausência dos desvios padrões populacionais, têm sido utilizado os desvios amostrais nos testes de inferências. Nestes casos, a distribuição normal tem sido substituída pela distribuição  $t$  de *Student*, que ajusta o número de observações independentes através do grau de liberdade [4,16,17].

## Testes de hipóteses

A presunção das diferenças significativas entre os valores hipotéticos e os valores reais de fenômenos, processos e parâmetros estatísticos, tem sido comumente feita através de testes de hipóteses. As inferências dos testes de hipótese têm sido validadas através de um critério com duas suposições: uma hipótese nula ( $H_0$ ) e uma hipótese alternativa ( $H_A$ ) [2]. A hipótese nula tem afirmado que haviam evidências suficientes que comprovavam que não existiam relações entre dois fenômenos medidos, e a hipótese alternativa, de que não haviam evidências suficientes para tal afirmação. As presunções, dependendo do tamanho amostral, têm sido feitas através de testes seguindo modelos de distribuições normal ou  $t$  de *Student*. Os valores do teste têm sido comparados com regiões críticas previamente tabeladas das distribuições, determinadas por níveis de significâncias (Equações 8 e 9) [7]. Através do nível de significância, também chamado de valor  $\alpha$ , tem sido possível determinar a probabilidade de se rejeitar a hipótese nula quando ela era verdadeira. Nessas análises, têm sido utilizado uma medida de evidência, relacionada com probabilidade de se obter resultados extremos sob a hipótese nula, chamado de  $p$ -valor, probabilidade de significância ou nível descritivo [18]. O  $p$ -valor tem sido interpretado como um índice de redução de confiabilidade, ou seja, quanto maior o seu valor, menos confiável tem sido a relação observada entre os parâmetros de interesse. De acordo com o teste, resultados de  $p$ -valor menores, ou iguais, do que os níveis de significância determinados, tem acarretado em rejeições das hipóteses nulas. Testes de inferências com resultados de  $p$ -valor significativos, mediante a níveis de significância determinados em 0,01, entretanto, ainda têm indicado riscos de 1% de falsas conclusões, ou seja, a probabilidade de erroneamente rejeitar a hipótese nula e de ocorrer um “falso positivo” [19].

$$Z_{calc} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \quad (8)$$

$$t_{calc} = \frac{\bar{x} - \mu}{S/\sqrt{n}} \quad (9)$$

onde  $Z_{calc}$  e  $t_{calc}$  referem-se aos valores das estatísticas dos testes seguindo as respectivas distribuições normal ou  $t$  de *Student*,  $\bar{x}$  a média amostral,  $\mu$  a média populacional,  $n$  ao número de observações independentes (distribuição  $t$  de *Student* com grau de liberdade de:  $n - 1$ ),  $\sigma$  ao desvio padrão populacional, e  $S$  ao desvio padrão amostral.

## Análises de regressão

Modelos estatísticos, através de métodos de regressão, têm propiciado a caracterização da relação entre as variáveis independentes e dependentes utilizadas na análise, chamadas respectivamente de “variáveis preditoras” e “variáveis respostas” [6]. As técnicas de regressão, amplamente utilizadas em modelagens estatísticas, têm sido utilizadas para estimar quantitativamente a contribuição das variáveis independentes na previsão das variáveis dependentes, avaliar a adequação das suposições, examinar a influência de observações atípicas (*outliers*) e também identificar a presença de efeitos lineares entre as variáveis, chamados de correlações [2,3]. Em geral, as relações entre as variáveis têm sido expressas em funções matemáticas, calculadas através do método de quadrados mínimos. O método tem proporcionado uma estimativa numérica dos interceptos verticais e dos ângulos de inclinação das retas das equações de regressão, em outras palavras, dos valores das constantes e dos coeficientes angulares, também chamados de declive ou *slope*, das funções lineares. As associações das respostas têm sido descritas através de coeficientes lineares simples ou complexos, ou seja, de primeira ou mais ordens (Equações 10 e 11). Os modelos de regressão têm sido referidos como lineares pois os coeficientes desconhecidos têm aparecido de forma linear, ou seja, como aditivos ou multiplicadores constantes dos

valores das variáveis preditoras [3]. Modelos de regressão têm sido diagnosticados através de medidas de ajuste, análises de inferências e visualmente através de gráficos, como os de dispersões e box-plotes [2,3,6].

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + e \quad (10)$$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + e \quad (11)$$

onde  $y$  refere-se a resposta prevista através do modelo ajustado,  $x...$  as variáveis preditoras,  $\beta_0$  ao valor do intercepto,  $\beta...$  aos coeficientes parciais lineares simples ou quadráticos (parabólicos),  $k$  ao número de variáveis preditoras, e  $e$  ao erro de ajustamento.

### Efeitos dos fatores

Em modelos estatísticos, as variáveis independentes, ou seja, as variáveis preditoras e controláveis, têm sido comumente chamadas de fatores. Assim como visto em modelos de regressão, as relações entre os fatores e as respostas têm sido quantitativamente calculadas e descritas na forma de funções polinomiais (Equação 12) [20]. Os efeitos dos fatores também têm sido utilizados em testes de inferências e nas estimativas de intervalo [3]. As interações ou correlações de efeitos, relações não lineares e lineares entre as respostas de três ou mais variáveis, têm ocorrido quando os resultados de um fator dependiam do estado de outro fator. Os efeitos das interações também têm sido medidos e descritos na forma equações matemáticas em modelos de regressão (Equação 13) [2]. Em modelos estatísticos balanceados, ou seja, com quantidades iguais de repetições para todas as combinações de fatores com valores igualmente espaçados, os coeficientes dos efeitos foram previamente tabulados, o que simplifica a derivação de funções polinomiais de segunda ou mais ordens (Equação 14) [3]. Os valores das variáveis preditoras têm sido utilizados de forma padronizada (uniformizada) e codificada, facilitando análises comparativas (Equação 15).

