



UEPB

**UNIVERSIDADE ESTADUAL DA PARAÍBA
CAMPUS I - CAMPINA GRANDE
CENTRO DE CIÊNCIAS E TECNOLOGIA
DEPARTAMENTO DE ESTATÍSTICA
CURSO DE GRADUAÇÃO EM ESTATÍSTICA**

LUCAS RODRIGUES DUTRA

UM TUTORIAL EM R PARA ANÁLISE FATORIAL 2^k COM E SEM REPETIÇÃO

**CAMPINA GRANDE - PB
2023**

LUCAS RODRIGUES DUTRA

UM TUTORIAL EM R PARA ANÁLISE FATORIAL 2^k COM E SEM REPETIÇÃO

Trabalho de Conclusão de Curso (Artigo) apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de bacharel em Estatística.

Orientador: Prof^a. Dr^a Ana Patrícia Bastos Peixoto de Oliveira

**CAMPINA GRANDE - PB
2023**

É expressamente proibido a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano do trabalho.

D978t Dutra, Lucas Rodrigues.
Um tutorial em R para análise fatorial 2k com e sem repetição [manuscrito] / Lucas Rodrigues Dutra. - 2023.
29 p. : il. colorido.

Digitado.

Trabalho de Conclusão de Curso (Graduação em Estatística) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2023.

"Orientação : Profa. Dra. Ana Patrícia Bastos Peixoto de Oliveira, Coordenação do Curso de Estatística - CCT. "

1. Experimento fatorial. 2. Recurso computacional. 3. Superfície de resposta. I. Título

21. ed. CDD 519.5

LUCAS RODRIGUES DUTRA

UM TUTORIAL EM R PARA ANÁLISE FATORIAL 2^k COM E SEM REPETIÇÃO

Trabalho de Conclusão de Curso (Artigo) apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de bacharel em Estatística.

Aprovado em: 22/06/2023.

BANCA EXAMINADORA



Prof.^a. Dr.^a Ana Patrícia Bastos Peixoto de
Oliveira (Orientador)

Universidade Estadual da Paraíba (UEPB)



Prof. Dr. Elias Dias Coelho Neto

Universidade Estadual da Paraíba (UEPB)



Prof. Me. Cleanderson Romualdo Fidelis

Universidade Estadual da Paraíba (UEPB)

Dedico este trabalho a Deus, meus pais e minha avó que sempre estiveram ao meu lado.

“ Você pode ter dados sem informação, mas não pode ter informação sem dados.”
(Daniel Keys Moran)

LISTA DE ILUSTRAÇÕES

Figura 1	–	Análise gráfica dos resíduos do modelo	18
Figura 2	–	Gráfico dos efeitos principais concentração e catalisador	19
Figura 3	–	Gráfico de interação entre concentração e catalisador	20
Figura 4	–	Superfície de resposta dos efeitos principais	21
Figura 5	–	Gráfico de Daniel para os efeitos dos fatores	23
Figura 6	–	Gráfico de Pareto para os efeitos estimados	24
Figura 7	–	Gráfico de barras para a relação do catalisador e da temperatura	26

LISTA DE TABELAS

Tabela 1	–	Planejamento 2 fatores com 3 repetições	12
Tabela 2	–	Planejamento 4 fatores sem repetição	13
Tabela 3	–	Planejamento 2^2 com 3 repetições	15
Tabela 4	–	Modelo de Regressão dos fatores	15
Tabela 5	–	Análise de variância	16
Tabela 6	–	Planejamento 2^4 sem repetição	22
Tabela 7	–	Significância dos efeitos	25
Tabela 8	–	Fatores de interação e a resposta	26

SUMÁRIO

1	INTRODUÇÃO	9
2	METODOLOGIA	9
2.1	Experimento Fatorial 2^k	10
2.2	Fatorial Fracionado 2^{k-p}	11
3	ANÁLISE EXPERIMENTAL	12
3.1	Planejamento com repetição via FrF2	14
3.1.1	<i>Modelo de regressão</i>	15
3.1.2	<i>Análise de variância - ANOVA</i>	16
3.1.3	<i>Pressupostos do modelo</i>	17
3.1.4	<i>Análise gráfica dos resíduos</i>	17
3.1.5	<i>Análise gráfica dos efeitos principais e de interação</i>	18
3.1.6	<i>Superfície de resposta</i>	20
3.2	Planejamento sem repetição	21
3.2.1	<i>Cálculo dos efeitos</i>	22
3.2.2	<i>Gráfico de Daniel</i>	23
3.2.3	<i>Gráfico de Pareto via método de Lenth</i>	23
3.2.4	<i>Significância dos efeitos</i>	24
4	CONCLUSÃO	27
	REFERÊNCIAS	27

UM TUTORIAL EM R PARA ANÁLISE FATORIAL 2^k COM E SEM REPETIÇÃO

Lucas Rodrigues Dutra*
Ana Patrícia Bastos Peixoto de Oliveira †

RESUMO

O experimento fatorial 2^k desempenha um papel de grande influência na identificação de problemas, sendo uma categoria direta dos experimentos fatoriais. Ao definir as variáveis, são aplicados métodos estatísticos para obter os resultados que impactam o interesse esperado. No entanto, é importante considerar que o número de fatores e repetições pode resultar em altos custos financeiros e exigir um período prolongado para análise. Felizmente, com o avanço da tecnologia computacional, há uma ampla gama de softwares disponíveis que podem contribuir significativamente para o desenvolvimento nessa área. Dentre eles, destacam-se o R, SAS, SPSS, entre outros. Essas ferramentas computacionais têm a capacidade de agilizar e aprimorar o processo de análise, tornando-o mais eficiente e acessível. Para tanto, foram feitas análises em experimentos com e sem repetição, em que, para o caso com repetições, foi analisado o efeito da concentração do reagente e a quantidade do catalisador na conversão em um processo químico. Enquanto no ensaio sem repetição, o objetivo era estudar uma reação química, desejando saber como o rendimento seria afetado. O resultado apresentado no experimento com repetição deu-se com os fatores principais concentração do reagente no nível alto e o catalisador no nível baixo como melhores respostas com relação ao rendimento do processo. Na análise realizada sem repetição após a aplicação dos testes o experimento contando com os fatores temperatura, catalisador, concentração e a interação entre a temperatura em 60°C e o catalisador “A” apresentaram os melhores resultados para obter o melhor rendimento.

Palavras-chaves: experimento fatorial; recurso computacional; superfície de resposta.

ABSTRACT

The factorial design 2^k plays a very influential role in identifying problems, being a direct category of factorial designs. When defining the variables, statistical methods are applied to obtain the results that impact the expected interest. However, it is important to consider that the number of factors and repetitions can result in high financial costs and require an extended period for analysis. Fortunately, with the advancement of computer technology, there is a wide range of software available that can significantly contribute to development in this area. Among them, stand out the R, SAS, SPSS, among others. These computational tools have the ability to streamline and improve the analysis process, making it more efficient and accessible. For that, analyzes were carried out in experiments with and without repetition, in which, for the case with repetitions, the effect of the concentration of the reagent and the amount of the catalyst in the conversion in a chemical process was analyzed. While in the non-repetition test, the objective was to study a chemical reaction, wanting to know how the yield would be affected. The result presented in the experiment with repetition occurred with the main factors reagent concentration in the high level and the catalyst in the low level as better answers in relation to the yield of the process. In the analysis carried out without repetition after the application of the tests, the experiment with the factors temperature, catalyst, concentration and the interaction between the temperature at 60°C and the catalyst “A” presented the best results for get the best yield.

Keywords: factorial experiment; computational resource; response surface.

* Aluno do curso Estatística, Depto Estatística, UEPB, Campina Grande, PB, lucas.pb.rodrigues@hotmail.com

† Prof.^a do curso Estatística, Depto Estatística, UEPB, Campina Grande, PB, anapatricia@servidor.uepb.edu.br

1 INTRODUÇÃO

O uso de métodos e técnicas estatísticas sempre foi de extrema importância em diversas áreas, inclusive na indústria, contribuindo para a tomada de decisões (Olguín e Fearn (1997)). No entanto, antes da ampla adoção da tecnologia computacional, esses estudos eram realizados manualmente, o que resultava em um tempo maior de execução e aumentava as chances de erros. Com os avanços tecnológicos, tornou-se possível desenvolver softwares estatísticos que permitem obter resultados precisos de maneira rápida e prática. Isso possibilitou aos usuários lidar com análises mais complexas, gerar gráficos e elaborar relatórios a partir de grandes conjuntos de dados. Além disso, esses softwares facilitaram a execução de tarefas com maior eficiência e confiabilidade.

