

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
DEPARTAMENTO DE QUÍMICA

MATHEUS ALMEIDA FÁVERO

**PRINCÍPIOS DO PLANEJAMENTO DE EXPERIMENTOS (DOE) NO
CONTEXTO DE ANALYTICAL QUALITY BY DESIGN**

São Carlos – SP
2025

MATHEUS ALMEIDA FÁVERO

**PRINCÍPIOS DO PLANEJAMENTO DE EXPERIMENTOS (DOE) NO
CONTEXTO DE ANALYTICAL QUALITY BY DESIGN**

Trabalho de conclusão de curso apresentado ao
Departamento de Química da Universidade Federal de
São Carlos para obtenção do título de bacharel em
Química.

Orientador: Prof. Dr. Edenir Rodrigues Pereira Filho

São Carlos – SP
2025

**FUNDAÇÃO UNIVERSIDADE FEDERAL DE SÃO CARLOS****DEPARTAMENTO DE QUÍMICA - DQ/CCET**

Rod. Washington Luís km 235 - SP-310, s/n - Bairro Monjolinho, São Carlos/SP, CEP 13565-905

Telefone: (16) 33518206 - <http://www.ufscar.br>

DP-TCC-FA nº 19/2025/DQ/CCET

Graduação: Defesa Pública de Trabalho de Conclusão de Curso**Folha Aprovação (GDP-TCC-FA)****FOLHA DE APROVAÇÃO****MATHEUS ALMEIDA FAVERO****PRINCÍPIOS DO PLANEJAMENTO DE EXPERIMENTOS (DOE) NO CONTEXTO DE ANALYTICAL QUALITY BY DESIGN****Trabalho de Conclusão de Curso****Universidade Federal de São Carlos – Campus São Carlos**

São Carlos, 03 de julho de 2025

ASSINATURAS E CIÊNCIAS

Cargo/Função	Nome Completo
Orientador	Prof. Dr. Edenir Rodrigues Pereira Filho
Membro da Banca 1	Prof. Dr. Manoel Gustavo Petrucelli Homem
Membro da Banca 2	Me. Renato Cesar De Souza



Documento assinado eletronicamente por **Ricardo Samuel Schwab, Professor(a)**, em 04/07/2025, às 09:49, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site <https://sei.ufscar.br/autenticacao>, informando o código verificador **1909363** e o código CRC **045922F6**.

Referência: Caso responda a este documento, indicar expressamente o Processo nº 23112.001933/2024-38

SEI nº 1909363

Modelo de Documento: Grad: Defesa TCC: Folha Aprovação, versão de 02/Agosto/2019

AGRADECIMENTOS

Agradeço, em primeiro lugar, à minha família, que mesmo com todas as dificuldades, me apoiou e apoia nos percursos que decido fazer, com todos os desvios que aparecem. Esse agradecimento vai especialmente à minha irmã Sabrina, que desde antes da matrícula, me deu tranquilidade para que eu pudesse me dedicar ao que realmente queria.

Aos meus amigos, com quem compartilhei tempo, conversas e risadas dentro e fora da faculdade, e que nunca esquecerei dos nossos momentos diários no RU. Com vocês, se aventurar por uma cidade nova, distante de tudo que eu conhecia, se tornou uma experiência muito mais leve e divertida. Em especial, agradeço aos meus amigos mais próximos: Ana Júlia, Débora, Diego, Érika, Fernanda, Laís, Leonardo, Letícia, Marcos Vinícius, Pamela, Paula, Vitória.

Aos meus companheiros de casa, que me permitiram uma vivência caótica, mas sempre engraçada, e que me permitiram um senso de união que levo até hoje.

À minhas colegas de trabalho, Vanessa e Julia, que me ensinaram muito e se tornaram uma família nos anos que estivemos juntos. Foram as duas que me instigaram a entender mais sobre regulamentos e estatística, que serviu como base para iniciar este trabalho.

Ao meu orientador, Edenir, que me recebeu de portas abertas, mas me manteve com os pés no chão, e que me ajudou da melhor forma a tornar este trabalho possível.

A Gabrielle Janasi, que me ajudou a acreditar e a enfrentar este tão temido trabalho.

E a mim mesmo, por não desistir.

RESUMO

A indústria farmacêutica tem priorizado, cada vez mais, a qualidade no desenvolvimento de produtos, não somente no lançamento, mas ao longo de todo o ciclo de vida do fármaco, com maior foco na eficácia do tratamento e segurança dos pacientes. Para atender a uma sociedade com demandas crescentes e urgentes, mostra-se necessário um sistema que permita melhorias ágeis e com poucos impedimentos. A *Quality by Design* (QbD), abordado neste trabalho, é uma ferramenta que permite a indústria farmacêutica acompanhar com agilidade as necessidades da saúde humana. Com o surgimento e a reformulação de produtos, é imprescindível um modelo eficaz de desenvolvimento fabril e analítico, que permita uma maior confiabilidade durante a aprovação regulatória. O Planejamento de Experimentos (DoE – *Design of Experiments*) apresenta uma alta viabilidade como uma ferramenta estatística para otimizar processos e viabilizar uma produção industrial robusta e eficiente. Dada a ampla gama de aplicações da QbD, surgiram subcategorias específicas, sendo a AQbD (Analytical Quality by Design) a abordagem dedicada à aplicação dessa ferramenta no desenvolvimento analítico. Este trabalho tem como objetivo a apresentação do DoE aplicado em um contexto de AQbD, apresentando aplicações e abordagens no desenvolvimento de fármacos de modo a se manter com altos padrões de qualidade, de maneira mais eficaz e com economia de recursos.

Palavras-chave: *quality by design*, planejamento de experimentos, quimiometria, desenvolvimento analítico, química analítica, indústria farmacêutica.

ABSTRACT

The pharmaceutical industry has increasingly prioritized quality in product development, not only at the time of launch but throughout the entire drug lifecycle, with a greater focus on treatment efficacy and patient safety. To meet the growing and urgent demands of society, it is necessary to implement a system that enables agile improvements with minimal obstacles. Quality by Design (QbD), addressed in this work, is a tool that allows the pharmaceutical industry to respond effectively to evolving healthcare needs. With the emergence and the reformulation of products, an effective model for both manufacturing and analytical development is essential to ensure greater reliability during regulatory approval. Design of Experiments (DoE) has shown high viability as a statistical tool to optimize processes and enable robust and efficient industrial production. Given the wide range of QbD applications, specific subcategories have emerged, with AQbD (Analytical Quality by Design) being the approach focused on applying this framework to analytical development. This work aims to present the application of DoE within the QbD framework, highlighting its applications and approaches in drug development in order to maintain high quality standards, while increasing efficiency and reducing resource consumption.

Keywords: quality by design, design of experiments, chemometrics, analytical development, analytical chemistry, pharmaceutical industry.

ÍNDICE DE ILUSTRAÇÕES

Figura 1 – Ciclo de vida do procedimento analítico.....	17
Figura 2 – Exemplo de diagrama de Ishikawa para avaliação de risco para definição de parâmetros do procedimento analítico com potencial impacto na seletividade.	20
Figura 3 – Visão geral dos estágios do planejamento experimental para aplicação em DoE..	22
Figura 4 – Um sistema que relaciona as variáveis de entrada com as saídas (respostas).....	22
Figura 5 – Interpretação geométrica em um planejamento fatorial 2^2 dos efeitos (a) principal de vazão de solvente, (b) principal de composição da fase móvel e (c) da interação entre as variáveis.....	26
Figura 6 – Interpretação geométrica em um planejamento fatorial 2^3 dos efeitos (a) principais e (b) das interações de segunda ordem entre as variáveis.	28
Figura 7 – Exemplos de efeitos confundidos em uma distribuição de níveis nas corridas de um planejamento fatorial fracionário 2^{4-1}	30
Figura 8 – Exemplos mais comuns de superfícies de resposta para duas variáveis com mapa de calor, de (a) equação linear; (b) equação quadrática com ponto máximo; (c) com ponto mínimo e (d) com ponto de sela.	33
Figura 9 – Exemplos de (a) superfície de resposta, (b) gráfico de contorno e (c) região de concepção operacional do método (MODR) para resultados $\geq 80\%$	35
Figura 10 – (a) Superfície de regressão e (b) gráfico de contorno para um modelo de duas variáveis sem termos de interação e quadráticos.....	38
Figura 11 – (a) Superfície de regressão e (b) gráfico de contorno para um modelo de duas variáveis com termos de interação entre as variáveis.....	39
Figura 12 – (a) Superfície de regressão e (b) gráfico de contorno para um modelo de duas variáveis com termos quadráticos e de interação entre as variáveis.....	40
Figura 13 – Exemplo gráfico da aproximação entre os pontos experimentais e a equação de reta que define o sistema com mínima distância entre todos os pontos.	41
Figura 14 – Exemplo genérico de matrizes de respostas, variáveis, coeficientes e resíduos para i variáveis e n experimentos.	43
Figura 15 – Demonstração dos cálculos para obter as matrizes (a) X^tX , (b) $(X^tX)^{-1}$, (c) X^tY e (d) $B = (X^tX)^{-1} X^tY$ de uma regressão linear múltipla de 2 variáveis sem termos de interação ou quadráticos.....	44
Figura 16 – Exemplo de estrutura das matrizes Y , X , B e E para um planejamento de (a) uma variável, (b) duas variáveis sem interação e termo quadrático e (c) duas variáveis.....	46

Figura 17 – Decomposição da variância das respostas experimentais contra a média ($y_i - \bar{y}$) em termos da variância das respostas previstas contra a média ($y_i - \bar{y}$) e das respostas experimentais contra as respostas previstas ($y_i - \hat{y}_i$) em uma equação linear.	48
Figura 18 – Curva de distribuição F para sistema com 95% de grau de confiança.....	51
Figura 19 - Demonstração dos cálculos para obter os valores do intervalo de confiança dos coeficientes de uma regressão linear.	55
Figura A1 – Pontos experimentais para geração do planejamento composto central (a) para duas variáveis e (b) para três variáveis.....	65
Figura A2 - Pontos experimentais para geração do planejamento Doehlert (a) para duas variáveis e (b) para três variáveis, com as possíveis projeções do sólido no plano.	66
Figura A3 – Exemplo de movimentação de níveis em um planejamento Doehlert de duas variáveis.....	67
Figura A4 - Pontos experimentais para geração do planejamento Box-Behnken de três variáveis.....	68
Figura A5 – Exemplo de matriz para o Método dos Mínimos Quadrados Ponderados contendo (a) as variâncias de cada ponto experimental e (b) os pesos que cada ponto representa.....	70

ÍNDICE DE QUADROS

Quadro 1 – Exemplo de Perfil Analítico Alvo (ATP) para determinação do teor de um Insumo Farmacêutico Ativo (IFA) no medicamento.	18
Quadro 2 – Exemplo de planejamento fatorial de experimentos em um caso 2^2	24
Quadro 3 – Exemplo de planejamento fatorial de experimentos em um caso 2^3	27
Quadro 4 – Quantidade de experimentos e efeitos de acordo com a quantidade de variáveis (k) em um planejamento fatorial completo 2^{kc}	29
Quadro 5 – Distribuição de níveis nas corridas de um planejamento fatorial fracionário 2^{4-1}	30
Quadro 6 – Quantidade de parâmetros e experimentos entre os Planejamentos Composto Central, Doehlert e Box-Behnken de acordo com a quantidade de variáveis.	37
Quadro 7 – Representação teórica dos valores avaliados na tabela ANOVA.	49
Quadro A1 – Exemplos de faixas reportáveis para procedimentos analíticos comuns.	62

LISTA DE ABREVIATURAS E SIGLAS

ANOVA	Análise de Variância / <i>Analysis Of Variance</i>
ANVISA	Agência Nacional de Vigilância Sanitária
AQbD	<i>Analytical Quality by Design</i>
ATP	Perfil Analítico Alvo / <i>Analytical Target Profile</i>
BBD	Planejamento Box-Behnken / <i>Box-Behnken Design</i>
CCD	Planejamento Composto Central / <i>Central Composite Design</i>
CQA	Atributos Críticos de Qualidade / <i>Critical Quality Attributes</i>
DD	Planejamento Doehlert / <i>Doehlert Design</i>
DoE	Planejamento de Experimentos / <i>Design of Experiments</i>
FMEA	Análise de Modos de Falhas e Efeitos / <i>Failure Mode and Effect Analysis</i>
GL	Graus de liberdade
H₀	Hipótese nula
HPLC	Cromatografia Líquida de Alta Eficiência / <i>High-Performance Liquid Chromatography</i>
IC	Intervalo de Confiança
ICH	<i>International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use</i>
IFA	Insumo Farmacêutico Ativo
LOD	Limite de Detecção / <i>Limit of Detection</i>
LOQ	Limite de Quantificação / <i>Limit of Quantification</i>
MMQ	Método dos Mínimos Quadrados
MMQO	Método dos Mínimos Quadrados Ordinários
MMQP	Método dos Mínimos Quadrados Ponderados
MODR	Região de Concepção Operacional do Método / <i>Method Operable Design Region</i>
MQ	Média Quadrática
OVAT	Uma-Variável-por-Vez / <i>One-Variable-At-a-Time</i>
QbD	<i>Quality by Design</i>
QTPP	Perfil-Alvo de Qualidade do Produto / <i>Quality Target Product Profile</i>
RDC	Resolução da Diretoria Colegiada
RMN	Espectroscopia de Ressonância Magnética Nuclear
SQ	Soma Quadrática
SST	Teste de Adequação do Sistema / <i>System Suitability Test</i>
WHO	<i>World Health Organization</i>

SUMÁRIO

1. INTRODUÇÃO.....	12
2. OBJETIVOS.....	16
3. DESENVOLVIMENTO	17
3.1. Perfil analítico alvo (ATP – <i>Analytical Target Profile</i>)	17
3.2. Seleção de tecnologia	19
3.3. Análise de risco	19
3.4. Identificação dos valores e/ou intervalos dos parâmetros	20
3.4.1. Triagem.....	22
3.4.1.1. Planejamento Fatorial completo	23
3.4.1.2. Planejamento Fatorial fracionário	29
3.4.2. Otimização.....	32
3.4.2.1. Superfície de resposta.....	32
3.4.3. Ajuste do modelo matemático	37
3.4.3.1. Regressão Linear Múltipla.....	37
3.4.3.2. Método dos Mínimos Quadrados	41
3.4.3.3. Análise de variância (ANOVA).....	47
3.4.3.4. Teste F e p-valor	51
3.4.3.5. Coeficiente de determinação e coeficiente de correlação.....	53
3.4.3.6. Intervalo de confiança	54
3.5. Definição da Estratégia de Controle do Procedimento Analítico.....	55
3.6. Validação	56
4. CONSIDERAÇÕES FINAIS	56
REFERÊNCIAS BIBLIOGRÁFICAS	58
APÊNDICE 01 – ATRIBUTOS DE DESEMPENHO	61
APÊNDICE 02 – PLANEJAMENTOS DE OTIMIZAÇÃO	65
APÊNDICE 03 – MÉTODO DOS MÍNIMOS QUADRADOS PARA CASOS COM DADOS HETEROCEDÁSTICOS.....	69

1. INTRODUÇÃO

Antes de um fármaco ser comercializado, é preciso garantir que o produto está de acordo com a necessidade de um medicamento de qualidade, eficaz e, principalmente, que traga segurança ao paciente. Esses são requisitos muito importantes para a indústria farmacêutica e todos os seus processos são planejados e realizados para garantir essa qualidade ao produto destinado à venda.