$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + e \quad (12)$$

$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{1 \leq i < j \leq k} \beta_{ij} x_i x_j + e \quad (13)$$

$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i=1}^k \beta_{ii} x_i^2 + \sum_{1 \leq i < j \leq k} \beta_{ij} x_i x_j + e \quad (14)$$

$$z_i = \frac{x_i - \bar{x}_i}{S_i} \quad (15)$$

onde  $y$  refere-se a resposta prevista,  $k$  ao número de variáveis,  $\beta_0$  ao termo constante,  $\beta_i$  ao coeficiente linear simples,  $\beta_{ij}$  ao coeficiente de interação,  $\beta_{ii}$  ao coeficiente quadrático (segunda ordem),  $x_i$  e  $x_j$

as variáveis preditoras,  $e$  ao resíduo,  $z_i$  a uma variável preditora padronizada,  $\bar{x}_i$  a média amostral, e  $S_i$  ao desvio padrão amostral.

### Coefficientes de determinação

O coeficiente de determinação, também chamado de coeficiente explanatório ou  $R^2$ , tem sido interpretado como a medida da porcentagem das respostas que era explicado pelas equações de regressão. Em outras palavras, o coeficiente de determinação tem quantificado a proporção da adequação entre os valores observados em relação aos valores propostos pelos modelos estatísticos [4,21]. O coeficiente de determinação ajustado ( $R^2a$ ), por sua vez, tem sido definido como um ajustamento do  $R^2$ , pois tem considerado o número de observações independentes, o grau de liberdade. O coeficiente de determinação ajustado tem sido utilizado para comparar modelos estatísticos com diferentes números de variáveis preditoras [2]. A qualidade do ajuste da equação de regressão, de acordo com o coeficiente de determinação  $R^2$ , tem sido calculado como indicado na equação 16. O cálculo do  $R^2$  ajustado tem sido feito como demonstrado na equação 17. Em geral, os valores de  $R^2$  e  $R^2a$  têm variado entre 1 e 0, e expressos na forma de porcentagens.

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (16)$$

$$R^2a = 1 - \frac{n - 1}{n - k - 1} (1 - R^2) \quad (17)$$

onde  $R^2$  refere-se ao coeficiente de explicação,  $R^2a$  ao coeficiente de explicação ajustado,  $y_i$  ao valor experimental observado na variável resposta em função do nível  $i$  de uma variável preditora,  $\bar{y}$  a média geral dos valores observados,  $\hat{y}_i$  ao valor estimado ou previsto através do modelo ajustado para uma variável resposta em função do nível  $i$  de uma variável preditora,  $n$  ao número de observações, e  $k$  ao número de variáveis preditoras do modelo.

### Intervalos de confiança e previsão

O intervalo de confiança, um tipo de estimativa por intervalo calculado a partir das observações, tem sido utilizado para inferir informações referentes a precisão dos estimadores, ou seja, da estatística utilizada para estimar os parâmetros [3]. Em modelos estatísticos, os intervalos têm sido associados a um nível de confiança, cuja amplitude estimava a frequência em que o valor de um parâmetro de interesse estaria incluso em infinitas repetições hipotéticas [19]. Os intervalos de confiança, além de terem sido usados para representar frequências de repetição em torno de estimativas de incertezas, também têm indicado a imprecisão das médias das respostas [22,23]. Intervalos de confiança de 95%, por exemplo, tem indicado que a média do valor do parâmetro de interesse tenderia a estar incluso dentro das extremidades calculadas em 95% das vezes, caso o procedimento fosse repetido inúmeras vezes [24].

O intervalo de previsão, ou predição, também tem sido interpretado como uma estimativa de intervalo calculada a partir de observações já realizadas, porém, tem indicado valores em que uma nova observação provavelmente iria cair, mediante a uma determinada probabilidade. Esse tipo intervalo tem propiciados aos modelos estatísticos estimativas mais precisas em relação a ocorrências de eventos e respostas experimentais [2]. Em modelos de regressão, ambos os intervalos têm sido frequentemente representados através de bandas em torno das retas das equações. O cálculo de um intervalo com nível de confiança  $100(1-\alpha)\%$  de um tamanho amostral grande, normalmente distribuído com desvio padrão

populacional conhecido, tem sido feito como descrito na equação 18. O intervalo de predição, similar ao intervalo de confiança, tem sido calculado como indicado na equação 19.

$$\hat{\theta} = \bar{x} \mp z_{\alpha} \frac{\sigma}{\sqrt{n}} \quad (18)$$

$$\hat{\theta} = \bar{x} \pm z_{\alpha} \sigma \quad (19)$$

onde  $\hat{\theta}$  refere-se a estimativa de intervalo;  $\bar{x}$  a média amostral,  $Z$  ao valor tabelado da distribuição normal,  $\alpha$  a probabilidade do nível de confiança,  $\sigma$  ao desvio padrão populacional, e  $n$  ao número de observações independentes.

### Análises de variância

Modelos estatísticos tem se utilizado de técnicas analíticas que possibilitam estimar diferenças significativas entre as médias de grupos de tratamento através do cálculo de suas variâncias. O método, chamado de ANOVA, tem possibilitado separar ou particionar a variação observável dos delineamentos em dois componentes: causas de variações atribuíveis e causas incontroláveis, e supor se haviam diferenças significativas entre eles [3]. Causas atribuíveis têm sido consideradas como fontes de variações possíveis de serem medidas e controláveis, enquanto as incontroláveis, fontes de variações randômicas [4,7]. Em geral, a decomposição das variabilidades em amostras aleatórias e independentes, com frequências normalmente distribuídas e variâncias populacionais iguais, tem sido feita através da Soma Total de Quadrados (Equações 20 e 21). A Soma de Quadrados do erro e a Soma de Quadrados do modelo de Regressão, chamados respectivamente de Resíduo e modelo ajustado, têm sido calculados através das equações 22 e 23.

$$SQ_{tot} = SQ_{reg} + SQ_{res} \quad (20)$$

$$SQ_{tot} = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (21)$$

$$SQ_{res} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (22)$$

$$SQ_{reg} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad (23)$$

onde  $SQ_{tot}$  refere-se a soma dos quadrados devido a variação Total,  $SQ_{reg}$  a soma de quadrados devido a parâmetros considerados no modelo (Regressão),  $SQ_{res}$  a soma de quadrados devido a erros de ajustamento (Resíduo),  $y_i$  ao valor observado na variável resposta em função do nível  $i$  de uma variável preditora,  $\bar{y}$  a média geral dos valores observados,  $\hat{y}_i$  ao valor previsto através do modelo ajustado para uma variável resposta em função do nível  $i$  de uma variável preditora, e  $n$  ao número de observações independentes.