No planejamento de experimentos, uma técnica bastante comum é o experimento fatorial 2^k . Essa técnica pode ser aplicada em diversas áreas, como engenharia, física, química, entre outras, com o objetivo de avaliar o efeito simultâneo de dois ou mais fatores em um determinado processo. Além disso, ela também pode ser utilizada para determinar as melhores combinações de níveis dos fatores, contribuindo para a otimização do processo. Quando todas as possíveis combinações de fatores são testadas, temos o chamado experimento fatorial 2^k completo. No entanto, em casos em que a quantidade de fatores é muito grande, utiliza-se uma variação do fatorial completo chamada fatorial fracionado. Essa técnica permite reduzir o número de ensaios necessários sem perder informações relevantes (MONTGOMERY; RUNGER, 2010).

Os experimentos fatoriais 2^k podem ser realizados com ou sem repetição. Quando são utilizados experimentos sem repetição, os recursos disponíveis são limitados, o que significa que cada combinação de fatores é testada apenas uma vez. Por outro lado, nos experimentos com repetição, há recursos suficientes disponíveis para testar cada combinação várias vezes, o que permite realizar análises estatísticas mais precisas. Durante a realização de um experimento fatorial 2^k é possível analisar os resultados usando diferentes métodos como regressão múltipla, análise de variância e superfície de resposta. A escolha da técnica depende das suposições feitas sobre os dados e dos objetivos do experimento. Um dos métodos mais utilizados é a ANOVA (análise de variância) que, segundo Govaerts et al. (2020), permite determinar se há diferenças significativas entre os tratamentos. Além da ANOVA, outras técnicas podem ser utilizadas, como regressão múltipla e superfície de resposta, dependendo das características dos dados e dos objetivos específicos do experimento (MEAD, 1990)

A partir dessas considerações, o objetivo deste trabalho foi desenvolver um tutorial utilizando o software R (R Core Team (2022)) no IDE, do inglês integrated development environment ou ambiente de desenvolvimento integrado RStudio (RStudio et al. (2020)). Esse enfoque se justifica pelo fato de que a maioria dos trabalhos realizados nessa área faz uso de aplicativos estatísticos como Minitab (Minitab (2021)) e Statistica (Statsoft et al. (2004)). Neste estudo, os experimentos fatoriais 2^k , tanto com repetição quanto sem repetição, foram empregados para investigar o comportamento dos rendimentos em dois processos químicos.

2 METODOLOGIA

A técnica de planejamento de experimentos teve suas primeiras aplicações práticas com Fisher na década de 1930, representando uma grande inovação que obteve sucesso inicialmente na experimentação agrícola. No entanto, o interesse por essa técnica se expandiu para diversas áreas, incluindo a indústria química e os processos industriais em geral. O uso dessa técnica permitiu às empresas alcançar ganhos significativos de rendimento e produtividade, trazendo benefícios substanciais para suas operações (ALMEIDA, 2007).

De fato, a técnica de planejamento de experimentos poderia ter sido aplicada há mais

tempo, mas sua difusão foi inicialmente limitada pela dificuldade em lidar com os complexos cálculos envolvidos. No entanto, com o avanço da tecnologia, programas computacionais como Minitab (Minitab (2021)), Statistica (Statsoft et al. (2004)), SPSS (IBM Corp. (2021)) e R (R Core Team (2022)) foram desenvolvidos, tornando mais acessível a resolução de problemas de experimentação. Esses softwares permitiram uma abordagem mais fácil e bem-sucedida em diferentes áreas, facilitando a aplicação da técnica de planejamento de experimentos e contribuindo para seu maior alcance e impacto (PAIVA et al., 2004).

2.1 Experimento Fatorial 2^k

Nos experimentos, é essencial estudar os efeitos de dois ou mais fatores. Dentro desse contexto, o experimento fatorial se destaca como uma das abordagens mais eficazes, pois nele todas as combinações possíveis dos níveis dos fatores são testadas, seja de forma completa ou replicada. Um exemplo comum é quando temos um fator “A” com níveis (-,+), e um fator “B” com níveis (-,+), ao organizar esses fatores em um experimento fatorial, dizemos que eles são “cruzados”. Isso significa que cada nível do fator “A” é combinado com cada nível do fator “B”, permitindo a análise conjunta dos efeitos desses fatores e possíveis interações entre eles. Essa abordagem de cruzamento de fatores no experimento fatorial é fundamental para compreender as relações entre os fatores e seus efeitos no processo em estudo.

Conforme mencionado por Nóbrega (2010), a notação 2^k indica que um experimento fatorial consiste de k fatores, cada um com dois níveis. No entanto, à medida que o valor de k aumenta, o número total de observações no experimento também aumenta. Isso ocorre porque o número de combinações possíveis entre os níveis dos fatores aumenta exponencialmente. Considerando a repetição dos tratamentos, em alguns casos pode se tornar inviável repetir todas as combinações devido a restrições de tempo e recursos econômicos. No entanto, em certos experimentos, são utilizadas repetição dos tratamentos. Nesses casos, a representação do experimento é definida como $n2^k$, em que n representa o número de repetições dos tratamentos. É importante encontrar um equilíbrio entre a quantidade de repetições e as limitações de recursos, levando em consideração os objetivos do experimento e a confiabilidade dos resultados desejada.

Nos experimentos fatoriais 2^k , os níveis de cada fator são frequentemente denominados como “baixo” e “alto”, ou “ausente” e “presente”. Eles também podem ser representados pelos sinais “-” e “+”, ou pelos números “-1” e “1”. Essa codificação é utilizada para indicar a configuração dos níveis dos fatores durante o experimento. Além disso, os fatores em um experimento podem ser classificados como quantitativos ou qualitativos. Fatores quantitativos possuem níveis representados em uma escala numérica, enquanto fatores qualitativos têm valores que podem ser separados em diferentes categorias e se distinguem por características não numéricas (BRUNS; NETO; SCARMÍNIO, 2001).

Essa diferenciação entre fatores quantitativos e qualitativos é relevante na análise dos resultados e na interpretação dos efeitos dos fatores no experimento. A escolha adequada da codificação e compreensão da natureza dos fatores são essenciais para uma correta interpretação dos resultados obtidos.

A equação geral da regressão linear múltipla é representada da seguinte forma:

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_px_p + \varepsilon, \quad (1)$$

em que, y é a variável dependente, os β 's são os parâmetros do modelo que representam o efeito de cada variável independente em y , x 's são as variáveis independentes e ε o erro experimental. O objetivo da regressão linear múltipla é estimar os valores dos coeficientes que melhor se ajustam os dados e permitir fazer previsões sobre a variável resposta.

Um exemplo da representação do modelo de regressão de um experimento fatorial contendo dois fatores pode ser escrito como:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon, \quad (2)$$

em que, y é a variável resposta, os β 's são os parâmetros desconhecidos que serão determinados os valores, x_1 é a variável que representa o fator A, x_2 é a variável que representa o fator B, $x_1 x_2$ representa a interação de dois fatores e ε o erro experimental.

Segundo Montgomery (2017), os efeitos principais, também conhecidos como efeitos de um fator, são definidos pela maneira como a resposta varia quando o nível desses fatores são alterados, mantendo constantes os níveis dos demais fatores. Em outras palavras, os efeitos principais medem o impacto individual de cada fator na resposta do experimento. Enquanto o efeito de interação ocorre quando a resposta não pode ser atribuída apenas aos efeitos principais dos fatores, mas é influenciada pela combinação específica desses fatores. Para determinar o efeito de interação, é necessário adicionar as colunas de interação, que são formadas multiplicando as colunas dos efeitos principais. A análise dos efeitos principais e das interações é essencial para entender como os fatores influenciam a resposta do experimento. Os efeitos principais nos fornecem informações sobre o impacto individual de cada fator, enquanto as interações revelam se a combinação de diferentes fatores tem um efeito conjunto significativo na resposta.