Para que o fármaco tenha a garantia de que irá cumprir com a especificação e que estará íntegro para o consumo pelo paciente, há sempre um extenso projeto que é realizado por meses ou até anos. Esse projeto engloba diversas etapas fabris, analíticas e regulatórias, com participação de um time multidisciplinar envolvido na tomada de decisão. O objetivo é garantir a qualidade e segurança do produto. Este projeto passa desde uma elaboração planejada do produto e testes do método de fabricação, até análises em diversos intervalos de tempo durante e após a produção.

Durante todo esse processo temos uma grande quantidade de análises, com procedimentos analíticos que devem apresentar muitos detalhes. O procedimento analítico é o que refere às etapas necessárias para realizar em cada teste analítico e deve ser detalhado o suficiente para que um analista consiga realizá-lo, com os devidos treinamentos. As análises no produto variam de acordo com a forma farmacêutica (sólido, líquido, injetável, entre outros) e, em todos os casos, os testes mais importantes são os de doseamento e impurezas relacionadas. Esses testes são realizados em produtos de todas as formas farmacêuticas, pois são justamente as análises que apresentarão os resultados da quantidade de Insumo Farmacêutico Ativo (IFA). O IFA é o responsável pela ação do medicamento no corpo, e é necessário identificar se há a degradação deste material ao longo do tempo e/ou condições extremas.

O desenvolvimento de métodos para doseamento e impurezas de um composto pode se apresentar desafiador, pois precisa forçar uma degradação com condições de stress extremas. Essas condições podem estar relacionadas com alta temperatura e umidade. Assim, o método em questão deve garantir desde sua elaboração que é capaz de determinar a degradação do fármaco em seu tempo de prateleira (TOME, 2019).

Pensando nisso, apenas o resultado da análise não fornece subsídios suficiente para garantir a qualidade e segurança do paciente. Para que os resultados apresentados pelas análises possam ser confiáveis para liberação do fármaco para comercialização, é necessário ter plena

confiança nos métodos utilizados para estas análises. Uma das características que mais atesta sobre a qualidade do método analítico é sua robustez, que pode ser definida como “uma medida da capacidade de permanecer inalterada por pequenas, mas deliberadas variações nos parâmetros do método e provém uma indicação da sua confiabilidade durante o uso de rotina” (ICH Q14, 2023). Para alcançar essa indispensável confiança, é realizado um estudo do método analítico, também conhecido como validação, para comprovar que a análise fornece resultados fidedignos e de qualidade e definir um *design space* para o procedimento. *Design space* pode ser definido como uma “combinação e interação multidimensionais de atributos de material e parâmetros de processo que garantem qualidade” (ICH Q8(R2), 2009). Esse parâmetro é alcançado em diversas etapas com foco em obter maior compreensão sobre todos os processos de análise e quais atributos e parâmetros que podem interferir ou causar riscos para o resultado. Em diversas situações são consideradas a qualidade de reagentes, ou condições operacionais como a vazão de solvente em um sistema cromatográfico. Para métodos analíticos, também pode ser afirmado que o *design space* inclui qualquer combinação de variáveis que demonstrou garantir qualidade nos dados produzidos pelo método analítico (TOME, 2019).

As validações na indústria farmacêutica, tanto de processos quanto de métodos, são regidas por diversos órgãos regulatórios, cada um com sua área de atuação e escopo. Os medicamentos devem ser submetidos para concessão de registro no órgão regulatório do país onde serão comercializados, com descrições detalhadas sobre seu processo produtivo, especificações e um relatório compilando os resultados obtidos no processo de lotes comerciais de validação, de acordo com a necessidade. Este registro assegura que a empresa comprovou estar apta e com condições de fornecer produtos de qualidade, de forma consistente, com segurança para a saúde e bem-estar do paciente (MOURE, 2023).

Para compostos químico-farmacêuticos em território nacional temos a Agência Nacional de Vigilância Sanitária (ANVISA) como órgão regulatório brasileiro e a *International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use* (ICH) e *World Health Organization* (WHO) em jurisdição global. Entre estes órgãos, um assunto em implementação extensiva para novas validações é a *Quality by Design* (QbD). A QbD é uma abordagem a ser utilizada durante a elaboração de novos produtos e métodos analíticos e é definida pela ICH como uma abordagem científica e sistemática ao desenvolvimento que começa com objetivos pré-definidos e enfatiza o entendimento do processo e do produto baseado em ciência sólida e gerenciamento de riscos de qualidade. Assim, o objetivo da utilização da QbD é reduzir a variação, permitindo estratégias de controle (ICH

Q8(R2), 2009). Como a QbD é uma ferramenta de desenvolvimento de métodos que pode ser utilizada em diversos processos farmacêuticos, podemos vê-la em uso nos processos de fabricação, de análises e até mesmo da preparação de amostras ou avaliação de estudos de estabilidade, porém neste trabalho o foco será no desenvolvimento de métodos analíticos, referido como *Analytical Quality by Design* (AQbD).

Nos órgãos nacionais, também há guias para o tratamento estatístico da validação analítica. Para a ANVISA, este se refere ao Guia 10 de 2017, elaborado para recomendar modelos e procedimentos estatísticos para cumprimento da Resolução RDC nº 166/2017. Nele, são apresentados alguns testes estatísticos para análise dos dados experimentais, porém também são aceitos outros, desde que sejam provados adequados.

Para entender a melhoria que o uso da AQbD pode proporcionar, é melhor entender primeiro como é feita a validação tradicionalmente. Em casos não otimizados pelos conceitos da AQbD, também conhecidos como abordagem tradicional, a análise é feita somente no produto final, após todo o processo de fabricação, com baixa compreensão sobre o processo e seus parâmetros críticos (**MOURE, 2023**). Usando uma análise por Cromatografia Líquida de Alta Eficiência (HPLC) como exemplo, normalmente os métodos analíticos são desenvolvidos com abordagens de tentativa e erro ou abordagem One-Variable-At-a-Time (OVAT), onde um parâmetro é variado em experimentos consecutivos até alcançar um parâmetro desejável de confiança no resultado, como a resolução entre os picos cromatográficos. Em casos com muitos parâmetros que afetam a análise, pode ser que o analista verifique o impacto de cada parâmetro, mesmo os com baixo impacto no método, pois é deficiente na definição prévia de quais são os parâmetros mais relevantes e críticos, aumentando muito a quantidade de experimentos e, conseqüentemente, o tempo necessário para o desenvolvimento de novos métodos. Por outro lado, para reduzir a quantidade de experimentos, o analista pode relevar a interação entre variáveis que podem acabar impactando no resultado. Além disso, pode ser que não seja levado em consideração o uso de diferentes equipamentos e diferentes analistas entre o laboratório de desenvolvimento e o de rotina. Com isso, é possível concluir que o método não terá robustez, pois qualquer alteração pequena nos parâmetros iniciais durante o uso na rotina que não foram avaliados, tornará o resultado desconhecido e sem confiabilidade (**TOME, 2019**).

Por isso é dito que métodos desenvolvidos por OVAT possuem alto risco e, avaliando esses aspectos pelo prisma regulatório, o procedimento não terá flexibilidade para que possa justificar uma alteração de parâmetros sem a necessidade de revalidação e nova submissão para o órgão fiscalizador. Como afirmado anteriormente, estes processos geralmente levam meses

ou até anos, ocasionando um impacto muito grande no processo analítico e a depender, até mesmo no processo fabril.

Os problemas citados do OVAT, como o excesso de experimentos com baixa eficiência na compreensão do método, além da falta de flexibilidade por não levar em conta variações no ambiente, poderiam ser mitigados pela criação do *design space*, avaliando a aplicabilidade do método, com foco nos parâmetros e aplicações. Assim, são evitados diversos riscos com o uso de um conhecimento prévio sobre o composto e/ou experiência do analista ao elaborar um plano da validação. Com um *design space* bem definido, alterações dentro do intervalo estipulado não são consideradas como uma mudança do método em si, isentando de uma revalidação completa do método.

A criação deste *design space* é um dos pontos críticos da aplicação da AqBd. A AqBd é uma abordagem sistemática para o desenvolvimento de métodos analíticos que busca sempre a maior compreensão do produto e da relação entre o desempenho do método analítico e suas variáveis, utilizando de um estudo com base em ciência e gerenciamento de riscos aplicável por todo o ciclo de vida do método analítico (**ICH Q8(R2), 2009**) (**ICH Q14, 2023**). Para isso, é necessário montar um fluxo com metas bem definidas do que será esperado e desejado durante o desenvolvimento do método, para que possa aproveitar da experiência prévia de analistas. Além disso, as análises de risco são importantes para identificar uma série de etapas que permitirão aumentar o conhecimento sobre o produto, cumprindo com as metas ou atualizá-las conforme necessário.

Quando realizado de maneira correta, a AqBd fornece um alto grau de robustez ao processo, reduzindo resultados fora de especificação e fora de tendência, e que acarreta também em uma maior flexibilidade regulatória, por possuir margens de trabalho que não necessitam de uma revalidação e alteração do processo pós-registro, caso seja implementado um *design space* de qualidade.

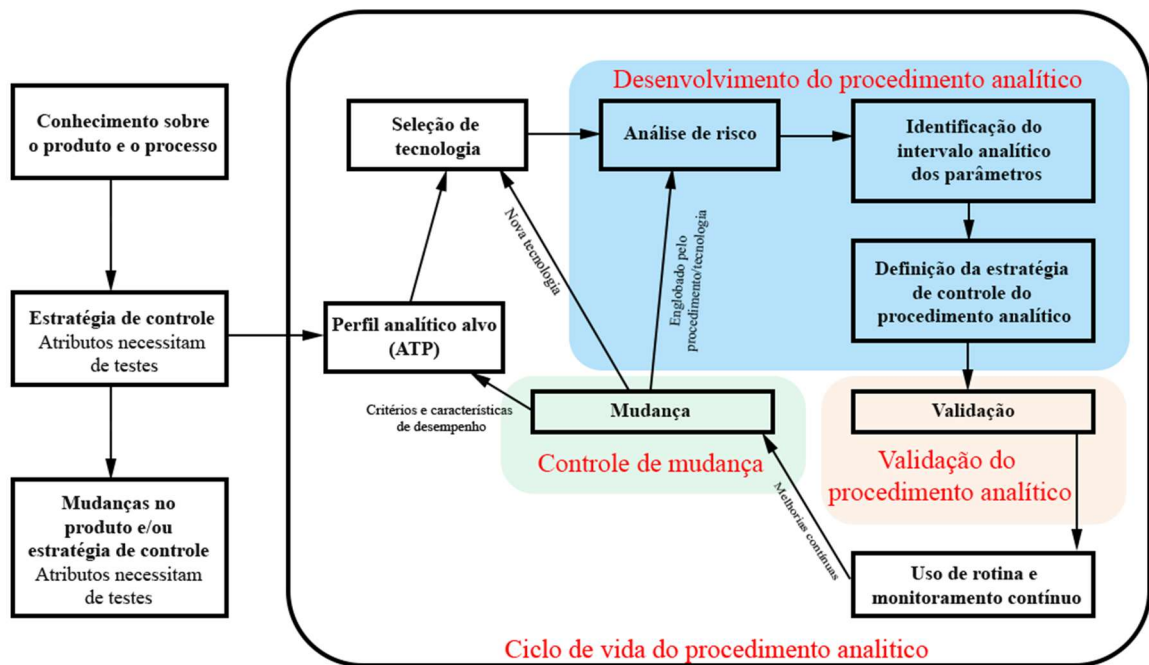
2. OBJETIVOS

O presente trabalho teve como objetivo descrever e avaliar, a partir do embasamento teórico e empírico, o uso do sistema de AQBd (*Analytical Quality by Design*) e seus principais requisitos e indicadores, com foco no DoE (*Design of Experiments*), para o desenvolvimento de novos métodos analíticos na indústria farmacêutica.

3. DESENVOLVIMENTO

As etapas da AQBd podem ser divididas, como visto na Figura 1, em criação do Perfil Analítico Alvo (ATP – *Analytical Target Profile*); seleção de tecnologia; análise de risco; identificação do intervalo analítico dos parâmetros; definição da estratégia de controle; e validação do método analítico (ICH Q14, 2023). As etapas após a validação são de uso de rotina do método e eventuais melhorias que possam ser realizadas, que ocasionam no retorno ao ciclo de vida a partir do ATP, e a de controle de mudança, que trata de um assunto com caráter regulatório e que não será discorrido neste trabalho.

Figura 1 – Ciclo de vida do procedimento analítico



Fonte: Elaborado pelo autor. Adaptado de ICH Q14, 2023.

3.1. Perfil analítico alvo (ATP – *Analytical Target Profile*)

O ATP consiste em uma descrição do objetivo do procedimento analítico, com detalhes dos atributos do produto a serem analisados e as características relevantes de desempenho junto aos seus critérios de aceitação.

Antes da criação do ATP, é necessário ter um conhecimento geral do produto e do processo e de estratégias de controle de atributos que devem ser avaliados durante a produção, garantindo a qualidade. As características do processo e do produto são estudadas de forma aprofundada para aumentar a compreensão, e geralmente inicia pela criação de um Perfil-Alvo

de Qualidade do Produto (QTPP – *Quality Target Product Profile*), que resume as características qualitativas desejáveis para garantir a qualidade esperada do produto. Essa informação servirá como guia dos atributos a serem testados na etapa de análise e as metas a serem buscadas durante o desenvolvimento do método analítico, considerando até mesmo a escolha da tecnologia analítica, tais como cromatografia, espectrometria, procedimentos manuais, entre outros.

Dentro disso, há a identificação dos Atributos Críticos de Qualidade (CQA – *Critical Quality Attributes*), que são os mais importantes e que em última instância, irão garantir a qualidade do produto. Assim, esses atributos devem ter um potencial impacto clínico do produto, como o doseamento, impurezas, dissolução, entre outros (MOURE, 2023).

No ATP podem ser definidos atributos de desempenho, como limite de detecção, limite de quantificação, seletividade, exatidão e precisão, considerando principalmente os CQAs definidos. A definição dos atributos de desempenho de acordo com a ICH pode ser vista no Apêndice 01.