### Testes de significância da ANOVA

Nas análises de variância, a probabilidade de se obter afirmações conclusivas sobre as diferenças entre as médias dos grupos de tratamento, em geral, tem sido calculada através da razão entre o Quadrado

Médio da Regressão e do Resíduo, denominado teste  $F$  (Equação 24). As estimativas do teste têm sido comparadas com regiões críticas da distribuição *Fisher-Snedecor*, mediante a níveis de significância previamente determinados [2,25]. Os testes de ANOVA têm sido comumente utilizados para medir a magnitude das significâncias estatísticas, ou seja, das estimativas relacionadas ao "grau de certeza" entre as associações dos grupos. Em análises de variância com múltiplas variáveis, o teste  $F$  tem sido feito de forma independente para cada fator e interação, incluindo as covariáveis, blocos e curvaturas [12].

$$F_{calc} = \frac{QMreg}{QMres} \quad (24)$$

onde  $F_{calc}$  refere-se ao valor da estatística do teste seguindo a distribuição  $F$ ,  $QMreg$  ao quadrado médio da Regressão, e  $QMres$  ao quadrado médio do Resíduo.

### Testes de falta-de-ajuste (*lack-of-fit*)

Os resíduos dos modelos estatísticos, importantes parâmetros de confiabilidade, têm sido utilizados em métodos de inferências que possibilitam validar a adequação dos modelos hipotéticos [4]. As estimativas têm sido feitas utilizando-se da variação entre as repetições dentro de cada tratamento, ou seja, das flutuações randômicas entre as réplicas autênticas [20]. Os cálculos têm sido feitos particionando a Soma dos Quadrados dos Resíduos em dois componentes úteis para testes de hipóteses mediante a determinados níveis de significâncias, chamados de Erro Puro e Falta-de-Ajuste (*lack-of-fit*) (Equações 25 e 26). Testes de hipótese de falta-de-ajuste estatisticamente significantes ( $p \leq \alpha$ ) têm indicado, em geral, uma variabilidade residual maior do que a do erro puro (Equação 27). Dependendo do critério de significância, testes significativos de falta-de-ajuste têm indicado um modelo com baixo grau de adequação em função dos resultados experimentais, requerendo, em alguns casos, a utilização de designs mais complexos na descrição do comportamento das respostas [4].

Ressalta-se que o teste de falta-de-ajuste da soma de quadrados tem sido altamente sensível quanto a desvios de normalidade, tamanhos amostrais muito grandes ou pequenos, e quanto a presença de observações influenciadoras, aquelas com valores atípicos, chamadas de "outliers" [26]. Durante os últimos anos, notáveis pesquisas têm sido feitas envoltas no desenvolvimento de novas técnicas de diagnósticos e transformações estruturais de modelos estatísticos. Métodos para validar as suposições propostas por diferentes tipos de modelos, como a bondade de ajuste (*goodness-of-fit*), índices de ajustes incrementais (*incremental fit indices*), índices de parcimônia (*parsimony fit indices*), entre outros, foram descritos na literatura [26,27,28,29,30].

$$SQep = \sum_{i=1}^n (y_{ij} - \bar{y}_i)^2 \quad (25)$$

$$SQres = SQfa + SQep = \sum_{i=1}^n (y_{ij} - \hat{y}_i)^2 = \sum_{i=1}^n (\bar{y}_{ij} - \hat{y}_i)^2 + \sum_{i=1}^n (y_{ij} - \bar{y}_i)^2 \quad (26)$$

$$F_{calc} = \frac{SQfa}{SQep} \quad (27)$$

onde  $SQep$  refere-se a soma dos quadrados do Erro Puro,  $SQres$  a soma de quadrados do Resíduo,  $SQfa$  a soma de quadrados de Falta de Ajuste,  $F_{calc}$  ao valor da estatística do teste seguindo a distribuição  $F$ ,

$y_{ij}$  ao valor da repetição  $j$  observado na variável resposta em função do nível  $i$  de uma variável preditora,  $\bar{y}_i$  a média geral dos valores observados no nível  $i$  de uma variável preditora,  $\hat{y}_i$  ao valor previsto através do modelo ajustado para uma variável resposta em função do nível  $i$  de uma variável preditora,  $\bar{y}_{ij}$  ao valor médio das observações entre as repetições no nível  $i$  de uma variável preditora, e  $n$  ao número de observações independentes.

## Outliers

A presença de *outliers*, observações com valores atípicos em relação a outras observações obtidas sob a mesma condição, foi vista por afetar fortemente as estimativas de parâmetros, procedimentos de inferências e também os modelos de regressão [3]. Os *outliers*, geralmente com alto valor residual, têm sido identificados de forma visual através de gráficos e diagramas, como *box plots*, *plots* de resíduos, Quantil-Quantil de normalidade, e também por técnicas como o teste de Grubbs, valores de *Leverage*, distância de Cook e análises padronizadas e studentizadas dos resíduos, chamadas de “*standardized residuals*” e “*studentized residuals*” [2,6,31,32,33]. Métodos de acomodamento, como a coleta de mais dados, reespecificação do modelo, deleção, limitação (*winsorizing*) e aparo (*trimming*) das observações extremas, têm sido frequentemente utilizados para mitigar os efeitos de *outliers* [3,34]. As técnicas, entretanto, têm sido procedidas de forma cuidadosa, justificada e gradual, a partir de métodos passo a passo (*stepwise selection methods*), como a seleção pra frente (*forward selection*), eliminação para trás (*backward elimination*) e iteração gradual (*stepwise iteration*) [2,35].