Na realização de um experimento fatorial 2^k , o primeiro passo é estimar os efeitos dos fatores e observar a distribuição dos sinais. Com base nisso, é definido o modelo inicial que será trabalhado. Geralmente, é utilizado o modelo completo, que inclui todos os efeitos principais e as interações. Em seguida, são realizados testes estatísticos, como a análise de variância (ANOVA), para verificar a significância dos efeitos principais e da interação. A partir desses testes, é possível realizar o refinamento do modelo, identificando quais variáveis não são significativas e removendo-as do modelo completo. Na análise dos resíduos, são testadas as suposições do modelo, verificando se os resíduos se distribuem de forma aleatória e se atendem às pressuposições do modelo estatístico utilizado, o que garante a validade dos resultados. Por fim, a interpretação dos resultados é realizada de forma geral, utilizando gráficos e outras ferramentas. Os gráficos podem ajudar a visualizar os efeitos dos fatores e suas interações, facilitando a compreensão dos padrões e tendências observados nos dados.

No geral, esse processo de análise em um experimento fatorial 2^k envolve a estimação dos efeitos, a realização de testes estatísticos, o refinamento do modelo, a verificação das suposições e a interpretação dos resultados com o auxílio de gráficos. Isso permite uma compreensão mais completa dos efeitos dos fatores estudados e suas implicações para o processo em questão.

2.2 Fatorial Fracionado 2^{k-p}

Existem os experimentos fatoriais fracionados, também conhecidos como experimentos incompletos. Nesses casos, apenas uma fração dos tratamentos possíveis é testada. A escolha dessa fração pode tornar a obtenção das equações um pouco mais complexa, uma vez que é a partir delas que a estrutura de confundimento entre os efeitos principais e as interações é analisada (CIRILLO, 2015).

Nos experimentos fracionados, é possível desprezar alguns efeitos de interação de ordens mais altas. Isso significa que certas combinações específicas de fatores e seus níveis não são testadas, o que reduz a complexidade do experimento. Essa abordagem é aconselhada quando se tem interesse apenas nos efeitos principais e nas interações de ordens mais baixas, e os efeitos de interação de ordens mais altas são considerados menos relevantes ou negligenciáveis. Os experimentos fatoriais fracionados são uma alternativa viável quando o número total de

tratamentos possíveis é muito grande ou quando há limitações de recursos, tempo ou custo. Eles permitem uma redução no número de ensaios necessários, facilitando a realização do experimento.

No entanto, é importante destacar que a escolha de um experimento fracionado requer cuidado na interpretação dos resultados, pois a presença de efeitos de interação de ordens mais altas que foram desprezados pode afetar as conclusões do estudo. Portanto, é fundamental considerar os objetivos do experimento, as restrições de recursos e as implicações da escolha do experimento fracionado para a análise e interpretação dos resultados.

Os experimentos fatoriais fracionados são uma boa opção quando há uma grande quantidade de fatores a serem analisados e se deseja determinar quais deles são mais importantes. Essa abordagem permite reduzir o número de ensaios necessários em comparação com um experimento fatorial completo. Os experimentos fatoriais fracionados também são adequados quando há restrições de recursos financeiros ou de tempo. Ao realizar apenas uma fração dos ensaios, é possível economizar recursos financeiros e reduzir o tempo necessário para executar o experimento.

A representação dos experimentos fatoriais fracionados é expressa de duas maneiras comuns. A primeira é na forma de $1/2^p$ frações de 2^k , em que p representa o número de frações utilizadas. Por exemplo, um experimento fracionado $1/2^3$ seria uma oitava parte de um experimento fatorial completo 2^k . A segunda forma de representação é por meio de 2^{k-p} , em que p representa o número de frações utilizadas. Por exemplo, um experimento fracionado 2^{3-1} seria metade de um experimento fatorial completo 2^k (GIESBRECHT; GUMPERTZ, 2011).

Essas representações indicam a proporção de ensaios que são realizados em relação ao experimento fatorial completo. Elas permitem uma redução significativa no número total de ensaios, enquanto ainda permitem avaliar os efeitos principais e algumas das interações mais importantes.

3 ANÁLISE EXPERIMENTAL

Neste trabalho foram analisados dois exemplos práticos. A primeira base de dados foi retirada do quadro da Figura 6.1 do livro de Montgomery (2017), no qual foi considerado uma investigação sobre o efeito da concentração do reagente e a quantidade do catalisador na conversão em um processo químico. O objetivo do experimento era determinar se os ajustes para qualquer um desses dois fatores aumentariam o rendimento. Sendo a concentração do reagente, o fator A com dois níveis de interesse 15% e 25%. E o catalisador determinado como fator B, com o nível alto denotando o uso de 2 libras do catalisador e o nível baixo denotando o uso de apenas 1 libra. O experimento é replicado três vezes, portanto, há 12 execuções.

Tabela 1 – Planejamento 2 fatores com 3 repetições

Fator Concentração(A)	Fator Catalisador(B)	Combinação	Replicação		
			I	II	III
-	-	baixo,baixo	28	25	27
+	-	alto,baixo	36	32	32
-	+	baixo,alto	18	19	23
+	+	alto,alto	31	30	29

Fonte: Montgomery (2017)

A segunda base de dados foi retirada da Tabela 3.6 do livro “Como Fazer Experimentos” dos autores Bruns, Neto e Scarmínio (2001), sua composição seguiu um planejamento fatorial 2^4 . O objetivo era estudar uma reação química, desejando saber como o rendimento seria afetado. Sendo o primeiro fator a temperatura que passa de $40^{\circ}\text{C}(-)$ para $60^{\circ}\text{C}(+)$. O fator do efeito do catalisador A(-) e B(+). O fator da concentração de um reagente, classificada nos níveis $1,0M(-)$ e $1,5M(+)$ e o fator pH do meio reacional nos níveis neutro $7(-)$ e levemente ácido $6(+)$. O experimento não foi repetido, portanto, há 16 execuções.

Tabela 2 – Planejamento 4 fatores sem repetição

Ordem	Temperatura	Catalisador	Concentração	Ph	Rendimento
1	-	-	-	-	54
2	+	-	-	-	85
3	-	+	-	-	49
4	+	+	-	-	62
5	-	-	+	-	64
6	+	-	+	-	94
7	-	+	+	-	56
8	+	+	+	-	70
9	-	-	-	+	52
10	+	-	-	+	87
11	-	+	-	+	49
12	+	+	-	+	64
13	-	-	+	+	64
14	+	-	+	+	94
15	-	+	+	+	58
16	+	+	+	+	73

Fonte: Bruns, Neto e Scarmínio (2001)

No Rstudio, há uma variedade de pacotes disponíveis com funções para o planejamento de experimentos. Entre esses pacotes, destaca-se o pacote "FrF2" desenvolvido por Grömping (2014). Esse pacote fornece a estrutura necessária para trabalhar com experimentos fatoriais, permitindo a criação de experimentos com ou sem repetição. O pacote "FrF2" oferece recursos gráficos que facilitam a avaliação dos efeitos nos experimentos fatoriais. Ele também é adequado para a realização de experimentos fatoriais fracionados, em que apenas uma fração dos tratamentos é testada.

O pacote "FrF2" funciona em conjunto com os pacotes "DoE.base" e "DoE.wrapper". Esses pacotes complementares fornecem funcionalidades adicionais e permitem uma integração eficiente das ferramentas de planejamento de experimentos. Ao utilizar o pacote "FrF2" em combinação com os pacotes "DoE.base" e "DoE.wrapper", os usuários podem se beneficiar de recursos e funcionalidades para o planejamento e análise de experimentos fatoriais, incluindo a criação de experimentos com ou sem repetição, avaliação dos efeitos e visualização gráfica dos resultados.