A robustez é um atributo importante que deve ser buscado durante o desenvolvimento, mas é mais adequado que seja derivada dos outros atributos, e não necessariamente definida no ATP. O Quadro 1 descreve um ATP hipotético da determinação de doseamento de um IFA. O ATP é um documento em constante aprimoramento, que pode ser reavaliado conforme necessidade, por alterações nas características de desempenho ou um aumento do conhecimento sobre o produto e o procedimento analítico durante o desenvolvimento.

Quadro 1 – Exemplo de Perfil Analítico Alvo (ATP) para determinação do teor de um Insumo Farmacêutico Ativo (IFA) no medicamento.

Análise	Correlação com CQA	Justificativa
Doseamento do teor de IFA	O procedimento deve ser capaz de medir o teor do IFA de 90% a 110%	Eficácia do medicamento
Desempenho do método	Alvo	Justificativa
Exatidão Precisão	Incerteza padrão = 1,25%	Para garantir que a variação analítica em torno da média estimada esteja dentro da faixa de especificação mais ampla.
Seletividade	Nenhuma interferência de impurezas ou componentes da matriz no pico do IFA	Garantir que o IFA seja dosado de maneira inequívoca, com seletividade adequada (resolução mínima de 1,5) na presença de impurezas e componentes da matriz.
Faixa reportável	90,0% a 110,0% do declarado	Faixa estabelecida na monografia do produto, em compêndio oficial.

Fonte: Elaborado pelo autor. Adaptado de MOURE, 2023.

3.2. Seleção de tecnologia

Após a criação de um ATP, inicia-se uma fase prospectiva para reunir conhecimento prévio do produto e do procedimento. Coleta-se informações relevantes sobre as propriedades químicas do analito e da matriz, assim como de instrumentos e reagentes que possam ser utilizados durante o procedimento analítico. Esse conhecimento pode ser obtido de documentação interna da empresa, experiência ou referências científicas. Assim, é definida uma tecnologia capaz de atender os critérios desejados de qualidade descritos no perfil do produto.

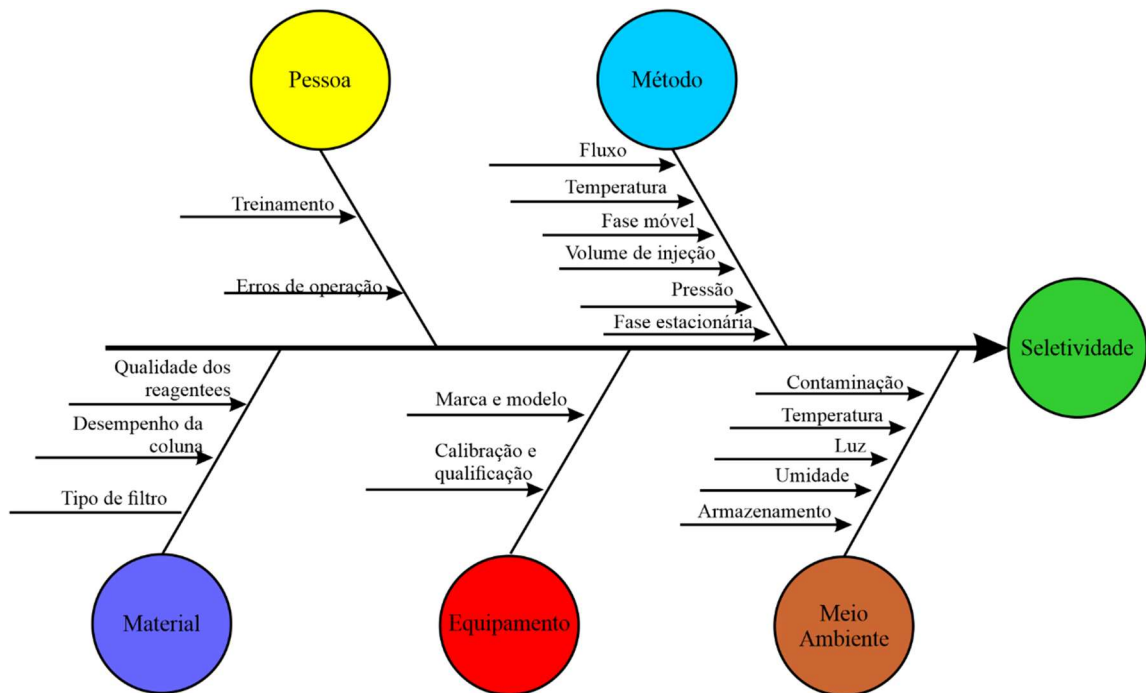
3.3. Análise de risco

Para definir o que será necessário estudar durante o desenvolvimento do método analítico, é necessário prosseguir com a fase prospectiva, agora utilizando de ferramentas mais robustas de definição de riscos. Nesta etapa, são definidos os atributos e os parâmetros do procedimento analítico.

Os atributos do procedimento analítico são os itens que estão mais próximos possíveis do desempenho do método, com uma correlação das respostas destes atributos à uma representação matemática da qualidade do método. Por exemplo, em um caso de cromatografia, podem ser definidos como atributos a resolução, fator de simetria do pico e número de pratos teóricos. Esses atributos normalmente estão definidos dentro de limites ou faixas de resultados para garantir a qualidade desejada da medição. Esses limites são definidos teoricamente por meio de literatura, ou podem até mesmo ser realizados experimentos durante esta fase de prospecção e seleção de tecnologia para ter uma melhor definição.

Já os parâmetros do procedimento analítico são definidos pela identificação dos possíveis itens que possam afetar um ou mais atributos do método analítico e são definidos geralmente com ferramentas de avaliação e gestão de risco, como o diagrama de Ishikawa, também conhecido como diagrama de espinha de peixe, Análise de Modos de Falhas e Efeitos (FMEA - *Failure Mode and Effect Analysis*), mapa de calor, uma combinação de métodos, entre outros. Essa etapa utiliza extensivamente o conhecimento prévio obtido embasado pela prospecção realizada até o momento. Na Figura 2 é mostrado um diagrama de Ishikawa hipotético para análise dos parâmetros relacionados à seletividade do método analítico, conseqüentemente relacionado ao atributo de resolução entre os picos do procedimento de cromatografia.

Figura 2 – Exemplo de diagrama de Ishikawa para avaliação de risco para definição de parâmetros do procedimento analítico com potencial impacto na seletividade.



Fonte: Elaborado pelo autor. Adaptado de MOURE, 2023.

3.4. Identificação dos valores e/ou intervalos dos parâmetros

Até este momento, foram definidos os atributos do produto e do procedimento analítico que serão a resposta-alvo do procedimento e a tecnologia usada para analisar tais atributos. Além disso, foram informados os parâmetros do procedimento analítico, que podem ser definidos como variáveis, pois serão os itens a serem alterados durante o desenvolvimento do método analítico para identificar a região ótima dos resultados esperados.

Com as informações obtidas, busca-se agora investigar as variáveis de modo a identificar as condições que permitam aumentar a confiança dos resultados obtidos entre cada análise durante a rotina, diminuindo a variância dos resultados. Para isso, é necessário modelar a forma de realizar os experimentos para criação e otimização do método analítico, garantindo que as variáveis serão amplamente testadas de forma mais eficaz possível, evitando, por exemplo, realizar um modelo OVAT. Uma das melhores ferramentas para isso é a utilização do DoE, que se baseia na conceitualização sistemática da combinação de variáveis de forma eficiente e combina experimentos multifatoriais com um modelo estatístico para mitigar a variabilidade e maximizar o resultado obtido.

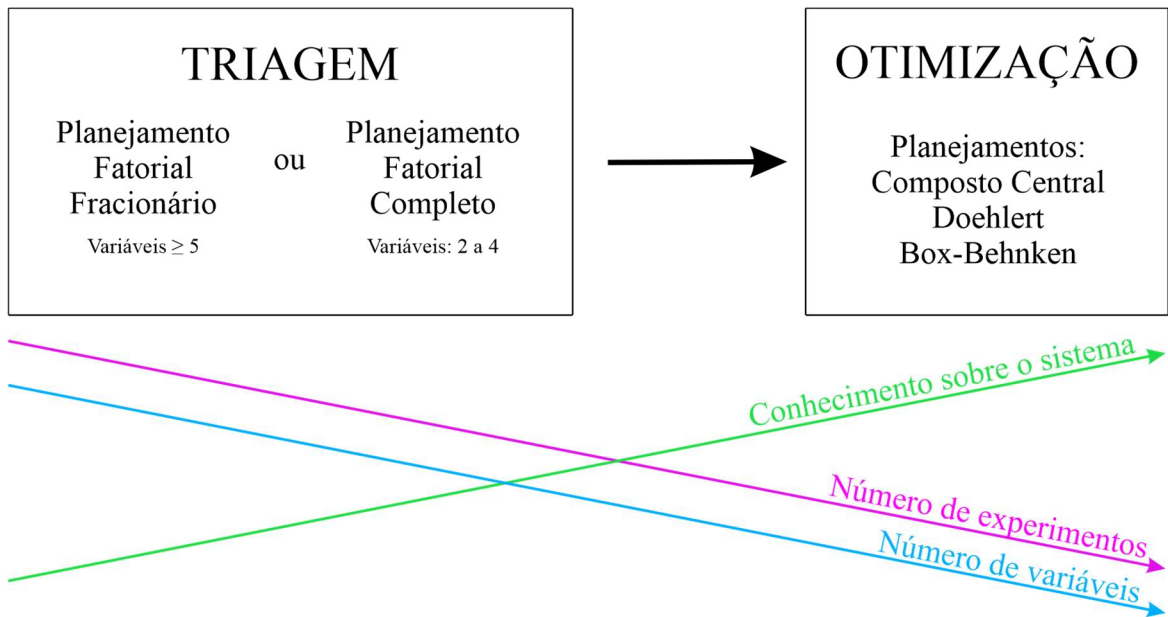
Seguindo esta ferramenta, primeiro são determinadas as variáveis (parâmetros) que possam impactar os resultados do método analítico (atributos), que foram definidas anteriormente na análise de risco. Considerando que o produto já estará finalizado, essas variáveis são relacionadas principalmente à análise e podem ser amplas como temperatura, pressão, tipo de solvente, pH, ou específicas da tecnologia definida para o projeto, como a vazão de solvente na cromatografia, ou o comprimento de onda monitorado na espectroscopia. Enquanto isso, a resposta avaliada pode ser definida de acordo com o método analítico utilizado, por meio dos sinais analíticos da análise, como absorvância, intensidade de emissão, sinal elétrico, ou por meio dos indicadores de desempenho como recuperação de analito, resolução entre picos cromatográficos, entre outros.

Após definidas as variáveis é necessário realizar um planejamento de experimentos para observar, com viés estatístico, se o experimento será adequado para obtenção de resultados confiáveis e reprodutíveis. Considerando que não se tenha conhecimento inicial suficiente para prosseguir diretamente para etapas de maior detalhamento, inicialmente é feita uma triagem, com um planejamento fatorial para que haja uma previsão de quais variáveis são realmente significativas para o método, sendo avaliada também a interação entre elas quando aplicável.

As variáveis e interações que não tenham impacto são desconsideradas, e fixadas em algum nível prático para uso no dia a dia. Já as variáveis significativas no experimento são analisadas com mais detalhamento na fase de otimização do método, por meio de modelos estatísticos com uma visão mais profunda do caso. Essa otimização servirá para definir o nível ótimo dessas variáveis, que permita extrair os resultados mais confiáveis durante o procedimento analítico, sempre seguindo como base as expectativas definidas no ATP.

A Figura 3 fornece uma visão ampla do racional por trás das etapas do DoE, onde o experimento inicia com uma grande quantidade de experimentos e variáveis, porém conforme os experimentos são realizados e o sistema se torna mais conhecido, os itens relevantes são afunilados e modelos mais refinados podem ser obtidos, com uma quantidade menor de experimentos e variáveis.

Figura 3 – Visão geral dos estágios do planejamento experimental para aplicação em DoE.

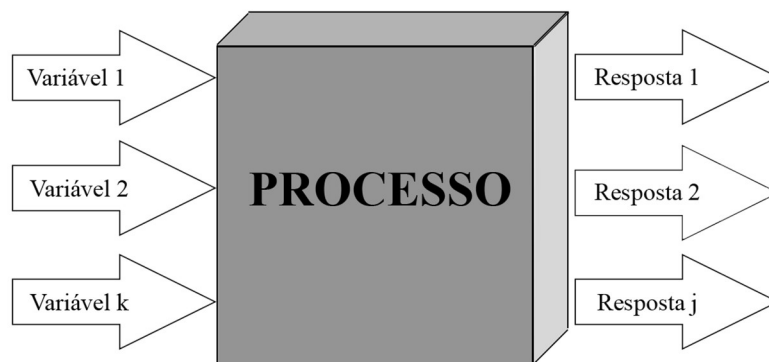


Fonte: Elaborado pelo autor. Adaptado de PARTE 12, 2024 2:00.

3.4.1. Triagem

Para determinar a influência de uma ou mais variáveis na resposta, assim como demonstrado na Figura 4, é necessário calcular qual a função que define o sistema do processo e opera com base nas variáveis como entrada e as respostas como saída.

Figura 4 – Um sistema que relaciona as variáveis de entrada com as saídas (respostas).



Fonte: Elaborado pelo autor. Adaptado de BRUNS, 2006.

As variáveis normalmente são os parâmetros do procedimento analítico obtidos anteriormente e podem ser definidos como variáveis independentes, e as respostas são os atributos, também definidas como variáveis dependentes. Após definidos, é realizado um planejamento fatorial de dois níveis para realizar uma triagem do impacto das variáveis, por

enquanto sem detalhar muito qual o nível desse impacto. Em casos com poucas variáveis ($n \leq 4$), é recomendado utilizar o planejamento fatorial completo, com um estudo de cada variável, enquanto em casos com mais variáveis ($n \geq 5$), o planejamento fatorial fracionário é o mais recomendado para economizar na quantidade de experimentos a serem realizados. Para somente uma variável, é possível realizar um sistema OVAT, porém não serão descritos neste trabalho.

Existem artigos demonstrando desenvolvimento de métodos com análises de 13 variáveis em 12 resultados (MOREIRA, 2020) para demonstrar que o DoE é uma ferramenta útil até mesmo para planejamentos complexos. Outros experimentos com exemplos práticos também são listados em artigos de revisão (ARAÚJO, 2021) (NOVAES, 2017).

3.4.1.1. Planejamento Fatorial completo

A primeira etapa do planejamento fatorial é especificar os níveis de cada variável, sendo numéricas para variáveis quantitativas ou categóricas para as qualitativas. Por exemplo, para uma análise em HPLC, pode ser definido uma vazão de solvente em diversos valores, enquanto o tipo de coluna utilizada em diversos tipos. A quantidade de níveis pode ser definida caso a caso, porém para este trabalho será descrito sobre o caso mais comum, com 2 níveis para cada variável (2^n , n = número de variáveis).