## Diagramas de Pareto

Diagramas de Pareto, gráficos que contêm histogramas e linhas, têm sido utilizados para ordenar as frequências acordo com suas ocorrências. Em modelos estatísticos, os diagramas de Pareto têm sido utilizados para ordenar os efeitos individuais e combinados das variáveis de acordo as magnitudes de suas significâncias estatísticas. A descrição dos efeitos em valores absolutos e uniformizados, também chamados de valores padronizados, tem possibilitado a inclusão de linhas de referências que indicam as magnitudes mínimas para efeitos estatísticos significativos mediante a determinados níveis  $\alpha$  de significância [4]. Em modelagens estatísticas, os diagramas de Pareto têm possibilitado a priorização de variáveis de maior importância de forma simplificada e frequentemente utilizados em gráficos de controle de qualidade [36].

## Superfície de resposta e plotes de contorno

A relação de dois ou mais fatores tem sido desenhada em gráficos com duas ou três dimensões, chamados de plotes de contorno ou de superfície de resposta. Os gráficos, construídos através de funções matemáticas da relação entre as variáveis independentes e dependentes, têm sido utilizados para simplificar a interpretação dos resultados empíricos, principalmente em experimentos com múltiplos fatores [3,5]. Em investigações científicas, gráficos de contorno e de superfície de resposta têm sido frequentemente utilizados na avaliação de importantes características de modelos estatísticos, como o comportamento dos resultados, sensibilidade das variáveis, e efeitos de interações. Na avaliação de processos industriais, gráficos de superfície de resposta e de contorno têm sido utilizados no aprimoramento da qualidade produtos e na estimação das condições ótimas de operação de sistemas [4,37].

## Modelos com múltiplo-fatores

Modelos estatísticos têm sido cada vez mais utilizados em experimentos com múltiplos fatores, chamados também de multi-fatores. A aplicação de modelos estatísticos apropriados nesse tipo de

estudo tem permitido uma eficiente caracterização dos efeitos individuais e também das interações das variáveis em função das respostas, além de propiciar a aquisição de preciosas estimativas em torno da variabilidade dos erros de ajuste [20,38]. Modelos estatísticos com múltiplas variáveis têm sido comumente utilizados na construção de equações matemáticas que descrevem o comportamento das respostas experimentais em determinadas regiões de interesse [3]. As funções de primeira, segunda, terceira ou mais ordens, derivadas dos modelos estatísticos multi-fatores, têm sido utilizadas com sucesso na síntese de gráficos de superfície de respostas e plotes de contorno. Essa abordagem tem propiciado a estimação quantitativa do efeito dos fatores, e a otimização de variáveis de experimentos científicos e processos transformadores [4,39].

### **Modelos fatoriais completos**

A escolha entre os vários tipos de modelos estatísticos tem dependido do objetivo, quantidade de fatores envolvidos, recursos disponíveis, e especialmente, do número de experimentos requeridos [3,39]. O modelo Fatorial Completo (FC), um delineamento fatorial estatístico clássico, tem propiciado a combinação de todas as possibilidades de experimentos com múltiplo fatores através de uma matriz com coeficientes equidistantes e níveis predeterminados [40]. O design FC tem sido utilizado na caracterização dos efeitos individuais e combinados das variáveis, em estimações das variações de erros de medida e também a derivação de gráficos e funções matemáticas que representam o comportamento das respostas [5,6,12]. Os pontos de tratamento da matriz de um planejamento FC com três fatores e três níveis ( $3^3$ ) foram representados na figura 1. A função polinomial de segunda ordem gerada pelo modelo de regressão de um delineamento FC  $3^3$  foi descrita na equação 28.

De acordo com relatos, modelos FCs foram utilizados com sucesso na investigação e determinação das condições ideais de operação de diferentes tipos de processos, como em um procedimento de separação de fases de petróleo e em um sistema de dessalinização de água por eletro-diálise [41,42]. Os autores desses estudos informaram que os modelos matemáticos gerados pelos delineamentos FCs foram capazes de determinar de forma eficaz os fatores estatisticamente significativos nos sistemas. Os planejamentos estatísticos também propiciaram avaliações robustas sobre os efeitos das variáveis em função das respostas obtidas. As funções matemáticas construídas a partir das observações experimentais, de acordo com as descrições, foram capazes de serem utilizadas com sucesso para prever a eficiência e também as melhores condições operacionais dos processos.

Ara e colaboradores [43] e Ejim e colaboradores [44] também utilizaram o planejamento FC na otimização das variáveis experimentais que influenciavam um método espectrofotométrico de determinação de herbicidas e no entendimento dos efeitos da geometria interna de reatores de fluxo oscilatório, respectivamente. De acordo com os autores, os modelos estatísticos FCs propiciaram a determinação dos fatores que eram responsáveis por influenciar de forma significativa os processos, incluindo as interações de efeitos. Relatos de Chan e colaboradores [45] também indicaram que o delineamento FC foi aplicado com sucesso em um estudo de otimização dos parâmetros que influenciavam um procedimento de solda de metais. O modelo estatístico FC utilizado na investigação, de acordo com os autores, proporcionou a assunção das condições de processo que indicavam melhores taxas de produtividade e também de qualidade.