Inicialmente é representado o pacote de delineamento de experimentos DoE.base, o qual, cria experimento a partir de matrizes ortogonais. Além disso, ele fornece exportações para utilização de pacotes como FrF2 o qual carregar as funções que utilizam os experimentos fatoriais com dois níveis. Para carregamento da biblioteca utiliza-se o comando *library*.

```
> library (FrF2)
```

Para construção de um planejamento de experimento fatorial 2^k é necessário criar um vetor que é uma estrutura de dados do R, o qual tem como objetivo armazenar um conjunto de valores podendo ser numéricos ou do tipo caractere, o qual servirá para entrar com a função `FrF2` e armazenar o planejamento. Os seguintes argumentos precisam ser definidos:

- i) `nruns` - Número de execuções, deve ser uma potência de 2 (4 a 4096), se fornecido;
- ii) `nfactors` - É o número de fatores de 2 níveis a serem investigados;
- iii) `factor.names` - Um vetor de caracteres dos nomes dos fatores;
- iv) `replications` - Quantidade de replicas, por padrão é definido 1;
- v) `randomize` - Se TRUE (verdadeiro), o delineamento é randomizado, tido como padrão.

Existe ainda diversos outros argumentos que fazem parte do pacote, os quais são usados a depender da necessidade e objetivo do estudo, para obter uma gama com todos os argumentos e suas respectivas funções o Rstudio contém a guia ajuda(help) que é um recurso o qual permite ao usuário obter informações completas com definições, argumentações e ainda exemplos sobre funções e pacotes da linguagem R. Levando em consideração um exemplo para acessar todos os detalhes do pacote utiliza-se o comando `help`:

```
> help("FrF2")

# De outra forma

> ?FrF2
```

3.1 Planejamento com repetição via FrF2

Considerando os dados da Tabela 1 para formar o planejamento foi admitido *planej* para receber os dados a partir da argumentação definida com dois fatores (*nfactors*) e quatro execuções (*nruns*), no código tanto pode ser escrito de forma direta 4 ou como 2^2 . Os nomes dos fatores (*factor.names*) para este trabalho foram definidos como *A* e *B* com três repetições (*replications*) cada e sem aleatorização (*randomize*). Com isso é formada a matriz pronta para ser analisada.

```
> planej = FrF2(nfactors = 2,
               nruns = 2^2,
               factor.names = c("A", "B"),
               replications = 3,
               randomize = FALSE)
```

A variável resposta rendimento é armazenada com seus respectivos resultados

```
> rendimento = c(28,36,18,31,25,32,19,30,27,32,23,29)
```

Existem diversas formas para adicionar a variável resposta ao experimento, como por exemplo `planej$rendimento = rendimento`, ou então, utilizando o comando `add.response` do pacote `DoE.base` que é carregado junto com o pacote `FrF2`.

```
> planej = add.response(planej, rendimento)
```

Na Tabela 3 é apresentado por completo todas as colunas do planejamento, contando com a inclusão da resposta. As colunas de ordem e replicação são apenas anotações, já a de blocos, nesse caso, não será considerada. A partir daí, pode-se usar o planejamento gerado pela função FrF2 para realização do experimento e registro dos resultados. Baseado nos resultados, é possível ajustar um modelo para estudar os efeitos dos fatores e avaliar sua significância estatística.

Tabela 3 – Planejamento 2^2 com 3 repetições

Ordem	Replicação	A	B	Blocos	rendimento
1	1.1	-1	-1	.1	28
2	2.1	1	-1	.1	36
3	3.1	-1	1	.1	18
4	4.1	1	1	.1	31
5	1.2	-1	-1	.2	25
6	2.2	1	-1	.2	32
7	3.2	-1	1	.2	19
8	4.2	1	1	.2	30
9	1.3	-1	-1	.3	27
10	2.3	1	-1	.3	32
11	3.3	-1	1	.3	23
12	4.3	1	1	.3	29

Fonte: Elaborado pelo autor, 2023.

3.1.1 Modelo de regressão

No modelo linear é utilizado o comando *lm* e o *summary*, que indicam quanto ao modelo se os parâmetros estimados são significativos ou não, portanto, pode-se assumir que são diferentes de zero, também é retornado a medida do coeficiente de determinação R^2 e R^2 ajustado que determina o quanto da variação que está presente nos dados é explicada. Com isso é ajustado o modelo que será utilizado contendo a resposta e as variáveis independentes.

```
> modelo <- lm(rendimento ~ A * B, data = planej)
> summary(modelo)
```

A partir desse comando é possível verificar as estimativas dos parâmetros, o erro padrão associado a estimativa, a estatística t e o p-valor, para interpretação os asteriscos são apresentados para determinar se as estimativas são diferentes de zero e, quanto mais asteriscos, maior o nível de confiança.

Tabela 4 – Modelo de Regressão dos fatores

	Estimativa	Erro Padrão	tcalc	p-valor
(Constante)	27,5000	0,5713	48,135	3,84e-11 ***
A1	4,1667	0,5713	7,293	8,44e-05 ***
B1	-2,5000	0,5713	-4,376	0,00236 **
A1:B1	0,8333	0,5713	1,459	0,18278

Nota: R^2 Ajustado = 0,8666, p-valor = 0,05 . Fonte: Elaborado pelo autor, 2023.

Para o modelo de regressão apresentado na Tabela 4 as linhas representam a constante e as variáveis independentes explicadas na equação de regressão obtida, ou seja, para a segunda coluna da Tabela 4 apresenta-se os coeficientes (regressores) de cada variável independente. Em seguida, apresenta-se o erro padrão de cada coeficiente e estatística t. Na coluna p-valor indica-se a probabilidade de que a estatística t observada seja igual ou maior do que a estatística t esperada. O p-valor registrado nessa coluna tem que ser menor que o nível de significância escolhido. Na literatura estatística o valor de 5% é comumente usado como um padrão, porém pode ser modificado dependendo do contexto do problema e dos objetivos da análise.

Portanto considerando nível de significância de 5% apenas os efeitos principais concentração e catalisador foram estatisticamente significativos, para avaliarmos a qualidade do modelo, podemos usar as medidas de R^2 , que irá nos dizer a porcentagem da variação dos dados pode ser explicada pelas variáveis. O R^2 ajustado do modelo foi de 0,8666. Isso significa que 86,66% da variação do rendimento pode ser explicada pelas variáveis preditoras.

3.1.2 Análise de variância - ANOVA

A Análise de Variância, muito representada com a abreviação ANOVA, é uma análise que visa buscar a influência dos diferentes fatores na variância do experimento. O desenvolvimento do método ocorreu com Fisher (1925), com os experimentos balanceados, em que cada tratamento tem o mesmo número de repetições. Portanto, pode-se usar como uma definição para análise de variância (ANOVA), um método que decompõe a variância total e seus graus de liberdade em partes distribuídas a fatores controlados (os tratamentos), e a uma outra parte denominada de resíduo que são relativas a uma causa não controlada.

Sua composição é dada com os graus de liberdade definidos, iguais ao número de observações menos o número de parâmetros estimados. A soma de quadrados mede a variação dos dados. O quadrado médio sendo a razão entre a soma de quadrados, e a estatística F é baseada na razão dos quadrados médios. Com o modelo ajustado é utilizado as funções *aov* e *summary* para obter a anova.

```
> anov<- aov ( modelo )
> summary ( anov )
```

Tabela 5 – Análise de variância

	GL	SQ	QM	Fcalc	p-valor
A	1	208,33	208,33	53,191	8,44e-05 ***
B	1	75,00	75,00	19,149	0,00236 **
A:B	1	8,33	8,33	2,128	0,18278
Resíduos	8	31,33	3,92		

Fonte: Elaborado pelo autor, 2023.

Segundo Govaerts et al. (2020) a ANOVA é usada principalmente para analisar a influência estatística de um conjunto de fatores em uma resposta quantitativa e tendo como seu principal objetivo definir a importância e a significância de cada efeito apresentado. Com base nisso, na Tabela 5, é possível chegar a conclusão que considerando o p-valor ao nível de 5% de significância, os efeitos principais do modelo foram significativos, então, existe diferença significativa no rendimento a partir dos níveis de concentração(A) e do catalisador(B), e apenas na interação não houve diferença significativa.

3.1.3 Pressupostos do modelo

A avaliação dos pressupostos do modelo é uma etapa importante na análise estatística, com isso, a normalidade dos resíduos e homocedasticidade, são comumente verificados para garantir a adequação do modelo aos dados (FOX, 2015).