Em um exemplo, deve-se avaliar o impacto ao alterar a vazão de solvente de 0,6 para 0,8 mL/min, e da composição da fase móvel de 40%/60% etanol/água para 50%/50% etanol/água. Definido os níveis, os valores estipulados são codificados em nível alto (+1) e nível baixo (-1) para melhor visualização dos experimentos e são distribuídos de forma a obter pelo menos um resultado (R_1, R_2, \dots, R_j) com todas as possíveis combinações entre as duas variáveis, exibidas no Quadro 2. Considerando 2 níveis e 2 variáveis, este seria o planejamento fatorial em 2^2 , com 4 experimentos no mínimo.

A ordem de experimentos presente no quadro é puramente organizacional, pois devem ser feitos aleatoriamente para diminuir o erro devido influência externa, mas segue um padrão em que o primeiro experimento representa o nível baixo de todas as variáveis além do nível da primeira variável alterar a cada experimento, da segunda variável a cada 2 experimentos, da terceira variável a cada 4 experimentos, e assim em diante.

Quadro 2 – Exemplo de planejamento fatorial de experimentos em um caso 2².

Corrida	Variáveis				Resposta
	Vazão de solvente (1)		Composição da fase móvel (2)		
	mL/min	Código	Tipo	Código	
1	0,6	-1	40/60	-1	R ₁
2	0,8	+1	40/60	-1	R ₂
3	0,6	-1	50/50	+1	R ₃
4	0,8	+1	50/50	+1	R ₄

Fonte: Elaborado pelo autor.

O efeito principal de cada variável é definido pela média dos resultados ao alterar entre nível alto e baixo sem alterar outra variável, ou seja, para verificar o efeito da vazão de solvente será necessário avaliar a média entre os grupos de variação do nível da vazão enquanto mantém o nível alto da composição da fase móvel (experimentos 1 e 2) e o nível baixo da composição da fase móvel (experimentos 3 e 4). A equação (1.1.1) mostra como será calculado o efeito, considerando que \bar{y}_i é a média dos resultados obtidos na corrida i .

$$E_{vazão} = \frac{(\bar{y}_2 - \bar{y}_1) + (\bar{y}_4 - \bar{y}_3)}{2}$$

$$E_{vazão} = \frac{(R_2 - R_1) + (R_4 - R_3)}{2}$$

(1.1.1)

É possível reescrever a equação (1.1.1) de forma a agrupar os valores altos e baixos, sem levar em consideração o nível utilizado na outra variável, chegando à equação (1.1.2).

$$E_{vazão} = \frac{(\bar{y}_2 - \bar{y}_1) + (\bar{y}_4 - \bar{y}_3)}{2}$$

$$E_{vazão} = \frac{(\bar{y}_2 + \bar{y}_4) - (\bar{y}_1 + \bar{y}_3)}{2}$$

$$E_{fluxo} = \bar{y}_+ - \bar{y}_-$$

$$E_{vazão} = \left(\frac{R_2 + R_4}{2} \right) - \left(\frac{R_1 + R_3}{2} \right)$$

(1.1.2)

O mesmo deve ser feito para determinar o efeito da variação da composição da fase móvel, pela equação (1.1.3).

$$E_{\text{fase móvel}} = \bar{y}_+ - \bar{y}_-$$

$$E_{\text{fase móvel}} = \left(\frac{R_3 + R_4}{2} \right) - \left(\frac{R_1 + R_2}{2} \right)$$
(1.1.3)

Considerando este exemplo, supõe-se um caso em que a variação de uma das variáveis do nível baixo ao alto pode ser diferente quando feita com a outra variável em nível alto ou em nível baixo. Isso se deve à um outro efeito que pode ser relevante à resposta: a interação entre os dois efeitos.

Para calcular esse efeito de interação que considera as duas variáveis ao mesmo tempo, é necessário primeiro pensar no produto entre os níveis das duas. Uma multiplicação das duas variáveis nos níveis (+1,+1) e (-1,-1) nos fornece um nível positivo, enquanto dos níveis (-1,+1) e (+1,-1) nos fornece um nível negativo, visto na equação (1.1.4).

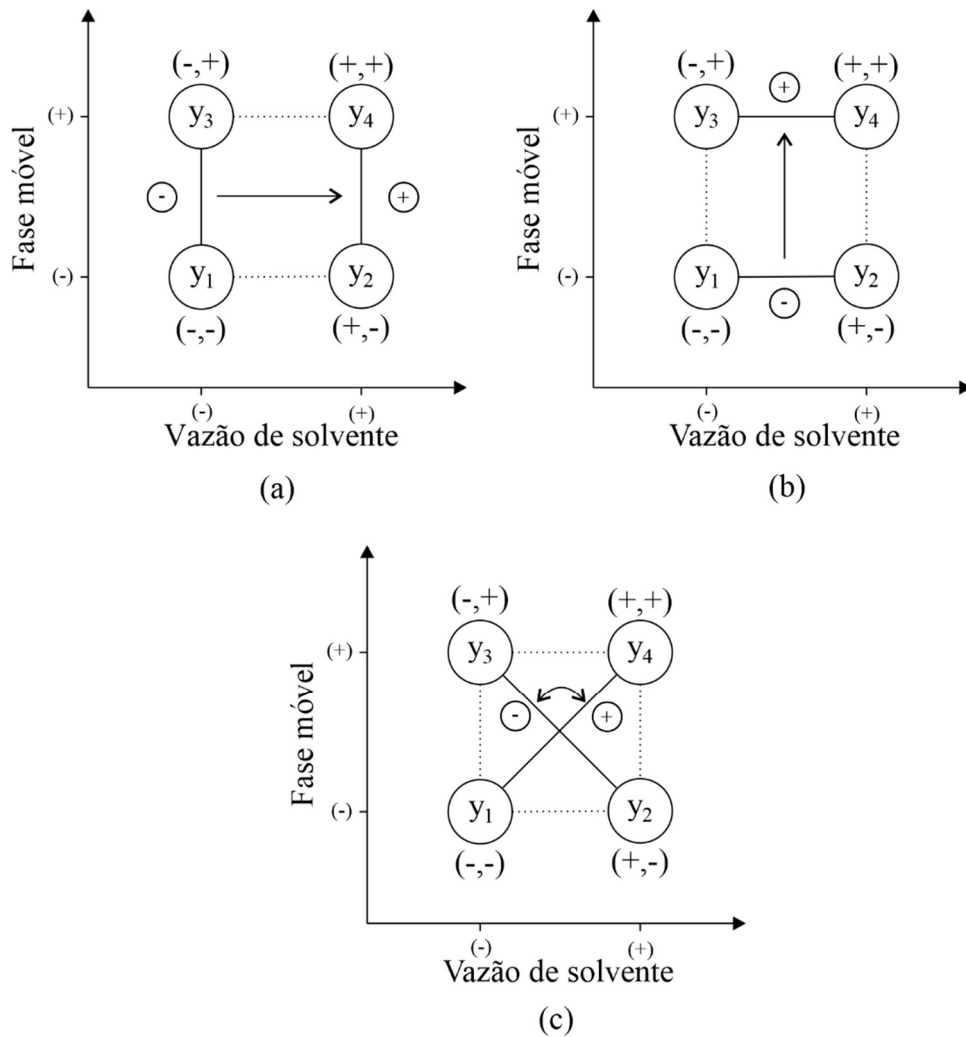
$$E_{\text{variação} \times \text{fase móvel}} = \bar{y}_+ - \bar{y}_-$$

$$E_{\text{variação} \times \text{fase móvel}} = \left(\frac{R_1 + R_4}{2} \right) - \left(\frac{R_2 + R_3}{2} \right)$$
(1.1.4)

Comumente o efeito principal das variáveis é mais relevante para o resultado do que a interação entre eles.

Essa interação pode fazer mais sentido quando visto geometricamente. Para fazermos isso, é necessário pensar em um sistema cartesiano em que cada eixo representa uma variável. Neste caso, um plano é o suficiente para cobrir as duas variáveis e pode ser visto na Figura 5.

Figura 5 – Interpretação geométrica em um planejamento fatorial 2^2 dos efeitos (a) principal de vazão de solvente, (b) principal de composição da fase móvel e (c) da interação entre as variáveis.



Fonte: Elaborado pelo autor. Adaptado de BRUNS, 2006.

A visualização do efeito das variáveis individualmente se baseia no deslocamento em seu próprio eixo, enquanto da interação entre as variáveis se baseia no deslocamento diagonal.

Mantendo o exemplo desenvolvido até o momento, pode-se estudar mais variáveis, aumentando a potência do planejamento fatorial completo, seja para 2^3 ou 2^4 .

Considerando que agora será estudada também a alteração de pressão no sistema, de 400 a 800 bar, será feita uma organização de experimentos descrita no Quadro 3. Agora serão necessários, no mínimo, 8 experimentos para avaliar todas as possíveis combinações.

Quadro 3 – Exemplo de planejamento fatorial de experimentos em um caso 2³.

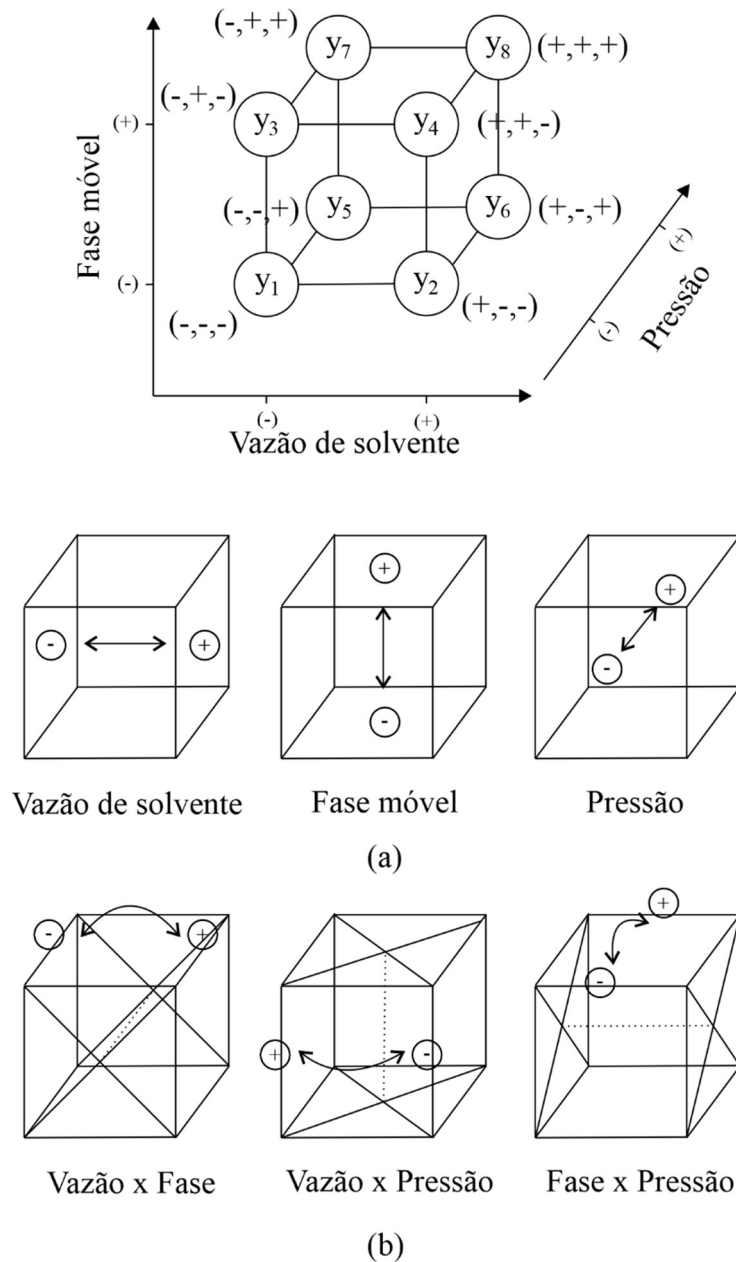
Corrida	Variáveis						Resposta
	Vazão de solvente (1)		Composição da fase móvel (2)		Pressão (3)		
	mL/min	Código	Tipo	Código	bar	Código	
1	0,6	-1	40/60	-1	400	-1	R ₁
2	0,8	+1	40/60	-1	400	-1	R ₂
3	0,6	-1	50/50	+1	400	-1	R ₃
4	0,8	+1	50/50	+1	400	-1	R ₄
5	0,6	-1	40/60	-1	800	+1	R ₅
6	0,8	+1	40/60	-1	800	+1	R ₆
7	0,6	-1	50/50	+1	800	+1	R ₇
8	0,8	+1	50/50	+1	800	+1	R ₈

Fonte: Elaborado pelo autor.

Para o caso de fatorial de duas variáveis, existia somente uma interação entre as duas variáveis, porém para um caso de fatorial de três variáveis, existe uma interação mais complexa entre eles. Para isso, é preciso analisar três tipos de efeito: o de primeira ordem é o efeito principal de cada variável (1, 2 e 3) que define a alteração na resposta com uma mudança do nível de uma variável individualmente; o de segunda ordem é a interação entre pares de variáveis (1*2, 1*3, 2*3) que define a diferença do efeito principal de uma variável enquanto outra variável esteja no nível baixo ou alto; e o de terceira ordem é a interação entre trios de variáveis (1*2*3), que define a diferença do efeito de interação de segunda ordem enquanto a variável restante esteja no nível baixo ou alto. Tanto para o efeito de segunda e terceira ordem, seu nível será o produto dos níveis das variáveis aplicáveis.

Essas interações também podem ser vistas geometricamente, conforme Figura 6, e, assim como as interações no fatorial de duas variáveis, os efeitos principais seguem uma alteração geométrica em um só eixo, enquanto as interações seguem alterações diagonais, com formação de planos de nível positivo e negativo.

Figura 6 – Interpretação geométrica em um planejamento fatorial 2^3 dos efeitos (a) principais e (b) das interações de segunda ordem entre as variáveis.



Fonte: Elaborado pelo autor. Adaptado de BRUNS, 2006.

Agora, para um caso de fatorial de quatro variáveis, segue-se a mesma lógica do fatorial de três variáveis, agora com, no mínimo, 16 experimentos e resultados com efeitos de quarto grau. Haverá quatro efeitos principais, seis de interações de segunda ordem, quatro de interações de terceira ordem e um de interações de quarta ordem.

A visualização geométrica em 4 dimensões já não pode ser representada, então comumente é utilizado uma representação de dois cubos 2^3 , um com a quarta variável em nível alto e outro em nível baixo.

3.4.1.2. Planejamento Fatorial fracionário

Na seção anterior, foi abordado o planejamento fatorial completo, em que os experimentos seguiam uma potência de 2^n (n = número de variáveis) e, até agora, não há nenhum problema evidente. Porém, como em qualquer potência, a quantidade de experimentos aumentará de forma exponencial. Ao analisar um sistema com poucas variáveis, de 3 ou 4 variáveis como tratado anteriormente, isso pode não ser um problema muito grave, pois 8 ou 16 experimentos podem não ser tão demandantes. Mas e ao analisar um sistema com mais variáveis?