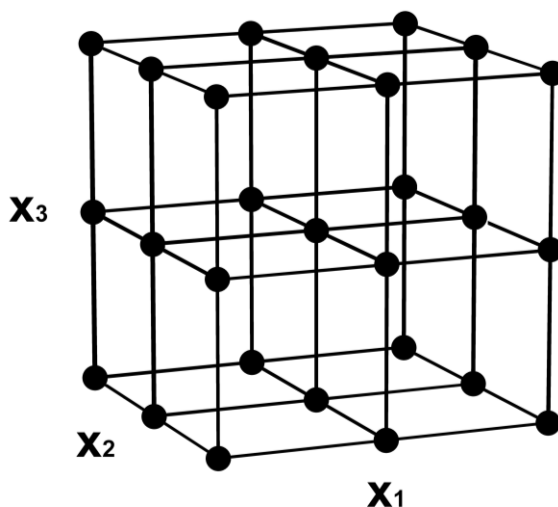


Figura 1 - Representação de um modelo estatístico Fatorial Completo (FC) com três níveis e três fatores ( $3^3$ ). Autoria própria.

$$\begin{aligned}
 y = & \beta_0 + \beta_1 x_1 + \beta_{11} x_1^2 + \beta_2 x_2 + \beta_{22} x_2^2 + \beta_3 x_3 + \beta_{33} x_3^2 + \beta_{12} x_1 x_2 + \beta_{122} x_1 x_2^2 \\
 & + \beta_{112} x_1^2 x_2 + \beta_{1122} x_1^2 x_2^2 + \beta_{13} x_1 x_3 + \beta_{133} x_1 x_3^2 + \beta_{113} x_1^2 x_3 \\
 & + \beta_{1133} x_1^2 x_3^2 + \beta_{23} x_2 x_3 + \beta_{233} x_2 x_3^2 + \beta_{223} x_2^2 x_3 + \beta_{2233} x_2^2 x_3^2 + e
 \end{aligned}
 \tag{28}$$

onde  $y$  refere-se a resposta prevista através do modelo ajustado,  $\beta_0$  ao valor do intercepto,  $x_1$ ,  $x_2$  e  $x_3$  as variáveis preditoras,  $\beta...$  aos coeficientes lineares parciais, e  $e$  ao erro de ajustamento.

### Modelos fatoriais fracionados

Devido a limitações envolvendo o tempo e recursos disponíveis, modelos fatoriais completos nem sempre têm sido possíveis de serem conduzidos. Nesses casos, têm sido frequentemente utilizados modelos Fatoriais Fracionados (FF), ou seja, modelos que se consistiam de um subconjunto dos pontos experimentais de um planejamento FC, em outras palavras, uma fração das corridas de tratamento [46]. A fracionalização dos modelos fatoriais tem sido feita através da utilização de efeitos “confusão”, cujos valores somente poderiam ser atribuídos à influencias combinadas de respostas, e não a respostas individuais únicas, chamados também de “*confounded*” ou “*aliased effects*” [3]. As frações das matrizes de planejamentos FFs têm sido cuidadosamente escolhidas, assegurando que os efeitos de interesse não estejam confundidos com outros de interesse ou, ao menos, estejam confundidos com efeitos de variáveis que não apresentavam magnitudes apreciáveis [3]. Esse tipo de aplicação tem possibilitado uma redução significativa do número de experimentações e, ainda sim, propiciado uma investigação compreensiva dos fatores sem a perda de informação relevante [46,47].

Delineamentos FFs têm sido utilizados com sucesso na identificação das variáveis significantes em experimentos científicos de reconhecimento e também na triagem de fatores em processos industriais. De acordo com Rezende e colaboradores [48], que utilizaram planejamentos FFs na otimização das variáveis de processos de pré-tratamento de biomassa destinados para a produção de bioetanol, o desenho experimental permitiu aos autores uma avaliação simultânea de um grande número de variáveis em um número reduzido de experimentos. Ma e colaboradores [49], por sua vez, também utilizaram o planejamento FF em sistemas aquosos de remoção de íons de Cobre por uma técnica de emulsão líquida por membranas. Os autores informaram que o modelo estatístico FF propiciou a identificação dos principais fatores envolvidos no processo, essencial para posterior otimização por método de superfície de resposta.

Planejamentos FFs também foram utilizados para otimizar as condições de extração de Zinco em alimentos de peixes para posterior caracterização por cromatografia de exclusão de tamanho, e avaliar os efeitos que cátions e ânions acarretavam em um processo de adsorção de Césio por óxido de grafeno. Os designs fracionados possibilitaram aos autores a determinação das variáveis mais importantes dos sistemas, e também a avaliação das interações entre um grande número de fatores em um número reduzido de corridas experimentais [50,51]. O delineamento FF também permitiu a Adio e colaboradores [52] determinar a influência que diferentes fatores causavam na remoção de Arsênio e Selênio da água utilizando nanopartículas de Ferro. Os autores relataram que o modelo estatístico gerado pelo design FF foi útil na identificação das variáveis que eram responsáveis por causar maior influência no processo, e também na determinação das condições que propiciavam maiores eficiências.

### Modelos compostos centrais rotacionais

O modelo Composto Central Rotacional (CCR), desenvolvido por Box e Hunter [53], tem se consistido de um design fatorial com as características de ortogonalidade e rotacionalidade (rotatabilidade), adquiridos pelas inserções de pontos estrelas (alpha) e centrais aos arranjos fatoriais. Essas condições têm propiciado as respectivas estimações independentes para os coeficientes dos modelos e variâncias idênticas para todos os pontos de tratamento, situados à mesma distância do centro [54]. Modelos CCRs têm possibilitado a obtenção de informações experimentais robustas em um número relativamente modesto de tratamentos, inclusive a quantificação dos efeitos individuais e combinados dos fatores, estimação da magnitude das variações de erros procedimentais, e também na construção de equações que expressavam respostas com curvaturas, ou seja, com funções de segunda ordem [3,17]. O modelo ajustado de designs estatísticos CCRs tem sido frequentemente utilizado na síntese de gráficos de contorno e de superfície de resposta [4]. Os pontos de tratamento da matriz de um planejamento estatístico fatorial CCR com três fatores, dois níveis e seis pontos estrelas foram representados na figura 2. A função polinomial de segunda ordem gerada pelo modelo de regressão do planejamento foi descrita na equação 29.