O teste de Normalidade verifica se os resíduos do modelo seguem uma distribuição normal. Existem diferentes métodos estatísticos para realizar esse teste, como o teste de normalidade de Kolmogorov-Smirnov, teste de Shapiro-Wilk ou teste de Anderson-Darling. Nesse ambiente, foi utilizado o teste Shapiro e Wilk (1965), o qual é considerado um dos mais utilizados para verificar a normalidade, ele é usado a partir da função *shapiro.test* com as seguintes hipóteses:

H_0 : os dados seguem uma distribuição normal

H_1 : os dados não seguem uma distribuição normal.

```
> shapiro.test(modelo$residuals)
```

Como p-valor do teste de Shapiro foi de 0,18, ou seja, maior que o nível de 5% de significância não rejeitamos a hipótese nula, portanto temos indícios para acreditar que os dados seguem uma distribuição normal.

O teste de homocedasticidade avalia se a variância dos resíduos é constante ao longo de todas as combinações dos níveis dos fatores. O pressuposto necessário à ser atendido na análise de variância é que os erros tenham variância comum, por isso é definido como hipóteses a serem testadas:

H_0 : A variância é constante (homocedasticidade)

H_1 : A variância não é constante (heterocedasticidade).

A homocedasticidade pode ser realizada através do teste de Breusch e Pagan (1979), carregando o pacote *lmtest* com o comando *library(lmtest)* e obtendo o resultado com a função *bptest*.

```
> bptest(modelo)
```

Realizando o teste de Breusch e Pagan foi obtido resultado de 0,24, então dado o nível adotado de 5% de significância não rejeitamos a hipótese nula, a variância presente segue constante.

3.1.4 Análise gráfica dos resíduos

A análise de diagnóstico envolve uma variedade de métodos que podem ser aplicados aos resíduos do modelo. Os resíduos representam as diferenças entre os valores observados e os valores previstos pelo modelo. Ao examinar os resíduos, é possível verificar se eles apresentam algum padrão sistemático, violações dos pressupostos do modelo ou outros comportamentos indesejáveis. Existem vários gráficos que podem ser utilizados na análise de diagnóstico, tais como apresentados na Figura 1 em que:

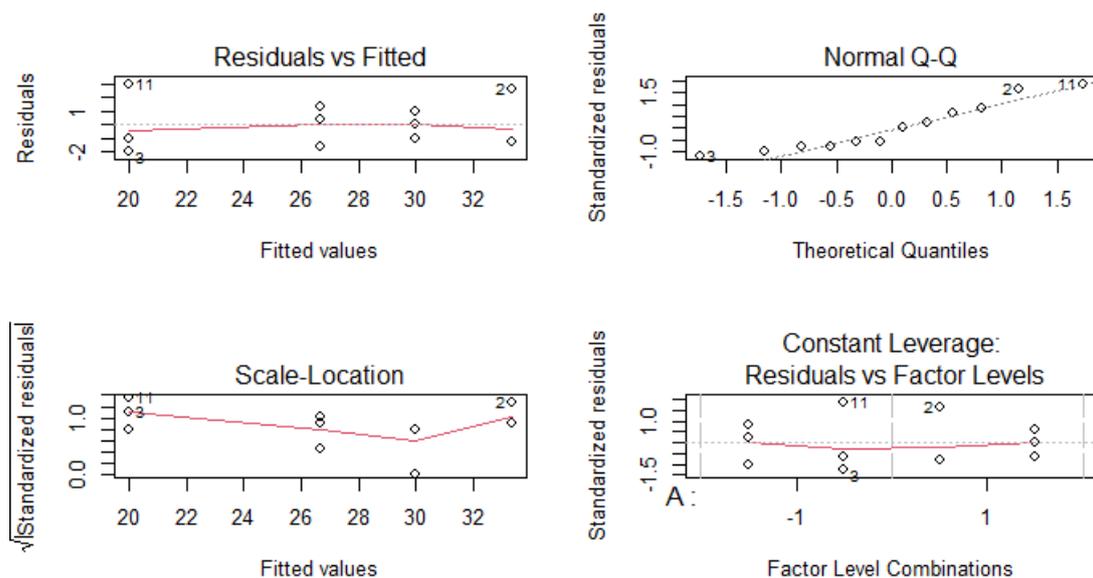
- i O gráfico - *Residual vs Fitted* (resíduos vs valores ajustados), mostra resultados de como se comporta a variância dos resíduos em relação aos valores ajustados. Sendo eficaz de modo a ter pontos espalhados aleatoriamente sem nenhum padrão óbvio como apresentado graficamente.

- ii O gráfico *-Scale-Location* (Escala-Localização) tem grande semelhança com o dos *residual vs. fitted*, mais simplifica a análise da suposição de homoscedasticidade. Ele obtém a raiz quadrada do valor absoluto dos resíduos padronizados em vez de plotar os próprios resíduos. Neste caso observou-se a homoscedasticidade de variâncias, reforçando assim o que foi apresentado no teste.
- iii O gráfico *Normal Q-Q* dos resíduos padronizados, é usado para verificar a normalidade. Tomamos como hipótese nula a normalidade dos resíduos contra os resíduos não normais. Para confirmação ainda é utilizado do teste de Shapiro. Nesse caso os resíduos estão próximos da reta os quais são considerados normais.
- iv O gráfico *- Constant Leverage* (alavancagem constante) é utilizado para identificar, caso haja, a presença de pontos influentes e outliers. Nenhum ponto foi retirado por sua influência no modelo

Com o comando `par(mfrow)` existe a facilidade de organizar vários gráficos no mesmo espaço de plotagem definindo quantos gráficos por linha e por coluna serão atribuídos, para plotar os gráficos dos resíduos utiliza-se a função `plot` do modelo. Quando terminar de usar a matriz de plotagem, é redefinido o parâmetro de plotagem de volta ao seu estado padrão.

```
> par(mfrow = c(2,2))
> plot(modelo)
> par(mfrow = c(1,1))
```

Figura 1 – Análise gráfica dos resíduos do modelo

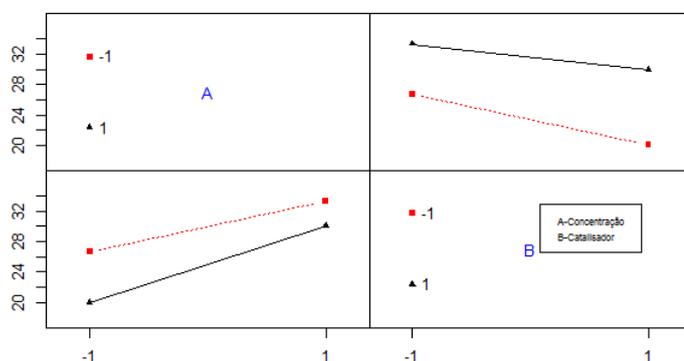


Fonte: Elaborado pelo autor, 2023.

3.1.5 Análise gráfica dos efeitos principais e de interação

O gráfico dos efeitos principais é geralmente utilizado em análises estatísticas para identificar o impacto de cada fator independente sobre a variável dependente. Ele mostra a média dos

Figura 3 – Gráfico de interação entre concentração e catalisador



Fonte: Elaborado pelo autor, 2023.

Na Figura 3 é possível constatar como já havia sido observado a partir da Tabela 5 que os efeitos de interação entre os fatores concentração e catalisador não são fortes suficientes, tendo o comportamento das linhas praticamente paralelas. Isso implica que o efeito de um fator na resposta é o mesmo independente do nível do outro fator. Se caso verificasse o cruzamento das duas linhas, haveria indícios de uma interação significativa entre os dois fatores.

3.1.6 Superfície de resposta

A superfície de resposta é uma abordagem estatística que visa encontrar a condição ótima de um processo ou sistema, considerando múltiplas variáveis independentes. Segundo Box e Draper (1987) a superfície de resposta é amplamente utilizada em experimentos e otimização de processos industriais, permitindo explorar a relação entre as variáveis independentes e a resposta desejada, sendo baseada em modelos matemáticos que descrevem a relação entre as variáveis independentes e a variável de interesse (ou resposta). Esses modelos podem ser ajustados utilizando técnicas estatísticas, como regressão linear ou regressão não linear, para representar as relações complexas entre as variáveis.