Um sistema com somente uma variável a mais, de 5 variáveis, já chega à 32 experimentos enquanto um maior, de 7 variáveis, necessita de 128 experimentos. Além disso, quanto mais variáveis a se analisar, maior a quantidade de efeitos de interação. Ao analisar o Quadro 4, podemos ver a quantidade total de experimentos e a quantidade de efeitos de cada ordem.

Quadro 4 – Quantidade de experimentos e efeitos de acordo com a quantidade de variáveis (k) em um planejamento fatorial completo 2^{kc} .

Quantidade de variáveis (k)	Quantidade de experimentos (n)	Quantidade de efeitos							Total de efeitos
		Ordem							
		1 ^a	2 ^a	3 ^a	4 ^a	5 ^a	6 ^a	7 ^a	
3	8	3	3	1	-	-	-	-	7
4	16	4	6	4	1	-	-	-	15
5	32	5	10	10	5	1	-	-	31
6	64	6	15	20	15	6	1	-	63
7	128	7	21	35	35	21	7	1	127

Fonte: Elaborado pelo autor. Adaptado de BRUNS, 2006.

Como os resultados são descritos por uma expansão em série de funções de efeitos (primários e interações), é esperado que os últimos termos (interações de maior ordem) sejam mínimos quando comparados aos primários. Em um fatorial 2^7 , por exemplo, é possível indagar se é necessário obter o valor de efeito de sétima ordem, já que terá uma influência mínima no resultado e necessita do dobro da quantidade de experimentos de um fatorial 2^6 . Para contornar esse problema, não é possível excluir totalmente alguma variável para diminuir o tamanho do projeto, pois pode ter impacto relevante nos efeitos primários.

Considerando isso, foi elaborado um modelo de planejamento fracionário para que, ao incluir uma variável no planejamento, ainda sejam analisadas todas as variáveis sem aumentar a quantidade de experimentos a serem realizados. Esse modelo é definido com um planejamento fatorial fracionário 2^{n-k} (n = número de variáveis, k = resolução do experimento).

Para a elaboração de um planejamento fatorial fracionário tem-se um projeto muito próximo do fatorial completo, com a alteração crucial de que a última variável seja definida como a interação entre todas as outras variáveis. Para facilitar a visualização, será analisado um experimento de 4 variáveis e as etapas poderão ser aplicadas em qualquer quantidade de variáveis. No Quadro 5 estão demonstrados todos os efeitos presentes em um fatorial fracionário de 4 variáveis. Neste caso, os níveis da variável 4 foram definidos como o produto entre as variáveis 1, 2 e 3.

Quadro 5 – Distribuição de níveis nas corridas de um planejamento fatorial fracionário 2^{4-1} .

Efeitos Corridas	Efeitos															
	1	2	3	4	12	13	14	23	24	34	123	124	134	234	1234	
1	-1	-1	-1	-1	1	1	1	1	1	1	-1	-1	-1	-1	1	
2	1	-1	-1	1	-1	-1	1	1	-1	-1	1	-1	-1	1	1	
3	-1	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1	-1	1	
4	1	1	-1	-1	1	-1	-1	-1	-1	1	-1	-1	1	1	1	
5	-1	-1	1	1	1	-1	-1	-1	-1	1	1	1	-1	-1	1	
6	1	-1	1	-1	-1	1	-1	-1	1	-1	-1	1	-1	1	1	
7	-1	1	1	-1	-1	-1	1	1	-1	-1	-1	1	1	-1	1	
8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	

Fonte: Elaborado pelo autor.

Ao olhar mais atentamente, como visto na Figura 7, alguns efeitos possuem a mesma configuração de níveis em todas as corridas.

Figura 7 – Exemplos de efeitos confundidos em uma distribuição de níveis nas corridas de um planejamento fatorial fracionário 2^{4-1} .

Efeitos Corridas	Efeitos	
	4	123
1	-1	-1
2	1	1
3	1	1
4	-1	-1
5	1	1
6	-1	-1
7	-1	-1
8	1	1

Efeitos Corridas	Efeitos	
	1	234
1	-1	-1
2	1	1
3	-1	-1
4	1	1
5	-1	-1
6	1	1
7	-1	-1
8	1	1

Efeitos Corridas	Efeitos	
	12	34
1	1	1
2	-1	-1
3	-1	-1
4	1	1
5	1	1
6	-1	-1
7	-1	-1
8	1	1

Fonte: Elaborado pelo autor.

Isso pode ser explicado pela relação com o elemento identidade. Ao multiplicar as colunas de sinais iguais entre si, o resultado sempre será uma coluna somente de sinais positivos. Ou seja, ao multiplicar os níveis da coluna 1 com a própria coluna 1, resultará em uma coluna somente de valores +1. Ao multiplicar essa nova coluna por outra, não haverá alteração desta segunda coluna, o que pode ser definido com o próprio elemento de identidade. Portanto, temos uma relação tal qual visto nas equações (1.2.1) e (1.2.2):

$$I = 11 = 22 = 33 = 44 \quad (1.2.1)$$

Para este caso, inicialmente tem-se que

$$\begin{aligned} 4 &= 123 \\ 4.4 &= 123.4 \\ I &= 1234 \end{aligned} \quad (1.2.2)$$

Nos casos citados na Figura 7, utilizando a equação geradora (1.2.2), consegue-se provar com as equações (1.2.3) e (1.2.4) que as relações entre os efeitos são possíveis:

$$\begin{aligned} I.1 &= 1234.1 \\ I.1 &= (1.1)234 \\ I.1 &= I.234 \\ 1 &= 234 \end{aligned} \quad (1.2.3)$$

e

$$\begin{aligned} I.12 &= 1234.12 \\ I.12 &= (1.1)(2.2)34 \\ I.12 &= I.I.34 \\ 12 &= 34 \end{aligned} \quad (1.2.4)$$

Neste caso, dizemos que essas interações são padrões de confundimento dos efeitos. Isto é, os efeitos de primeira ordem estão confundidos com os de terceira ordem ($1 = 234$, $2 = 134$, $3 = 124$ e $4 = 123$) e os efeitos de segunda ordem estão confundidos entre si ($12 = 34$, $13 = 24$, etc.).

E afinal, o que isso significa para o planejamento? A consequência destes confundimentos é que um efeito qualquer analisado no fatorial fracionário não é quantitativamente igual ao mesmo efeito ao realizar um fatorial completo, porque é a soma dos dois efeitos confundidos. Isso causa um impacto ao valor real de cada efeito, mas isso não é tão relevante nesse caso pelos seguintes motivos: (i) esta é uma etapa de triagem, onde serão definidas as variáveis mais relevantes para o resultado analisado e, após isso, feito uma otimização do modelo e (ii) é esperado que os efeitos de maior ordem não sejam tão relevantes quanto os de menor ordem, portanto uma leve diferença no valor do efeito geralmente não causa um deslocamento tão grande durante a análise estatística da significância dos efeitos. Então, normalmente, se descobrirá quais são os efeitos significativos de forma muito próxima a de um fatorial completo, com menos experimentos.

3.4.2. Otimização

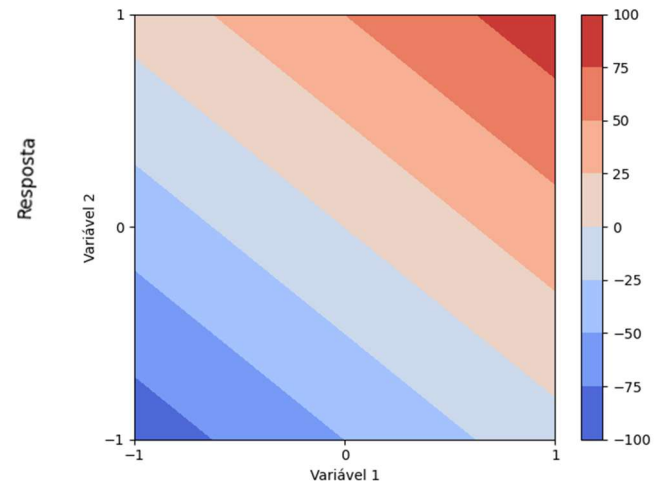
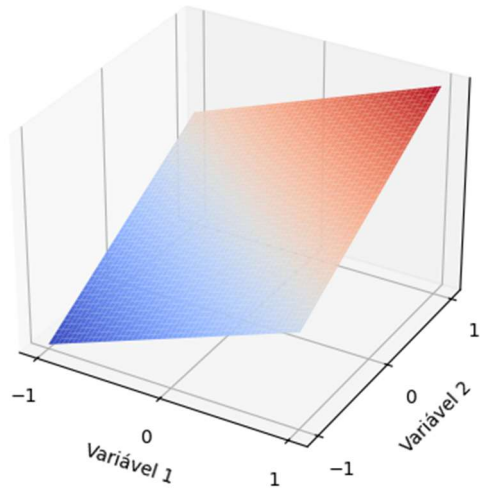
Após todo o processo de triagem do experimento, e com os efeitos significantes definidos, é necessário realizar um estudo mais aprofundado para identificar o ponto em que as variáveis irão maximizar a resposta, isto é, um ponto ótimo da curva de resposta.

3.4.2.1. Superfície de resposta

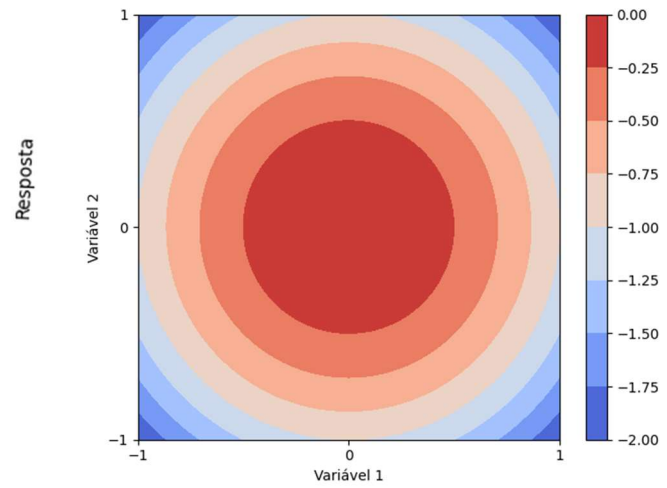
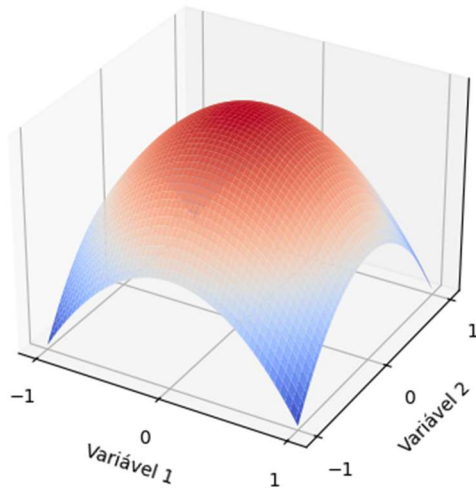
Dentre as técnicas mais utilizadas, há o método de superfícies de resposta. Esse método fornece uma visão geométrica com base no ajuste de equações e modelos matemáticos. Ao utilizar dos dados experimentais obtidos junto de análises de regressão, busca descrever o comportamento dos dados e realizar previsões estatisticamente válidas. Na Figura 8, pode-se ver exemplos de como normalmente essa superfície de resposta é apresentada, com visão geométrica, comumente com a inclusão de um mapa de calor.

Figura 8 – Exemplos mais comuns de superfícies de resposta para duas variáveis com mapa de calor, de (a) equação linear; (b) equação quadrática com ponto máximo; (c) com ponto mínimo e (d) com ponto de sela.

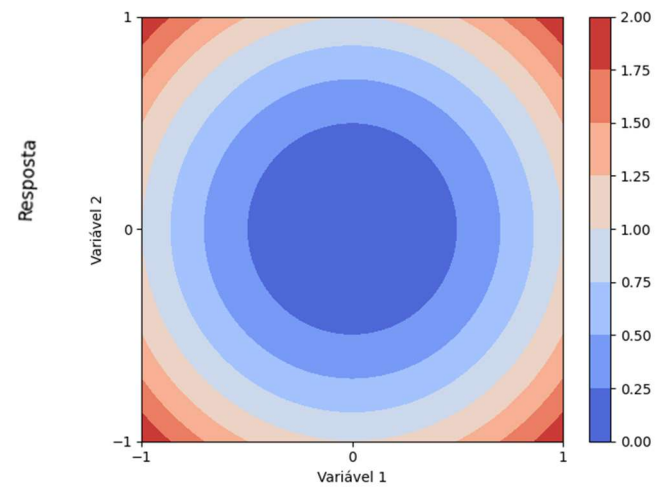
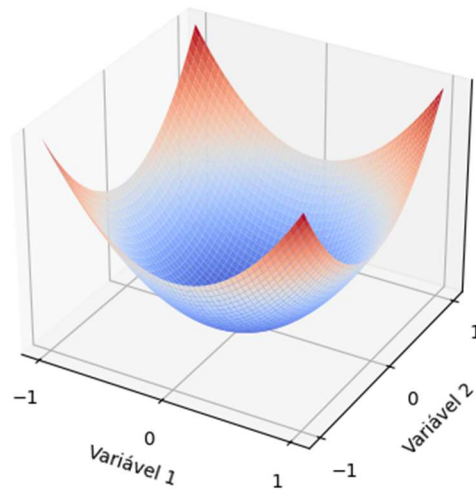
(a)



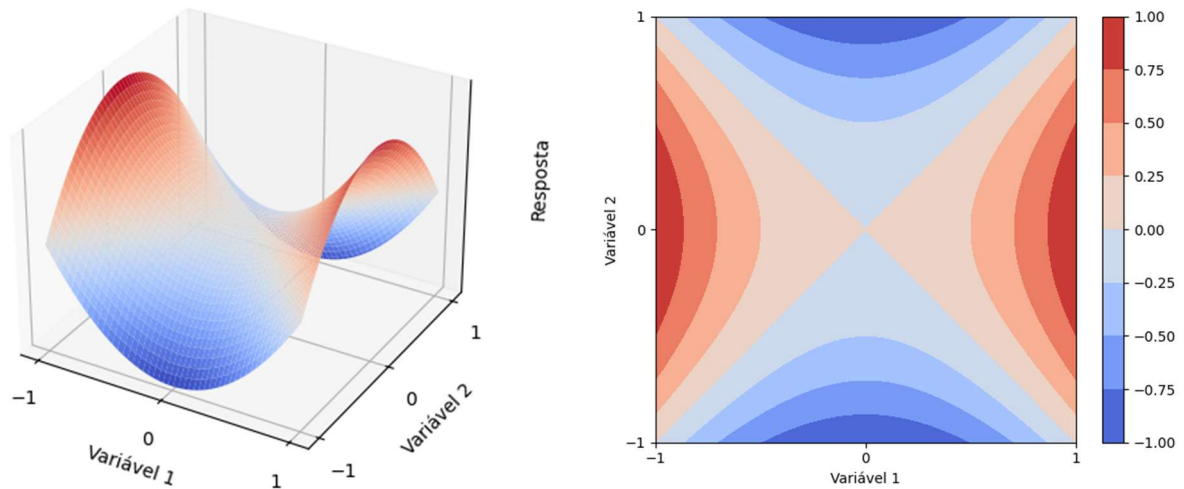
(b)