Relatos, como o de Ma e colaboradores [49], indicaram que planejamentos compostos centrais foram utilizados com sucesso na otimização dos fatores envolvidos na remoção de Cobre em sistemas aquosos por técnica de emulsão líquida por membranas. Outros relatos indicaram que delineamentos CCRs foram aplicados com sucesso em estudos de previsão e otimização dos fatores envolvidos em um processo de adsorção de corantes em nanopartículas imobilizadas em carvão ativado, e na otimização dos fatores envolvidos na formulação de fármacos para o tratamento de osteoporose [55,56]. Designs estatísticos CCRs também foram capazes de otimizar os parâmetros de um processo anaeróbio de tratamento de efluente em biorreatores [57]. O design CCR também foi possível de ser particionado com sucesso em blocos sem a perda das características de rotacionalidade e ortogonalidade em experimentos de adubação, uma característica altamente desejável [54]. De acordo com os relatos, as equações polinomiais construídas através dos modelos CCRs foram vistas por deter um bom ajuste em comparação com os dados experimentais, e possibilitaram a síntese de gráficos de superfície de resposta e plots de contorno de forma eficiente. Os autores afirmaram que os modelos estatísticos CCRs poderiam ser utilizados com sucesso em métodos de controle, previsão, e em procedimentos de aumento de escala (*scale-up*) de experimentos científicos e processos industriais. Aplicações de modelos estatísticos fatoriais em diferentes áreas e com diferentes finalidades publicadas na literatura foram descritas na tabela 1.

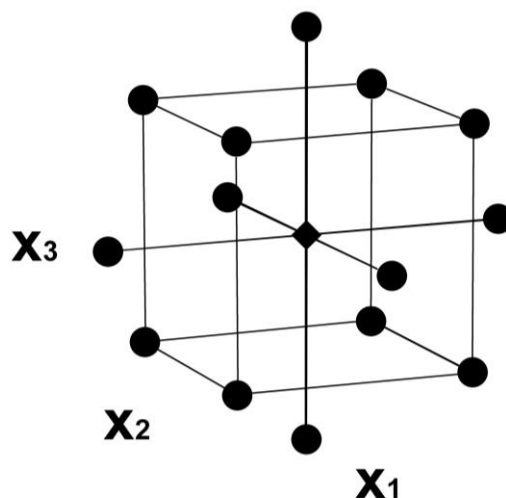


Figura 2 - Representação de um modelo estatístico fatorial Composto Central Rotacional (CCR) com dois níveis, três fatores e seis pontos estrelas ( $2^3$ ,  $\alpha = 8^{1/4} = 1,6818$ ). Autoria própria.

$$y = \beta_0 + \beta_1 x_1 + \beta_{11} x_1^2 + \beta_2 x_2 + \beta_{22} x_2^2 + \beta_3 x_3 + \beta_{33} x_3^2 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + e \quad (29)$$

onde  $y$  refere-se a resposta prevista através do modelo ajustado,  $\beta_0$  ao valor do intercepto,  $x_1$ ,  $x_2$  e  $x_3$  as variáveis preditoras,  $\beta...$  aos coeficientes lineares parciais, e  $e$  ao erro de ajustamento.

Tabela 1: Aplicações de modelos estatísticos fatoriais publicados em artigos na literatura.

Modelo	Área	Finalidade	Autores
Fatorial Completo	Engenharia química	Otimização de um processo de separação de fases de petróleo	[42]
Fatorial Completo	Engenharia ambiental	Otimização de um processo de dessalinização d'água por eletro-diálise	[41]
Fatorial Completo	Engenharia química	Entendimento do efeito da geometria da tubulação interna de um reator de fluxo oscilatório	[44]
Fatorial Completo	Engenharia	Otimização dos parâmetros de um processo de solda de metais	[45]
Fatorial Completo	Química	Optimização das variáveis experimentais de um método espectrofotométrico de determinação de herbicidas	[43]
Fatorial Fracionado	Ambiental	Avaliação dos efeitos que cátions e ânions acarretavam em um processo de adsorção de Césio por óxido de grafeno	[50]
Fatorial Fracionado	Biotecnologia	Otimização das variáveis de processos de pré-tratamento de biomassa destinados para a produção de bioetanol	[48]
Fatorial Fracionado	Química analítica	Otimização das condições de um método de extração de Zinco em alimentos de peixes para posterior caracterização por cromatografia de exclusão de tamanho	[51]
Fatorial Fracionado	Ambiental	Determinação da influência que diferentes fatores causavam na remoção de Arsênio e Selênio da água utilizando nanopartículas de Ferro	[52]

Fatorial Fracionado e Composto Central	Ambiental	Triagem e otimização dos fatores envolvidos na remoção de Cobre por técnica de emulsão líquida por membranas	[49]
Composto Central Rotacional	Agrícola	Experimentos de adubação em campo	[54]
Composto Central Rotacional	Engenharia ambiental	Otimização dos parâmetros de um processo de tratamento de efluente em biorreatores anaeróbios	[57]
Composto Central	Química	Previsão e otimização dos efeitos dos fatores envolvidos em um processo de adsorção de corantes em nanopartículas imobilizadas em carvão ativado	[55]
Composto Central	Farmacêutica	Otimização dos fatores envolvidos na formulação de fármacos para o tratamento de osteoporose	[56]

## Considerações finais

Através do uso de matrizes com todas as combinações possíveis, modelos estatísticos fatoriais têm sido extremamente úteis em experimentações científicas, amplamente utilizados na compreensão dos efeitos individuais e combinados de fatores, estimação da variabilidade de erros experimentais, otimização de parâmetros e também na construção de equações polinomiais que descrevem o comportamento das variáveis. A fracionalização dos delineamentos fatoriais tem sido utilizada com sucesso para reduzir de forma considerável o número de experimentos sem a perda de informação relevante. A propriedade tem sido extremamente útil na triagem de variáveis de interesse em experimentos com grande quantidade de fatores. As características de ortogonalidade e rotacionalidade (rotatabilidade) foram alcançadas com a adição de pontos estrelas (alpha) e centrais a matriz dos designs fatoriais. A condição tem proporcionado a obtenção de estimativas robustas em relação aos efeitos individuais e combinados dos fatores, respostas com curvatura, além da derivação de funções polinomiais quadráticas referentes ao comportamento das respostas. A partir de relatos da literatura, foi observado que planejamentos estatísticos fatoriais foram aplicados com sucesso em estudos da influência de diferentes tipos de fatores, estimação da magnitude de erros experimentais, síntese de gráficos de superfície de resposta e de contorno, e também na previsão e otimização das condições de diversas variáveis em experimentos científicos e processos industriais. A partir desta revisão, espera-se que os modelos estatísticos fatoriais sejam utilizados com maior frequência na análise de dados, modelagem matemática e otimização de processos.