Uma vez que o modelo é ajustado, é possível visualizar a resposta esperada em diferentes combinações das variáveis independentes, criando uma superfície tridimensional ou em forma de contorno. Essa superfície mostra as condições ótimas que maximizam ou minimizam a resposta, permitindo a identificação de pontos de máximo, mínimo ou valores-alvo. A análise da superfície de resposta permite explorar a área de estudo do experimento e encontrar as melhores configurações das variáveis independentes para otimizar a resposta desejada. Além disso, também é possível realizar análises de sensibilidade para avaliar o impacto das variações nas variáveis independentes na resposta.

Para gerar a superfície de resposta foi utilizados as variáveis *AI* e *BI* e a resposta *y* em sus condições originais, seguido do ajuste do modelo, daí então foi carregado o pacote *rsm* e a função *persp* para que então fosse plotado o gráfico de superfície de resposta apresentado na Figura 4.

```
# Fatores observados
> AI<- c(15,25,15,25,15,25,15,25,15,25)

> BI<- c(1,1,2,2,1,1,2,2,1,1,2,2)
```

```

# Variável resposta - Rendimento
> y = c(28,36,18,31,25,32,19,30,27,32,23,29)

# Modelo
> modelo1<- lm(y~A1*B1)

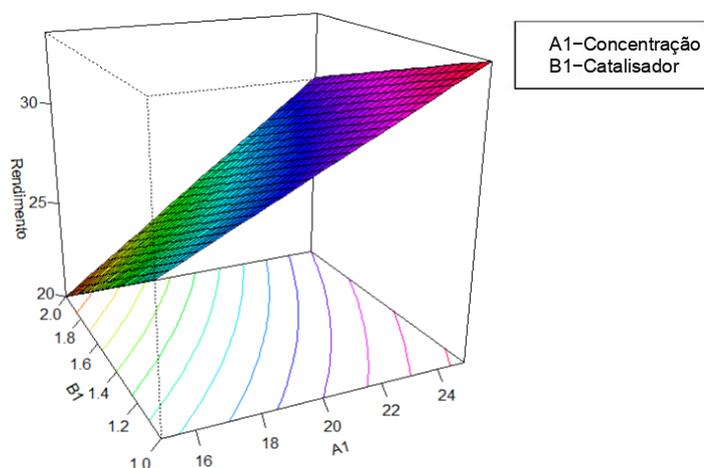
# Superfície
> library(rsm)

> persp(modelo1, B1 ~ A1, zlab = "Rendimento",
        col = rainbow(50),
        contours = "colors")

> legend("topright", legend=c("A1-Concentração",
                              "B1-Catalisador"),
        cex = 0.75, pt.cex = 0.75)

```

Figura 4 – Superfície de resposta dos efeitos principais



Fonte: Elaborado pelo autor, 2023.

Uma das grandes vantagens de utilizar a metodologia de superfície de resposta é observar como seus resultados são consistentes, mesmo sofrendo influências de erros e de condições não satisfatórias. Seu uso nos últimos anos na área de processamento ao em vez de procurar a resposta ótima se dá por encontrar a melhoria na resposta (MYERS; KHURI; CARTER, 1989).

A Figura 4 é a representação gráfica da superfície de resposta do rendimento do processo. Como contém apenas efeitos principais, ou seja, um modelo de primeira ordem a superfície é um plano. Ao verificar a disposição do modelo percebe-se que quando a concentração aumenta para o nível 25% e a quantidade do catalisador diminui para o nível de 1libra o rendimento atinge as melhores condições.

3.2 Planejamento sem repetição

A partir dos dados da Tabela 2 o planejamento foi realizado com dezesseis execuções (*nruns*) 2^4 que correspondem a quatro fatores (*nfactors*) com dois níveis cada, os fatores foram nomeados com por sua forma descrita (*factor.names*), cada combinação única dos níveis dos fatores foi testada apenas uma vez (*replications*). Além disso, a ordem das execuções foi determinada de forma pré-determinada (*randomize*) e não aleatória.

```
> planej2 = FrF2(nfactors = 4,
  nruns = 2^4,
  factor.names = c("temperatura", "catalisador", "concentração",
    "ph"),
  replications = 1,
  randomize = FALSE)
```

Armazenando em rendimento2 os valores correspondentes a variável resposta.

```
> rendimento2= c(54,85,49,62,64,94,56,70,52,87,49,64,64,94,58,73)
```

Para adicionar a variável resposta ao planejamento é utilizado o seguinte comando

```
> planej2 = add.response(planej2, rendimento2)
```

A Tabela 6, demonstra a matriz de planejamento pronta contendo todas variáveis que serão analisadas.

Tabela 6 – Planejamento 2^4 sem repetição

Ordem	Temperatura	Catalisador	Concentração	Ph	Rendimento2
1	-1	-1	-1	-1	54
2	1	-1	-1	-1	85
3	-1	1	-1	-1	49
4	1	1	-1	-1	62
5	-1	-1	1	-1	64
6	1	-1	1	-1	94
7	-1	1	1	-1	56
8	1	1	1	-1	70
9	-1	-1	-1	1	52
10	1	-1	-1	1	87
11	-1	1	-1	1	49
12	1	1	-1	1	64
13	-1	-1	1	1	64
14	1	-1	1	1	94
15	-1	1	1	1	58
16	1	1	1	1	73

Fonte: Elaborado pelo autor, 2023.

3.2.1 Cálculo dos efeitos

Para realizar análises e gráficos para experimentos não replicados, foi utilizado o pacote *unrepx* desenvolvido por Lenth e Lenth (2022) no ambiente do R. Esse pacote oferece funcionalidades específicas para tratar experimentos não replicados. Os efeitos dos fatores na resposta foram calculados a partir do método de Yates (1978), carregando o pacote com a função *library* e executando com a função *yates*.

```
> library(unrepx)
> efeitos<- yates(rendimento2)
```

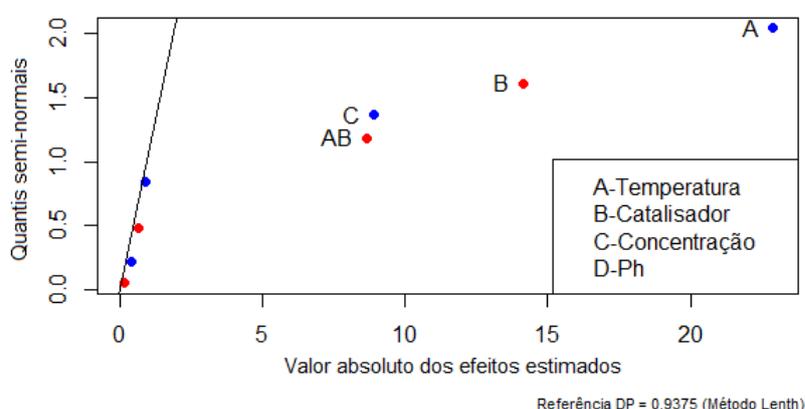
O método de Yates é um método estatístico usado para calcular os efeitos em um experimento fatorial 2^k . Geralmente ele é usado quando os dados apresentados são de baixa contagem, na maioria das vezes, são considerados k menor que cinco. Os resultados estimados dos efeitos está apresentado na primeira coluna da Tabela 7.

3.2.2 Gráfico de Daniel

A função *hnplot* do pacote *unrepx* permite gerar um gráfico semi-normal de efeitos, também conhecido como gráfico de Daniel (1959). Esse gráfico é amplamente utilizado para identificar possíveis efeitos significativos em um experimento.

```
> hnplot(efeitos , half = TRUE, method = "Lenth" , ID = > ME(efeitos))
> legend("bottomright" , legend=c("A-Temperatura" ,
    "B-Catalisador" ,
    "C-Concentração" ,
    "D-Ph"))
```

Figura 5 – Gráfico de Daniel para os efeitos dos fatores



Fonte: Elaborado pelo autor, 2023.