(c)



(d)



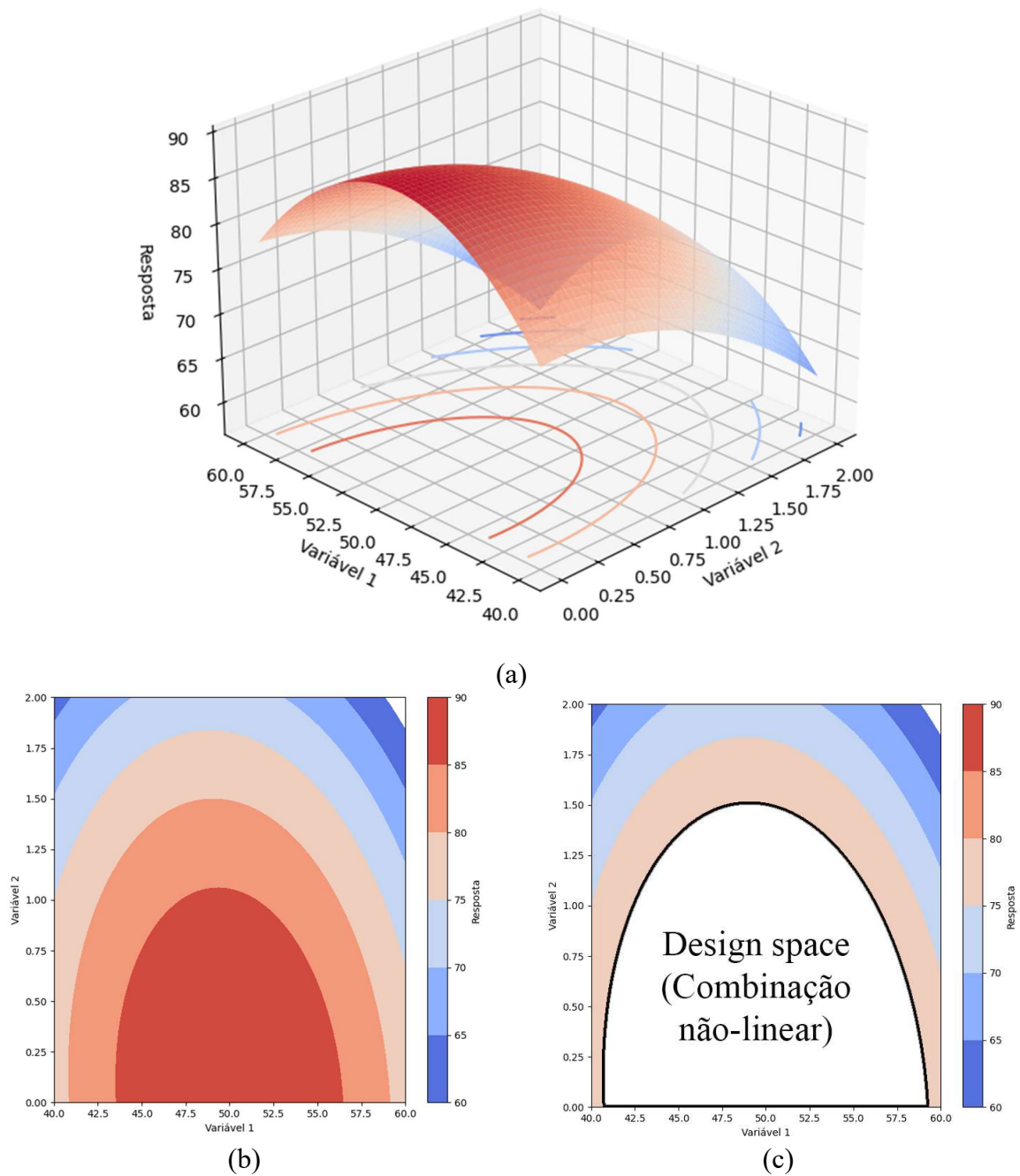
Fonte: Elaborado pelo autor.

Cada superfície de resposta representa somente uma resposta dependente da variação de duas variáveis. Em muitos casos, será necessário otimizar diversas respostas ao mesmo tempo. É possível fazer essa análise com uma sobreposição das superfícies ou das funções que a representam, para verificar a intersecção entre elas. A facilidade de realizar esse processo depende da similaridade entre os perfis de cada resposta.

Essa análise das superfícies de resposta é similar à construção do MODR. O MODR se refere à um intervalo dos parâmetros do procedimento onde os critérios de desempenho definidos no ATP são atendidos e a qualidade do resultado experimental é garantida. Normalmente, uma das simulações mais comuns para obtenção do MODR é a simulação estocástica de Monte-Carlo (REZVANI, 2014). A construção do MODR diminuirá a necessidade de uma revalidação no futuro pelo uso de parâmetros fora da região original esperada, pois pode fornecer uma região permitida comumente bem mais ampla do que o utilizado durante a rotina (ICH Q8, 2009) (MOURE, 2023).

Na Figura 9, vemos um exemplo de MODR criado a partir de uma combinação não-linear da equação da superfície de resposta que permite um resultado minimamente satisfatório. Toda essa região denominada como *design space* indica o intervalo dos parâmetros que fornece um valor acima do esperado (por exemplo, mais de 80% de dissolução).

Figura 9 – Exemplos de (a) superfície de resposta, (b) gráfico de contorno e (c) região de concepção operacional do método (MODR) para resultados $\geq 80\%$.



Fonte: Elaborado pelo autor. Adaptado de ICH Q8(R2), 2009.

Além dos modelos visuais da superfície de resposta, é também necessário analisar como essa região se comporta com ajuda da equação que forma esta superfície. A discussão sobre o modelo matemático utilizado para essa análise será melhor apresentada na seção Ajuste do modelo matemático (3.4.3). Nessa seção também será discutido o formato de álgebra matricial.

Para descrever corretamente a equação e a superfície, inicialmente é necessário obter corretamente o comportamento dos dados experimentais. Para isso, é realizado um planejamento experimental mais detalhado, que requer um estudo de, no mínimo, três níveis para cada variável. Em todos os modelos de planejamento que serão descritos, uma similaridade é o uso de um terceiro nível como a média dos níveis das variáveis e demonstrados como -1 (nível baixo), 0 (nível intermediário), e +1 (nível alto). Também são realizados experimentos em um ponto central, posicionado no nível intermediário de todas as variáveis simultaneamente e deve englobar, geralmente, cerca de 20% do total de experimentos, sendo normalmente o único ponto em que são feitas réplicas.

A necessidade de três níveis se deve à dimensionalidade do sistema. Em um planejamento, a quantidade de coeficientes não deve ser maior do que o número de combinações realizadas, pois não seria possível definir independentemente cada coeficiente (**BRUNS, 2006**). Para um modelo quadrático de duas variáveis, são definidos seis coeficientes (um para constante - β_0 ; dois de primeira ordem - β_i ; dois de segunda ordem - β_{ii} ; e um de interação entre as duas variáveis - β_{ij}) enquanto serão analisadas no mínimo sete combinações, a depender do modelo utilizado considerando três níveis para duas variáveis.

Os modelos mais utilizados para o ajuste de funções quadráticas são: Planejamento Composto Central (CCD – *Central Composite Design*), para casos em que todas as variáveis sejam igualmente importantes; Planejamento Doehlert (DD – *Doehlert Design*), para casos em que algumas variáveis sejam mais relevantes do que outras; e Planejamento Box-Behnken (BBD – *Box-Behnken Design*), para casos em que não seja possível testar as condições extremas (limites) das variáveis. Estes modelos de planejamento foram criados como uma solução para reduzir a quantidade de experimentos realizados em planejamentos fatoriais de três níveis, que demandavam muita experimentação mesmo para poucas variáveis. No Quadro 6, os três planejamentos são comparados entre si em relação à quantidade de experimentos necessários para definir a resposta para diferentes quantidades de variáveis. Uma descrição mais detalhada dos planejamentos de otimização se encontra no Apêndice 02.

Quadro 6 – Quantidade de parâmetros e experimentos entre os Planejamentos Composto Central, Doehlert e Box-Behnken de acordo com a quantidade de variáveis.

Variáveis	Número de parâmetros (p)	Número de experimentos (f)		
		Composto Central (CCD)	Doehlert (DD)	Box-Behnken (BBD)
2	6	9	7	-
3	10	15	13	13
4	15	25	21	25
5	21	43	31	41
6	28	77	43	61
7	36	143	57	85
8	45	273	73	113

Fonte: Elaborado pelo autor. Adaptado de PARTE 11, 2024 09:30.

3.4.3. Ajuste do modelo matemático

Todos os planejamentos percorridos até o momento tinham um objetivo em comum: obter dados experimentais suficientes para elaborar uma equação que definisse o comportamento da resposta de acordo com a dependência das variáveis e que será utilizada para formar uma superfície de resposta.

3.4.3.1. Regressão Linear Múltipla

Para definir matematicamente o formato da superfície deve-se analisar por meio de uma função, idealmente por meio de funções lineares. Para duas ou mais variáveis, é utilizado um modelo de regressão linear múltipla, descrita pela equação geral (3.1.1). Nessa equação, a resposta y é relacionada às variáveis x_i , também chamados de regressores, k se refere ao número de variáveis e os termos β se referem aos coeficientes da curvatura, com β_0 indicando o termo constante, β_i o coeficiente linear, também chamados de coeficientes parciais de regressão, e ε o resíduo (diferença entre o experimental e o previsto) (MONTGOMERY, 2021). Ao aplicar a equação (3.1.1) em um planejamento fatorial de duas variáveis, é possível obter a equação (3.1.2), que descreve um plano com gráfico de contorno composto de linhas retas, como visto na Figura 10.

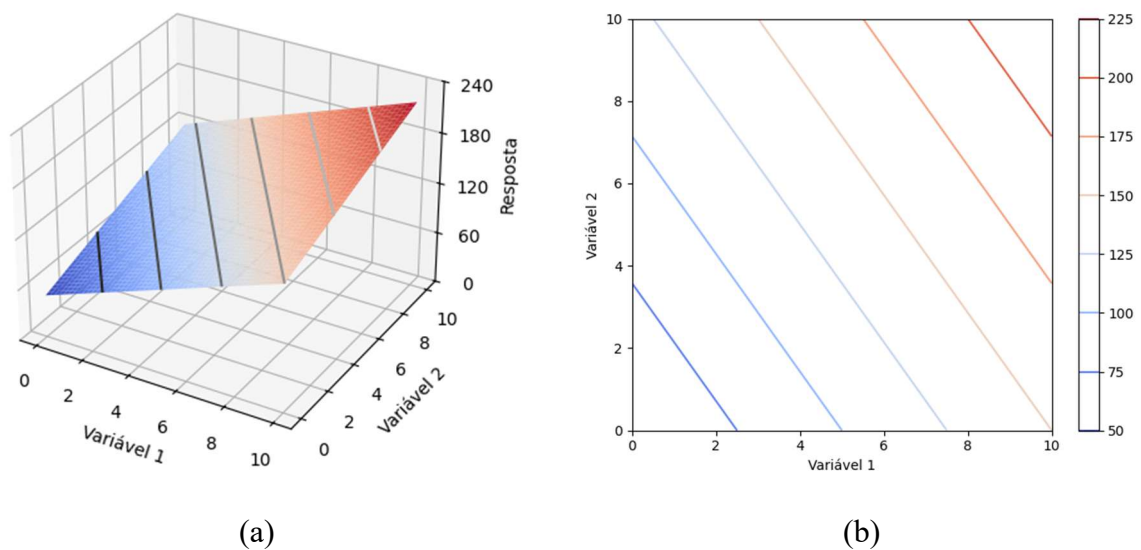
$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \varepsilon$$

(3.1.1)

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

(3.1.2)

Figura 10 – (a) Superfície de regressão e (b) gráfico de contorno para um modelo de duas variáveis sem termos de interação e quadráticos.



Fonte: Elaborado pelo autor. Adaptado de MONTGOMERY, 2021.

É possível, também, analisar a interação entre as variáveis pela adição de um coeficiente β_{ij} , representado nas equações geral (3.1.3) e para duas variáveis (3.1.4). Esse modelo irá criar um plano curvado com gráfico de contorno compostos de linhas curvas, como visto na Figura 11.

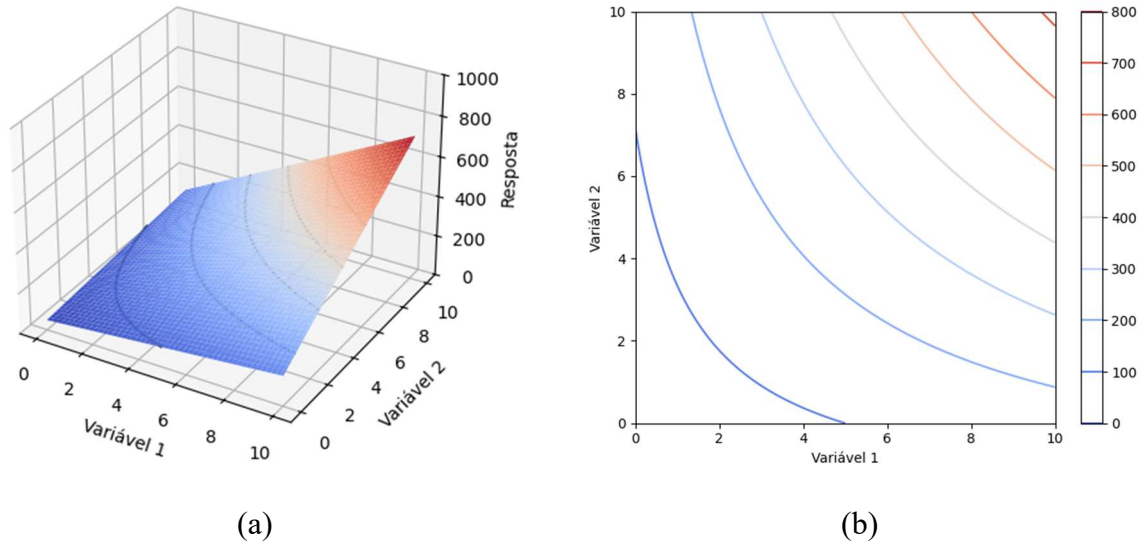
$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{1 \leq i < j}^k \beta_{ij} x_i x_j + \varepsilon$$

(3.1.3)

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon$$

(3.1.4)

Figura 11 – (a) Superfície de regressão e (b) gráfico de contorno para um modelo de duas variáveis com termos de interação entre as variáveis.