## Agradecimentos

Os autores P.H.M. e E.D.B. agradecem a CAPES pelo auxílio financeiro. P.H.M. agradece C.T.M., J.F.M., G.M. e C.M. pelo vital suporte, carinho, colaboração e carinho. O presente trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES), Código de Financiamento 001.

## Referências bibliográficas

- [01] ZANCAN, Glaci T. Educação científica: uma prioridade nacional. **São Paulo em perspectiva**, v. 14, n. 3, p. 3-7, 2000.
- [02] CHRISTENSEN, Ronald. **Analysis of variance, design, and regression: applied statistical methods**. CRC Press, 1996.
- [03] MASON, Robert L.; GUNST, Richard F.; HESS, James L. **Statistical design and analysis of experiments: with applications to engineering and science**. John Wiley & Sons, 2003.

- [04] RODRIGUES, Maria Isabel; IEMMA, Antônio Francisco. **Planejamento de experimentos e otimização de processos: uma estratégia sequencial de planejamentos**. Casa do Pão Editora, 2005.
- [05] HINKELMANN, Klaus; KEMPTHORNE, Oscar. **Design and analysis of experiments**. New York: Wiley, 1994.
- [06] MONTGOMERY, Douglas C. **Design and analysis of experiments**. John Wiley & Sons, 2017.
- [07] MORETTIN, Pedro Alberto; BUSSAB, Wilton Oliveira. **Estatística básica**. Saraiva Educação SA, 2017.
- [08] YULE, G. Udny. On the theory of correlation. **Journal of the Royal Statistical Society**, v. 60, n. 4, p. 812-854, 1897.
- [09] PEARSON, Karl. Contributions to the mathematical theory of evolution. **Philosophical Transactions of the Royal Society of London. A**, v. 185, p. 71-110, 1894.
- [10] FISHER, Ronald A. XV.—The correlation between relatives on the supposition of Mendelian inheritance. **Earth and Environmental Science Transactions of the Royal Society of Edinburgh**, v. 52, n. 2, p. 399-433, 1919.
- [11] ZWILLINGER, Daniel; KOKOSKA, Stephen. **CRC standard probability and statistics tables and formulae**. Crc Press, 1999.
- [12] SHESKIN, David J. **Handbook of parametric and nonparametric statistical procedures**. Crc Press, 2003.
- [13] ALTMAN, Douglas G.; BLAND, J. Martin. Standard deviations and standard errors. **Bmj**, v. 331, n. 7521, p. 903, 2005.
- [14] TSAI, Chih-Ling; CAI, Zongwu; WU, Xizhi. The examination of residual plots. **Statistica Sinica**, p. 445-465, 1998.
- [15] GALTON, Francis. **Natural inheritance**. Macmillan and Company, 1894.
- [16] WILKS, **Mathematical Statistics**. Princeton Univ. 1944.
- [17] ONYIAH, Leonard C. **Design and analysis of experiments: classical and regression approaches with SAS**. CRC Press, 2008.
- [18] BROWNLEE, Kenneth Alexander. **Statistical theory and methodology in science and engineering**. New York: Wiley, 1965.
- [19] KALINOWSKI, Pawel; FIDLER, Fiona. Interpreting significance: the differences between statistical significance, effect size, and practical importance. **Newborn and Infant Nursing Reviews**, v. 10, n. 1, p. 50-54, 2010.
- [20] BEZERRA, Marcos Almeida et al. Response surface methodology (RSM) as a tool for optimization in analytical chemistry. **Talanta**, v. 76, n. 5, p. 965-977, 2008.

- [21] NAKAGAWA, Shinichi; SCHIELZETH, Holger. A general and simple method for obtaining R<sup>2</sup> from generalized linear mixed-effects models. **Methods in ecology and evolution**, v. 4, n. 2, p. 133-142, 2013.
- [22] NAKAGAWA, Shinichi; CUTHILL, Innes C. Effect size, confidence interval and statistical significance: a practical guide for biologists. **Biological reviews**, v. 82, n. 4, p. 591-605, 2007.
- [23] PATINO, Cecilia Maria; FERREIRA, Juliana Carvalho. Confidence intervals: a useful statistical tool to estimate effect sizes in the real world. **Jornal Brasileiro de Pneumologia**, v. 41, n. 6, p. 565-566, 2015.
- [24] COX, David Roxbee; HINKLEY, David Victor. **Theoretical statistics**. CRC Press, 1979.
- [25] HOAGLIN, David C.; MOSTELLER, Frederick; TUKEY, John W. (Ed.). **Fundamentals of exploratory analysis of variance**. John Wiley & Sons, 1991.
- [26] HOOPER, Daire; COUGHLAN, Joseph; MULLEN, Michael R. Structural equation modelling: Guidelines for determining model fit. **Electronic journal of business research methods**, v. 6, n. 1, p. 53-60, 2008.
- [27] D'AGOSTINO, Ralph B. **Goodness-of-fit-techniques**. CRC press, 1986.
- [28] MULAIK, Stanley A. et al. Evaluation of goodness-of-fit indices for structural equation models. **Psychological bulletin**, v. 105, n. 3, p. 430, 1989.
- [29] STEIGER, James H. Structural model evaluation and modification: An interval estimation approach. **Multivariate behavioral research**, v. 25, n. 2, p. 173-180, 1990.
- [30] MILES, Jeremy; SHEVLIN, Mark. A time and a place for incremental fit indices. **Personality and individual differences**, v. 42, n. 5, p. 869-874, 2007.
- [31] GRUBBS, Frank E. et al. Sample criteria for testing outlying observations. **The Annals of Mathematical Statistics**, v. 21, n. 1, p. 27-58, 1950.
- [32] COOK, R. Dennis. Detection of influential observation in linear regression. **Technometrics**, v. 19, n. 1, p. 15-18, 1977.
- [33] NETTER, John; WASSERMAN, William; KUTNER, M. H. **Applied linear statistical models: Regression, analysis of variance, and experimental designs**. RD Irwin: Homewood, IL, 1985.
- [34] TUKEY, John W. The future of data analysis. **The annals of mathematical statistics**, v. 33, n. 1, p. 1-67, 1962.
- [35] DRAPER, Norman R.; SMITH, Harry. **Applied regression analysis**. John Wiley & Sons, 1998.
- [36] PORTER, Stuart C.; VERSEPUT, Richard P.; CUNNINGHAM, Charles R. Process optimization using design of experiments. **Pharmaceutical technology**, v. 21, n. 10, p. 60-71, 1997.
- [37] NOORDIN, Mohd Yusof et al. Application of response surface methodology in describing the performance of coated carbide tools when turning AISI 1045 steel. **Journal of materials processing technology**, v. 145, n. 1, p. 46-58, 2004.