Segundo Daniel (1959) a interpretação gráfica se dá, inicialmente, determinando os valores absolutos dos efeitos estimados em relação aos valores críticos no momento em que um ponto é marcado fora da reta do gráfico, significa que este efeito representado é significativo e quanto mais longe o ponto estiver da reta, mais forte será sua indicação. Na Figura 5 são destacados os fatores temperatura, catalisador, concentração e a interação entre temperatura e catalisador, como os pontos que estão representados mais distante da reta, considerados estes de significância para o modelo.

3.2.3 Gráfico de Pareto via método de Lenth

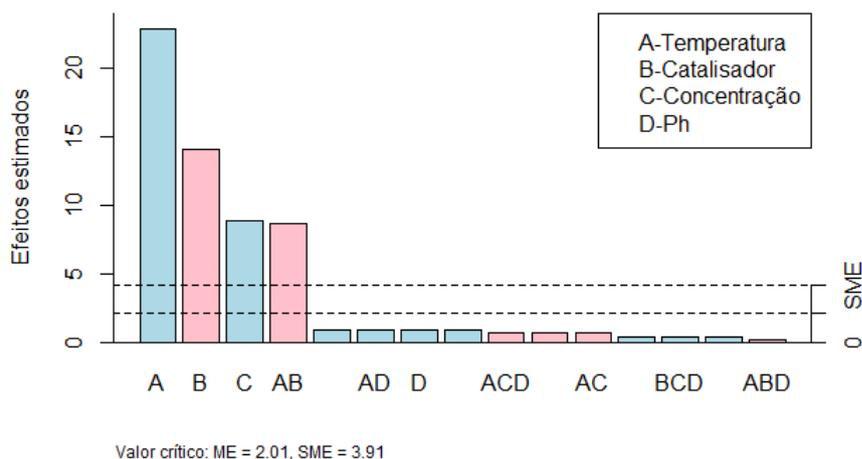
A interpretação do gráfico de Pareto permite identificar os fatores mais relevantes que afetam a resposta do experimento, permitindo uma melhor compreensão dos principais efeitos.

Uma ferramenta de análise que permite visualizar a frequência ou o impacto das ocorrências em um determinado conjunto de dados é o gráfico de Pareto, no R pode ser definido com a função *parplot* do pacote *unrepx*. Um dos métodos que podem ser usados é o de Lenth por se tratar de um método simples para análise dos fatoriais não replicados, Lenth definiu baseado no princípio da esparsidade dos efeitos, um SME erro marginal simultâneo e uma ME margem de erro como limites, e a representação dos efeitos são demonstradas a partir de um gráfico de barras.

```
> parplot(efeitos , method = "Lenth")
```

```
> legend("topright", legend=c("A-Temperatura",
                              "B-Catalisador",
                              "C-Concentração",
                              "D-Ph"))
```

Figura 6 – Gráfico de Pareto para os efeitos estimados



Fonte: Elaborado pelo autor, 2023.

O gráfico de Pareto exibe os efeitos estimados dos fatores em ordem decrescente de magnitude como apresentado na Figura 6, juntamente com as linhas verticais que representam os limites de significância sendo o ME e SME. Efeitos que ultrapassam essa linha são considerados estatisticamente significativos. Portanto é considerado significativo para o modelo da mesma forma que denotado na Figura 5 os efeitos dos fatores temperatura, catalisador, concentração e a interação entre temperatura e catalisador.

3.2.4 Significância dos efeitos

Para testar a significância dos efeitos em um experimento fatorial não replicado, é usual definir um nível de significância para o p-valor e verificar se o valor-p for menor que o limite de significância, conclui-se que o efeito é estatisticamente significativo.

A função *eff.test* facilita em um experimento não replicado a realização de inferências com relação as estimativas de efeito. Tendo como algumas de suas vantagens: a realização de testes estatísticos robustos, ser eficiente com relação ao tempo e ter flexibilidade permitindo sua aplicação a partir de diferentes métodos. No entanto, é importante destacar que os gráficos também tem suas vantagens. Eles apresentam uma visualização dos dados, permitindo uma análise detalhada além de facilitar a identificação de padrões e tendências.

```
> sig_efeitos <- eff.test(efeitos, method = "Lenth")
> sig_efeitos
```

Considerando os resultados apresentados a seguir na Tabela 7 foi possível observar que, levando em consideração tanto p-valor do erro marginal como o p-valor do erro marginal simultâneo as variáveis que tiveram efeitos significativos considerando um nível de 5% de significância foram temperatura, catalisador, concentração e a interação entre temperatura e catalisador. Todas apresentaram p-valor menor que 5%.

Tabela 7 – Significância dos efeitos

	Efeitos	Erro.Lenth	t.calc	P.valor	P.valorSimult.
A	22,875	0,9375	24,400	0,0000	0,0001
B	-14,125	0,9375	-15,067	0,0000	0,0001
C	8,875	0,9375	9,467	0,0002	0,0026
AB	-8,625	0,9375	-9,200	0,0003	0,0030
BD	0,875	0,9375	0,933	0,3257	0,9981
AD	0,875	0,9375	0,933	0,3257	0,9981
D	0,875	0,9375	0,933	0,3257	0,9981
ABC	0,875	0,9375	0,933	0,3257	0,9981
ACD	-0,625	0,9375	-0,667	0,5390	1,0000
BC	-0,625	0,9375	-0,667	0,5390	1,0000
AC	-0,625	0,9375	-0,667	0,5390	1,0000
ABCD	0,375	0,9375	0,400	0,7131	1,0000
BCD	0,375	0,9375	0,400	0,7131	1,0000
CD	0,375	0,9375	0,400	0,7131	1,0000
ABD	-0,125	0,9375	-0,133	0,9008	1,0000

Fonte: Elaborado pelo autor, 2023.

Na análise da interação significativa observada no modelo, foi gerado um gráfico de barras para melhor representação dos resultados. Admitindo as variáveis originais, criado o *data.frame* e feito a agregação dos dados com a função *aggregate* e por fim é gerado o gráfico com a função *ggplot* do pacote *ggplot2*.

```
#Varáveis nos valores originais
>Temperatura<-c(40,60,40,60,40,60,40,60,40,60,40,60,40,60,40,60)

>Catalisador<-c("A","A","B","B","A","A","B","B","A","A","B","B","A","A","B",
,"B")
Rendimento <- c(54,85,49,62,64,94,56,70,52,87,49,64,64,94,58,73)

#Criando o data frame
> dados<- data.frame(Temperatura , Catalisador , Rendimento)

> dados

#Agregando os resultados
> dados_agregados<- aggregate(Rendimento ~ Temperatura+Catalisador ,
data=dados , mean)

#Carregando o pacote gráfico
> library(ggplot2)

#Gráfico de barras
> g_barras<- ggplot(data = dados_agregados ,
aes(x = Temperatura ,
y = Rendimento ,
fill = Catalisador))
+ geom_bar(stat="identity" ,
position="dodge" ,
colour = "black")

> g_barras
```

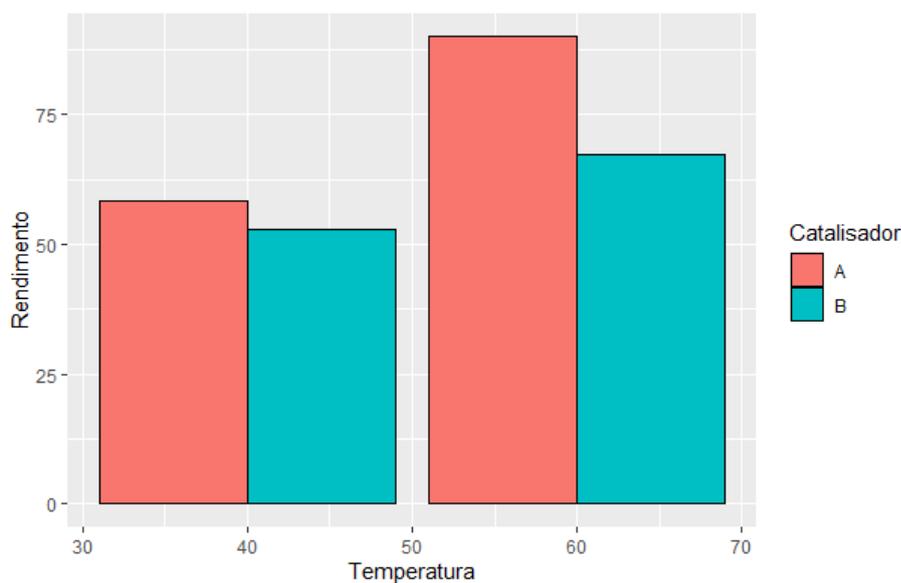
Tabela 8 – Fatores de interação e a resposta

Ordem	Temperatura	Catalisador	Rendimento
1	40	A	54
2	60	A	85
3	40	B	49
4	60	B	62
5	40	A	64
6	60	A	94
7	40	B	56
8	60	B	70
9	40	A	52
10	60	A	87
11	40	B	49
12	60	B	64
13	40	A	64
14	60	A	94
15	40	B	58
16	60	B	73

Fonte: Elaborado pelo autor, 2023.