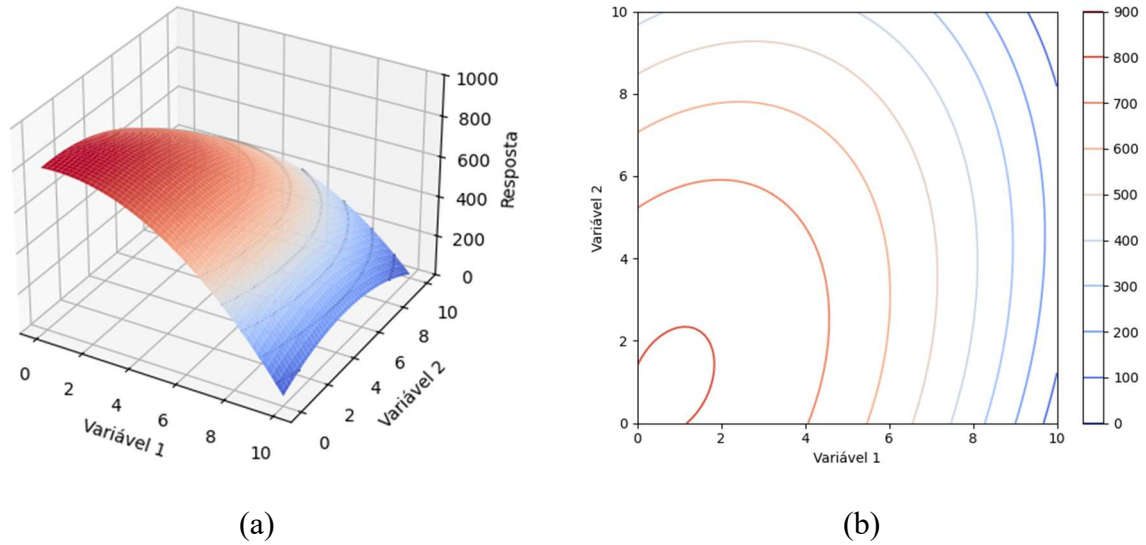
Fonte: Elaborado pelo autor. Adaptado de MONTGOMERY, 2021.

Para um melhor entendimento e detalhamento do comportamento dos dados na presença de curvatura, é possível utilizar um modelo quadrático. Os modelos empíricos normalmente são polinômios de segunda ordem (NOVAES, 2017). Esse modelo segue o modelo de regressão demonstrado nas equações (3.1.1) e (3.1.3), com adição dos termos quadráticos β_{ii} . Ao aplicar a equação geral (3.1.5) em um planejamento duas variáveis, é possível obter a equação (3.1.6). Esse modelo irá criar uma parábola com gráfico de contorno compostos de linhas curvas, como visto na Figura 12.

$$\hat{y} = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i=1}^k \beta_{ii} x_i^2 + \sum_{1 \leq i < j \leq k} \beta_{ij} x_i x_j + \varepsilon \quad (3.1.5)$$

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + \varepsilon \quad (3.1.6)$$

Figura 12 – (a) Superfície de regressão e (b) gráfico de contorno para um modelo de duas variáveis com termos quadráticos e de interação entre as variáveis.



Fonte: Elaborado pelo autor. Adaptado de MONTGOMERY, 2021.

A condição da superfície criada não ser linear não interfere na linearidade do método de regressão. Isso se deve pois o modelo deve ser linear nos coeficientes (β), independente do formato da superfície gerada. Uma equação com variáveis quadráticas ou de interação podem ser rearranjadas para manter um termo linear, assim como feito nas equações (3.1.7) e (3.1.8), desde que os coeficientes se mantenham lineares (MONTGOMERY, 2021).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon$$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

(3.1.7)

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + \varepsilon$$

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \varepsilon$$

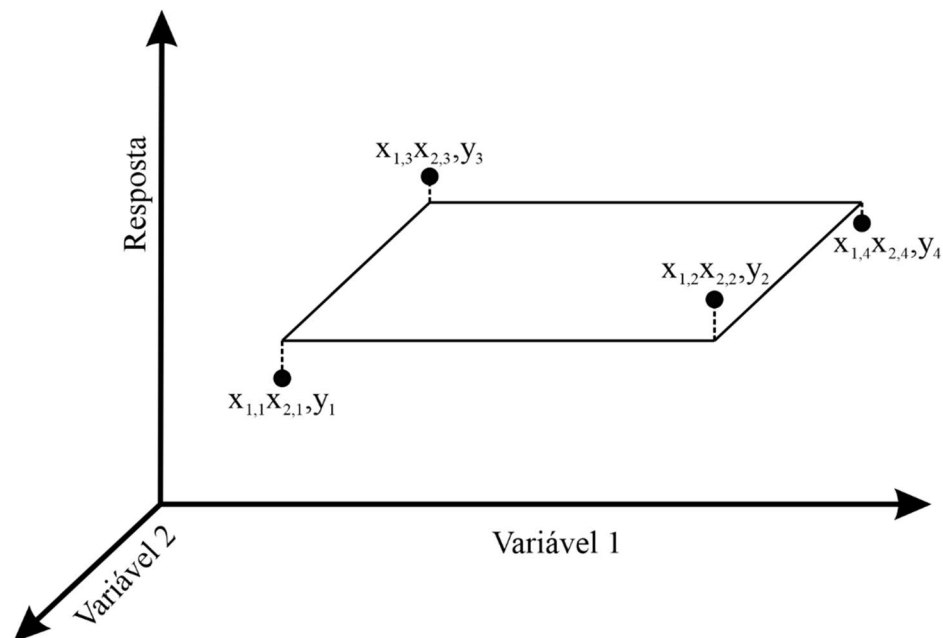
(3.1.8)

Agora, é necessário estudar modelos matemáticos que possam fornecer os valores dos coeficientes que se aproximem do valor ótimo de resposta por meio do uso dos dados obtidos pelo planejamento de otimização.

3.4.3.2. Método dos Mínimos Quadrados

Uma das formas de estimar os coeficientes da equação que descreve a superfície de resposta é por meio do Método dos Mínimos Quadrados (MMQ). O MMQ é um modelo de ajuste que busca minimizar a distância entre os dados experimentais obtidos durante o planejamento de triagem/otimização (y_n) e os dados previstos pela equação formada a partir da regressão (\hat{y}_n), aproximando ao máximo a reta/superfície de resposta que define o comportamento do sistema dos seus pontos experimentais. Na Figura 13 vemos um exemplo geral de reta formada na minimização da distância entre os seis pontos experimentais com uma equação que define o sistema.

Figura 13 – Exemplo gráfico da aproximação entre os pontos experimentais e a equação de reta que define o sistema com mínima distância entre todos os pontos.



Fonte: Elaborado pelo autor. Adaptado de MILLER, 2010.

A base para este método se dá pela consideração de que os erros experimentais possam ser tanto negativos quanto positivos se comparados à regressão. Por isso, o ideal é minimizar a soma dos quadrados dos resíduos, pois é um valor mais representativo, por analisar o grau do erro e não seu sentido (MILLER, 2010).

Ao adaptar a equação (3.1.1) para vários pontos experimentais, como visto na equação (3.2.1), o MMQ busca diminuir o valor do resíduo ϵ . Considerando que haverá diversos pontos experimentais (y_n), cada ponto haverá um resíduo (ϵ_n) para ser otimizado. Neste método, o termo a ser reduzido é a soma dos quadrados do resíduo (L), como visto na equação (3.2.2)

$$y_n = \beta_0 + \sum_{i=1}^k \beta_i x_{in} + \varepsilon_n$$

(3.2.1)

$$L = \sum_{n=1}^k \varepsilon_n^2 = \sum_{n=1}^k \left(y_n - \beta_0 - \sum_{i=1}^k \beta_i x_{in} \right)^2$$

(3.2.2)

Ao derivar L a fim de minimizar o valor em relação aos coeficientes β , as estimativas dos coeficientes ($\hat{\beta}$) devem seguir o conjunto de equações (3.2.3), em uma quantidade de experimentos m e quantidade de variáveis k .

$$\begin{aligned} m\hat{\beta}_0 &+ \hat{\beta}_1 \sum_{n=1}^m x_{1n} &+ \hat{\beta}_2 \sum_{n=1}^m x_{2n} &+ \dots &+ \hat{\beta}_k \sum_{n=1}^m x_{kn} &= \sum_{n=1}^m y_n \\ \hat{\beta}_0 \sum_{n=1}^m x_{1n} &+ \hat{\beta}_1 \sum_{n=1}^m x_{1n}^2 &+ \hat{\beta}_2 \sum_{n=1}^m x_{1n}x_{2n} &+ \dots &+ \hat{\beta}_k \sum_{n=1}^m x_{1n}x_{kn} &= \sum_{n=1}^m x_{1n}y_n \\ \dots &\dots &\dots &\dots &\dots &\dots \\ \hat{\beta}_0 \sum_{n=1}^m x_{kn} &+ \hat{\beta}_1 \sum_{n=1}^m x_{kn}x_{1n} &+ \hat{\beta}_2 \sum_{n=1}^m x_{kn}x_{2n} &+ \dots &+ \hat{\beta}_k \sum_{n=1}^m x_{kn}^2 &= \sum_{n=1}^m x_{kn}y_n \end{aligned}$$

(3.2.3)

Esse conjunto de equações pode ser resolvido por qualquer método apropriado para resolução de um sistema de equações lineares. Neste trabalho, será discorrido a resolução utilizando de matrizes.

Para um modelo de regressão linear múltipla, é possível usar a equação (3.2.1) como base para definir um sistema de matrizes, descrita pela equação (3.2.4) e utilizando a Figura 14. Nas matrizes estão demonstradas as combinações de níveis entre as variáveis, coeficientes e resíduo com a resposta obtida na matriz Y.

$$Y = XB + E$$

(3.2.4)

Figura 14 – Exemplo genérico de matrizes de respostas, variáveis, coeficientes e resíduos para i variáveis e n experimentos.

$$\begin{array}{c}
 \left[\begin{array}{c} y_1 \\ y_2 \\ \vdots \\ y_n \end{array} \right] \\
 Y
 \end{array}
 \quad
 \begin{array}{c}
 \left[\begin{array}{cccccc} 1 & x_{1,1} & x_{2,1} & \dots & x_{i,1} \\ 1 & x_{1,2} & x_{2,2} & \dots & x_{i,2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1,n} & x_{2,n} & \dots & x_{i,n} \end{array} \right] \\
 X
 \end{array}
 \quad
 \begin{array}{c}
 \left[\begin{array}{c} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_i \end{array} \right] \\
 B
 \end{array}
 \quad
 \begin{array}{c}
 \left[\begin{array}{c} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{array} \right] \\
 E
 \end{array}$$

Fonte: Elaborado pelo autor. Adaptado de MONTGOMERY, 2021.

Para esta resolução, é preparada uma matriz X com cada variável ocupando uma coluna. Isso se deve para a multiplicação com a matriz B resultar em cada variável x_i multiplicando com seu respectivo coeficiente β_i . É necessário, então, uma matriz X contendo uma coluna completa de valores 1 (constante β_0), colunas para os valores de cada variável (β_i, β_j), para seus termos quadráticos, quando aplicáveis, (β_{ii}, β_{jj}) e para as interações (β_{ij}). O vetor Y contém os valores de resposta experimental para cada combinação de valores na matriz X .

Essas matrizes são utilizadas na equação (3.2.5) para resolver um vetor de coeficientes \hat{B} , onde cada termo b_i representa uma estimativa do coeficiente β_i da matriz B .

$$\hat{B} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_i \end{bmatrix} = (X^t X)^{-1} X^t Y \tag{3.2.5}$$

Para um melhor entendimento, a Figura 15 exemplifica as etapas de cálculos necessárias para obter o vetor \hat{B} em uma regressão linear sem termos quadráticos e de interação.

Figura 15 – Demonstração dos cálculos para obter as matrizes (a) X^tX , (b) $(X^tX)^{-1}$, (c) X^tY e (d) $\hat{B} = (X^tX)^{-1} X^tY$ de uma regressão linear múltipla de 2 variáveis sem termos de interação ou quadráticos.

$$X = \begin{bmatrix} 1 & x_{1,1} & x_{2,1} \\ 1 & x_{1,2} & x_{2,2} \\ \vdots & \vdots & \vdots \\ 1 & x_{1,n} & x_{2,n} \end{bmatrix} \quad X^t = \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_{1,1} & x_{1,2} & \dots & x_{1,n} \\ x_{2,1} & x_{2,2} & \dots & x_{2,n} \end{bmatrix}$$

$$X^tX = \begin{bmatrix} 1 * 1 + 1 * 1 + \dots + 1 * 1 & 1 * x_{1,1} + 1 * x_{1,2} + \dots + 1 * x_{1,n} & 1 * x_{2,1} + 1 * x_{2,2} + \dots + 1 * x_{2,n} \\ 1 * x_{1,1} + 1 * x_{1,2} + \dots + 1 * x_{1,n} & x_{1,1} * x_{1,1} + x_{1,2} * x_{1,2} + \dots + x_{1,n} * x_{1,n} & x_{1,1} * x_{2,1} + x_{1,2} * x_{2,2} + \dots + x_{1,n} * x_{2,n} \\ 1 * x_{2,1} + 1 * x_{2,2} + \dots + 1 * x_{2,n} & x_{1,1} * x_{2,1} + x_{1,2} * x_{2,2} + \dots + x_{1,n} * x_{2,n} & x_{2,1} * x_{2,1} + x_{2,2} * x_{2,2} + \dots + x_{2,n} * x_{2,n} \end{bmatrix}$$

$$X^tX = \begin{bmatrix} n & \sum x_1 & \sum x_2 \\ \sum x_1 & \sum x_1^2 & \sum x_1x_2 \\ \sum x_2 & \sum x_1x_2 & \sum x_2^2 \end{bmatrix}$$

(a)

$$X^tX * (X^tX)^{-1} = I$$

$$\begin{bmatrix} n & \sum x_1 & \sum x_2 \\ \sum x_1 & \sum x_1^2 & \sum x_1x_2 \\ \sum x_2 & \sum x_1x_2 & \sum x_2^2 \end{bmatrix} * \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

(b)

$$X^t = \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_{1,1} & x_{1,2} & \dots & x_{1,n} \\ x_{2,1} & x_{2,2} & \dots & x_{2,n} \end{bmatrix} \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$X^t Y = \begin{bmatrix} 1 * y_1 + 1 * y_2 + \dots + 1 * y_n \\ x_{1,1} * y_1 + x_{1,2} * y_2 + \dots + x_{1,n} * y_n \\ x_{2,1} * y_1 + x_{2,2} * y_2 + \dots + x_{2,n} * y_n \end{bmatrix}$$

$$X^t Y = \begin{bmatrix} \sum y \\ \sum x_1 y \\ \sum x_2 y \end{bmatrix}$$

(c)

$$(X^t X)^{-1} = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{bmatrix} \quad X^t Y = \begin{bmatrix} \sum y \\ \sum x_1 y \\ \sum x_2 y \end{bmatrix}$$

$$\hat{B} = (X^t X)^{-1} X^t Y = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} a_{1,1} * \sum y + a_{1,2} * \sum x_1 y + a_{1,3} * \sum x_2 y \\ a_{2,1} * \sum y + a_{2,2} * \sum x_1 y + a_{2,3} * \sum x_2 y \\ a_{3,1} * \sum y + a_{3,2} * \sum x_1 y + a_{3,3} * \sum x_2 y \end{bmatrix}$$

(d)

Fonte: Elaborado pelo autor. Adaptado de MONTGOMERY, 2021.