- [38] ANDERSON, Mark J.; WHITCOMB, Patrick J. Design of experiments. **Kirk-Othmer Encyclopedia of Chemical Technology**, p. 1-22, 2000.
- [39] WEISSMAN, Steven A.; ANDERSON, Neal G. Design of experiments (DoE) and process optimization. A review of recent publications. **Organic Process Research & Development**, v. 19, n. 11, p. 1605-1633, 2014.
- [40] FISHER, Ronald Aylmer. Design of experiments. **Br Med J**, v. 1, n. 3923, p. 554-554, 1936.
- [41] CAMACHO, Lucy Mar; FOX, Jessica A.; AJEDEGBA, Johnson O. Optimization of electro dialysis metathesis (EDM) desalination using factorial design methodology. **Desalination**, v. 403, p. 136-143, 2017.
- [42] ARAUJO, C. M. B. et al. The use of factorial design to evaluate systematic changes in the operating conditions of a three-phase separator using simulation as a tool. **Brazilian Journal of Petroleum and Gas**, v. 12, n. 1, 2018.
- [43] ARA, Behisht et al. Spectrophotometric determination of metribuzin herbicide with p-dimethylamino-benzaldehyde using factorial designs for optimization of experimental variables. **Journal of Saudi Chemical Society**, v. 20, p. S566-S572, 2016.
- [44] EJIM, Louisa N. et al. A factorial approach to understanding the effect of inner geometry of baffled meso-scale tubes on solids suspension and axial dispersion in continuous, oscillatory liquid–solid plug flows. **Chemical Engineering Journal**, v. 308, p. 669-682, 2017.
- [45] CHAN, Lydia et al. Optimisation of weld overlay cladding parameters using full-factorial design of experiment. In: **Materials Science Forum**. Trans Tech Publications, v. 880, p. 54-58, 2017.
- [46] BOX, George EP; HUNTER, J. Stuart. The 2<sup>k</sup>—p fractional factorial designs. **Technometrics**, v. 3, n. 3, p. 311-351, 1961.
- [47] GUNST, Richard F.; MASON, Robert L. Fractional factorial design. **Wiley Interdisciplinary Reviews: Computational Statistics**, v. 1, n. 2, p. 234-244, 2009.
- [48] REZENDE, Camila A. et al. Optimization of biomass pretreatments using fractional factorial experimental design. **Biotechnology for biofuels**, v. 11, n. 1, p. 206, 2018.
- [49] MA, Hao; KÖKKILIÇ, Ozan; WATERS, Kristian E. The use of the emulsion liquid membrane technique to remove copper ions from aqueous systems using statistical experimental design. **Minerals Engineering**, v. 107, p. 88-99, 2017.
- [50] LE, Giang Thi Truc; GRISDANURAK, Nurak; SREEARUNOTHAI, Paiboon. Seawater Ions Effect on the Adsorption of Cesium by Reduced Graphene Oxide Using Fractional Factorial Design. In: **Proceedings of the 7th International Conference on Informatics, Environment, Energy and Applications**, p. 77-81, 2018.
- [51] SILVA, Marta Sofia et al. Speciation of zinc in fish feed by size exclusion chromatography coupled to inductively coupled plasma mass spectrometry—using fractional factorial design for method optimisation and mild extraction conditions. **Journal of Chromatography B**, v. 1104, p. 262-268, 2019.

- [52] ADIO, Salawu Omobayo et al. Arsenic and selenium removal from water using biosynthesized nanoscale zero-valent iron: a factorial design analysis. **Process Safety and Environmental Protection**, v. 107, p. 518-527, 2017.
- [53] BOX, George EP & HUNTER, JS. Multi-factor experimental designs for exploring response surfaces. **The Annals of Mathematical Statistics**, v. 28, n. 1, p. 195-241, 1957.
- [54] CONAGIN, Armando. Delineamento composto central com duas estrelas. **Pesquisa Agropecuária Brasileira**, v. 17, n. 9, p. 1261-1269, 1982.
- [55] GHAEDI, Abdol Mohammad et al. Optimization and modeling of simultaneous ultrasound-assisted adsorption of ternary dyes using copper oxide nanoparticles immobilized on activated carbon using response surface methodology and artificial neural network. **Ultrasonics sonochemistry**, v. 51, p. 264-280, 2019.
- [56] OZ, Umut Can et al. Development and Optimization of Alendronate Sodium Loaded PLGA Nanoparticles by Central Composite Design. **Macromolecular Research**, p. 1-10, 2019.
- [57] PEREIRA, Erlon Lopes et al. Central Composite Rotatable Design for Startup Optimization of Anaerobic Sequencing Batch Reactor Treating Biodiesel Production Wastewater. **Journal of Environmental Chemical Engineering**, p. 103038, 2019.