A Tabela 8 representa a estrutura de dados criada a partir do *data frame* (dados) contendo as variáveis de interação significativa: temperatura, catalisador e a variável resposta rendimento, para que seja possível realizar a agregação dos dados e a análise gráfica.

Figura 7 – Gráfico de barras para a relação do catalisador e da temperatura



Fonte: Elaborado pelo autor, 2023.

Foi apresentado na Figura 7 um gráfico de barras para plotar as médias de cada nível do fator em relação à variável resposta, em que, foi observado a partir do momento que em conjunto o catalisador A e a temperatura em 60°C foram capazes de obter o rendimento no melhor nível.

4 CONCLUSÃO

Em suma, este estudo fornece um tutorial no desenvolvimento de uma aplicação prática para um experimento fatorial 2^k , via software R no ambiente Rstudio, construindo um código, implementando funções e gerando gráficos. Na primeira análise, após a realização de todos os testes, foi concluído que apenas os fatores principais tiveram efeito e, ajustando o aumento da concentração e a diminuição do catalisador, o rendimento do processo químico será maior. No segundo caso, no qual não houve repetição, os fatores que tiveram efeito significativo foram temperatura, catalisador, concentração e a interação entre a temperatura apresentada no nível alto 60°C e o catalisador “A” com melhores efeitos sobre o rendimento do processo. A partir disso é esperado que este trabalho sirva de auxílio para o desenvolvimento do conhecimento e prática para que futuras análises sejam realizadas.

REFERÊNCIAS

- ALMEIDA, F. R. d. V. Análises estatísticas e reconhecimento de padrão aplicados em diagnósticos de defeitos em rolamentos através da análise de vibração. 2007. Citado na página 9.
- BOX, G. E.; DRAPER, N. R. *Empirical model-building and response surfaces*. [S.l.]: John Wiley & Sons, 1987. Citado na página 20.
- BREUSCH, T. S.; PAGAN, A. R. A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the econometric society*, JSTOR, p. 1287–1294, 1979. Citado na página 17.
- BRUNS, R. E.; NETO, B. B.; SCARMÍNIO, I. Como fazer experimentos. *Editora da Unicamp, Campinas*, 2001. Citado 2 vezes nas páginas 10 e 13.
- CIRILLO, M. Otimização na experimentação: aplicações nas engenharias e ciências agrárias. *Lavras: UFLA*, 2015. Citado na página 11.
- DANIEL, C. Use of half-normal plots in interpreting factorial two-level experiments. *Technometrics*, Taylor & Francis, v. 1, n. 4, p. 311–341, 1959. Citado na página 23.
- FISHER, R. A. Applications of "student's" distribution. *Metron*, v. 5, p. 90–104, 1925. Citado na página 16.
- FOX, J. *Applied regression analysis and generalized linear models*. [S.l.]: Sage Publications, 2015. Citado na página 17.
- GIESBRECHT, F. G.; GUMPERTZ, M. L. *Planning, construction, and statistical analysis of comparative experiments*. [S.l.]: John Wiley & Sons, 2011. Citado na página 12.
- GOVAERTS, B. et al. The essentials on linear regression, anova, general linear and linear mixed models for the chemist. *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering*, Elsevier Inc., p. 431–463, 2020. Citado 2 vezes nas páginas 9 e 16.
- GRÖMPING, U. R package frf2 for creating and analyzing fractional factorial 2-level designs. *Journal of Statistical Software*, v. 56, p. 1–56, 2014. Citado na página 13.
- IBM Corp. IBM SPSS Statistics [versão 28.0]. Armonk, NY, 2021. Citado na página 10.

- LENTH, R.; LENTH, M. R. Package ‘unrepx’. 2022. Citado na página 22.
- MEAD, R. *The design of experiments: statistical principles for practical applications*. [S.l.]: Cambridge university press, 1990. Citado na página 9.
- MINITAB, L. Minitab. *Obtido de <http://www.minitab.com/en-US/products/minitab>*, 2021. Citado 2 vezes nas páginas 9 e 10.
- MONTGOMERY, D. C. *Design and analysis of experiments*. [S.l.]: John wiley & sons, 2017. Citado 2 vezes nas páginas 11 e 12.
- MONTGOMERY, D. C.; RUNGER, G. C. *Applied statistics and probability for engineers*. [S.l.]: John wiley & sons, 2010. Citado na página 9.
- MYERS, R. H.; KHURI, A. I.; CARTER, W. D. Response surface methodology: 1966–1988. *Technometrics*, Taylor & Francis, v. 31, n. 2, p. 137–157, 1989. Citado na página 21.
- NÓBREGA, M. P. *Estudo comparativo de gráficos de probabilidade normal para análise de experimentos fatoriais não replicados*. Dissertação (Mestrado) — Universidade Federal do Rio Grande do Norte, 2010. Citado na página 10.
- OLGUÍN, J.; FEARN, T. A new look at half-normal plots for assessing the significance of contrasts for unreplicated factorials. *Journal of the Royal Statistical Society Series C: Applied Statistics*, Oxford University Press, v. 46, n. 4, p. 449–462, 1997. Citado na página 9.
- PAIVA, A. P. d. et al. Estudo da minimização de erro nas medições de concentração de emulsões por titração karl-fisher utilizando-se projeto de experimentos. Universidade Federal de Itajubá, 2004. Citado na página 10.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2022. Disponível em: <<https://www.R-project.org/>>. Citado 2 vezes nas páginas 9 e 10.
- RSTUDIO, T. et al. Rstudio: integrated development for r. *Rstudio Team, PBC, Boston, MA URL <http://www.rstudio.com>*, 2020. Citado na página 9.
- SHAPIRO, S.; WILK, M. *An analysis of variance test for normality (complete samples)*. *Biometrika*, 52 (3–4), 591–611. 1965. Citado na página 17.
- STATSOFT, I. et al. Statistica (data analysis software system). *Version*, v. 7, p. 1984–2004, 2004. Citado 2 vezes nas páginas 9 e 10.
- YATES, F. *The design and analysis of factorial experiments*. [S.l.]: Imperial Bureau of Soil Science Harpenden, UK, 1978. Citado na página 22.

AGRADECIMENTOS

Primeiramente agradeço a Deus, por ter me dado força e sabedoria para seguir durante todo período desta graduação.

Aos meus pais, Marliete Rodrigues e Severino Ramos, por todo amor e incentivo nos momentos mais difíceis, e todo sacrifício para me dar uma boa educação.

A minha orientadora Prof^a Dr^a Ana Patricia Bastos Peixoto de Oliveira por toda dedicação durante todo esse processo, sem ela não teria conseguido, a qual tenho uma grande admiração.

Aos amigos da universidade Ana Maria, Eduardo, Débora, Fagna, Filype, Zelma e a todos os amigos que passaram durante o curso, por todos os momentos de aprendizado e companheirismo.

Aos meus amigos Edna, Everaldo, Débora, Fagner, e Wendell que sempre estiveram junto apoiando e ajudando.

Aos meus secretários, Geraldo, Júnior e Sérgio, minha gratidão por ter me possibilitado quando precisei chegar atrasado por assistir uma aula online pela manhã ou sair mais cedo para pegar o ônibus, deixando minhas demandas no setor.

A todos os professores do Departamento de Estatística, por todas aulas e ensinamentos, meu muito obrigado a todos.