Na Figura 16 temos os exemplos de matrizes que devem ser preparadas para modelos de uma ou duas variáveis. Planejamentos com mais variáveis seguem o mesmo princípio.

Figura 16 – Exemplo de estrutura das matrizes Y, X, B e E para um planejamento de (a) uma variável, (b) duas variáveis sem interação e termo quadrático e (c) duas variáveis.

$$\begin{array}{c}
 \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \\
 \text{Y}
 \end{array}
 \begin{array}{c}
 \begin{bmatrix} 1 & x_{1,1} \\ 1 & x_{1,2} \\ \vdots & \vdots \\ 1 & x_{1,n} \end{bmatrix} \\
 \text{X}
 \end{array}
 \begin{array}{c}
 \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \\
 \text{B}
 \end{array}
 \begin{array}{c}
 \begin{bmatrix} \varepsilon_0 \\ \varepsilon_1 \end{bmatrix} \\
 \text{E}
 \end{array}$$

(a)

$$\begin{array}{c}
 \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \\
 \text{Y}
 \end{array}
 \begin{array}{c}
 \begin{bmatrix} 1 & x_{1,1} & x_{2,1} \\ 1 & x_{1,2} & x_{2,2} \\ \vdots & \vdots & \vdots \\ 1 & x_{1,n} & x_{2,n} \end{bmatrix} \\
 \text{X}
 \end{array}
 \begin{array}{c}
 \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} \\
 \text{B}
 \end{array}
 \begin{array}{c}
 \begin{bmatrix} \varepsilon_0 \\ \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} \\
 \text{E}
 \end{array}$$

(b)

$$\begin{array}{c}
 \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \\
 \text{Y}
 \end{array}
 \begin{array}{c}
 \begin{bmatrix} 1 & x_{1,1} & x_{2,1} & x_{11,1} & x_{22,1} & x_{12,1} \\ 1 & x_{1,2} & x_{2,2} & x_{11,2} & x_{22,2} & x_{12,2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1,n} & x_{2,n} & x_{11,n} & x_{22,n} & x_{12,n} \end{bmatrix} \\
 \text{X}
 \end{array}
 \begin{array}{c}
 \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_{11} \\ \beta_{22} \\ \beta_{12} \end{bmatrix} \\
 \text{B}
 \end{array}
 \begin{array}{c}
 \begin{bmatrix} \varepsilon_0 \\ \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_{11} \\ \varepsilon_{22} \\ \varepsilon_{12} \end{bmatrix} \\
 \text{E}
 \end{array}$$

(c)

Fonte: Elaborado pelo autor.

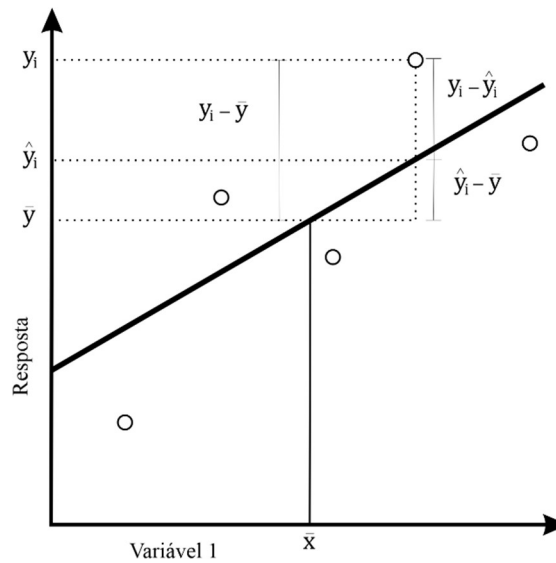
Após demonstrado o princípio do MMQ, deve-se lembrar que este método possui duas aplicações, de acordo com a variância dos dados em todos os pontos experimentais, chamados de MMQ Ordinários (MMQO) para casos em que os dados apresentem variâncias uniformes durante toda a curva da variável e a resolução segue conforme explicado até o momento e MMQ Ponderados (MMQP), para casos em que os dados apresentem variâncias desiguais conforme altera-se a variável. Uma melhor explicação de como avaliar a uniformidade das variâncias pelo Teste de Cochran e quais as alterações feitas em casos de MMQP é feita no Apêndice 03.

3.4.3.3. Análise de variância (ANOVA)

Uma dúvida comum após essa quantidade de cálculos é como garantir que a equação que consegue definir todo o sistema de diversas variáveis está realmente correta. Para isso, é necessário avaliar estatisticamente os valores obtidos experimentalmente (real) contra os obtidos pelo uso da equação (previsto), por meio de técnicas bem conhecidas no meio estatístico, como a Análise de Variância (ANOVA – *ANalysis Of VAriance*). A ANOVA se baseia na decomposição da variância total do sistema em diversos componentes menores que descrevem os efeitos causados pela mudança de nível das variáveis e pela aleatoriedade intrínseca aos erros experimentais.

Os componentes podem ser separados inicialmente pela Variância Total, que define a variância individual dos resultados experimentais contra a média. Essa componente pode ser dividida em duas, a Variância de Regressão e a Variância do Resíduo. A Variância de Regressão define a variância individual dos resultados previstos contra a média, enquanto a Variância do Resíduo definiria tudo que não é explicado pelo modelo. Na Figura 17, é possível demonstrar graficamente essa decomposição dos dados, onde a Variância Total ($y_i - \bar{y}$) consegue ser explicada pela Variância de Regressão ($\hat{y}_i - \bar{y}$) e a Variância de Resíduo ($y_i - \hat{y}_i$).

Figura 17 – Decomposição da variância das respostas experimentais contra a média ($y_i - \bar{y}$) em termos da variância das respostas previstas contra a média ($\hat{y}_i - \bar{y}$) e das respostas experimentais contra as respostas previstas ($y_i - \hat{y}_i$) em uma equação linear.



Fonte: Elaborado pelo autor. Adaptado de BRUNS, 2006.

A Variância do Resíduo também pode ser dividida em mais dois componentes: a Variância de Falta de Ajuste e a Variância de Erro Puro. A Variância de Falta de Ajuste demonstra os resultados experimentais que não são explicados pela regressão e a Variância de Erro Puro define as variações aleatórias no experimento, inerentes à repetição do mesmo experimento, como o efeito da mudança de temperatura, dia, horário, analista, entre outros. A decomposição da Variância do Resíduo só é possível em um experimento com replicatas dos pontos experimentais, pois é o que permite avaliar se a variância em cada ponto experimental se dá pela presença dos fatores externos ou por uma baixa qualidade da regressão que representa o sistema (BRUNS, 2006) (NOVAES, 2017).

Os cálculos se baseiam, principalmente, no uso da Soma Quadrática (SQ) e da Média Quadrática (MQ). A SQ, neste caso, é a potência quadrática da diferença entre dois valores de resposta. As equações (3.3.1) e (3.3.2) demonstram as decomposições das SQs da variância a partir da Variância Total, onde y_i se refere à uma resposta obtida experimentalmente, \hat{y}_i uma resposta obtida por meio da equação de regressão, e \bar{y} se refere à média dos resultados experimentais. As equações apresentam um componente com índice j , para representar o número da replicata do mesmo nível i . As SQs de cada variância são representadas pelos termos SQ_T (SQ Total); SQ_R (SQ de Regressão); SQ_r (SQ do Resíduo); SQ_{FaJ} (SQ de Falta de Ajuste); e SQ_{EP} (SQ de Erro Puro). Os componentes SQ_r e SQ_{FaJ} não foram expandidos pois comumente são obtidos por meio da subtração dos outros dois termos das respectivas equações.

$$SQ_T = SQ_R + SQ_r$$

$$\sum_{i=1}^m \sum_{j=1}^{nr} (y_{i,j} - \bar{y})^2 = \sum_{i=1}^m (\hat{y}_i - \bar{y})^2 + SQ_r$$

(3.3.1)

$$SQ_r = SQ_{EP} + SQ_{FaJ}$$

$$SQ_r = \sum_{i=1}^m \sum_{j=1}^{nr} (y_{i,j} - \bar{y}_i)^2 + SQ_{FaJ}$$

(3.3.2)

Enquanto isso, a MQ é dependente da SQ e pode ser descrita pela divisão da equação (3.3.3), que mostra a divisão entre a SQ e os Graus de Liberdade (GL) desta soma quadrática.

$$MQ = \frac{SQ}{GL}$$

(3.3.3)

Segundo a decomposição das variâncias, é construído a tabela ANOVA, representada no Quadro 7, onde os componentes que definem a variância podem ser descritos.

Quadro 7 – Representação teórica dos valores avaliados na tabela ANOVA.

Fonte de variação	Soma quadrática (SQ)	Graus de liberdade (GL)	Média quadrática (MQ)
Total	$SQ_T = \sum_{i=1}^m \sum_{j=1}^{nr} (y_{i,j} - \bar{y})^2$	$n - 1$	-
Regressão	$SQ_R = \sum_{i=1}^m (\hat{y}_i - \bar{y})^2$	$p - 1$	$MQ_R = \frac{SQ_R}{p - 1}$
Resíduos	$SQ_r = \sum_{i=1}^m \sum_{j=1}^{nr} (y_{i,j} - \hat{y}_i)^2$	$n - p$	$MQ_r = \frac{SQ_r}{n - p}$
Erro Puro	$SQ_{EP} = \sum_{i=1}^m \sum_{j=1}^{nr} (y_{i,j} - \bar{y}_i)^2$	$n - m$	$MQ_{EP} = \frac{SQ_{EP}}{n - m}$
Falta de ajuste	$SQ_{FaJ} = \sum_{i=1}^m (\hat{y}_i - \bar{y}_i)^2$	$m - p$	$MQ_{FaJ} = \frac{SQ_{FaJ}}{m - p}$

Fonte: Elaborado pelo autor. Adaptado de BRUNS, 2006.

Para iniciar a construção da tabela ANOVA, primeiramente precisamos da equação que define o sistema, como descrito nos capítulos anteriores. Após definido os coeficientes, podemos obter os valores calculados (previstos) da resposta, demonstrados por \hat{y} . Serão definidas, então, as variâncias principais do sistema. Como visto na equação (3.3.1), temos a equação (3.3.4):

$$\begin{aligned}
 SQT &= SQR + SQr \\
 \sum_{i=1}^m \sum_{j=1}^{nr} (y_{i,j} - \bar{y})^2 &= \sum_{i=1}^m (\hat{y}_i - \bar{y})^2 + SQr \\
 SQr &= \sum_{i=1}^m \sum_{j=1}^{nr} (y_{i,j} - \bar{y})^2 - \sum_{i=1}^m (\hat{y}_i - \bar{y})^2
 \end{aligned}
 \tag{3.3.4}$$

Com a resolução da equação (3.3.4) temos o resultado da SQ_T e SQ_R por meio do cálculo e da SQ_r por meio da diferença entre eles. A SQ_r será dividida nos seus dois componentes, demonstrado na equação (3.3.5), onde será necessário usar o valor da média de cada conjunto de réplicas.

$$\begin{aligned}
 SQr &= SQEP + SQFaJ \\
 SQr &= \sum_{i=1}^m \sum_{j=1}^{nr} (y_{i,j} - \bar{y}_i)^2 + SQFaJ
 \end{aligned}
 \tag{3.3.5}$$

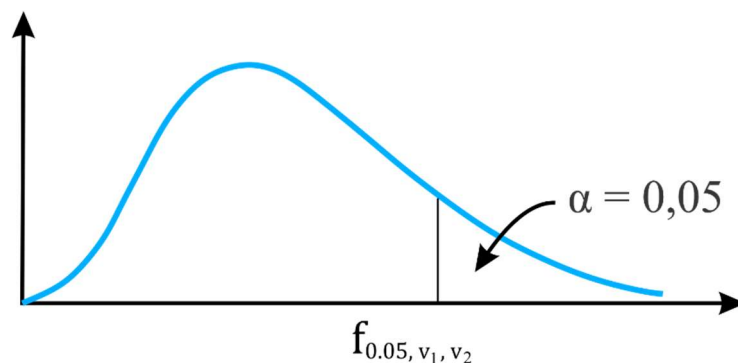
Após o cálculo de todas as somas quadráticas de variância, é possível montar uma tabela ANOVA como visto no Quadro 7.

A análise da tabela ANOVA permite avaliar a adequação do modelo estatístico que descreve o comportamento do sistema experimental. Nessa etapa, são considerados indicadores como o teste F, o valor de p, o coeficiente de determinação (R^2) e o coeficiente de correlação. Esses parâmetros devem ser analisados de forma integrada, uma vez que nenhum deles, isoladamente, é capaz de garantir a validade do modelo.

3.4.3.4. Teste F e p-valor

Após a construção da tabela ANOVA, é necessário realizar avaliações para garantir a confiança nos valores obtidos. O teste F representa um teste de hipóteses em relação à razão de duas variâncias e, nestes casos, analisa a hipótese de que as duas variâncias não apresentem diferença entre elas, também chamada de hipótese nula (H_0) (MINITAB, 2014). O Teste F se dá pela análise da distribuição F, demonstrada pela curva na Figura 18. Diferente de uma distribuição normal, esta não é simétrica. Cada combinação de GLs das variâncias e de nível de significância possui uma distribuição diferente. Os GLs da distribuição são demonstrados na sequência de numerador e denominador (MINITAB, 2019).

Figura 18 – Curva de distribuição F para sistema com 95% de grau de confiança.



Fonte: Elaborado pelo autor. Adaptado de MONTGOMERY, 2021.

Um valor isolado de F não demonstra muita informação. Para interpretar o valor F, é necessário calcular a probabilidade de observar uma estatística F tão alta quanto o valor obtido no estudo, isto é, qual a quantidade de vezes que uma hipótese nula apresentará um valor F maior do que o F calculado com os dados obtidos experimentalmente.

Esta probabilidade também pode ser interpretada de outra maneira, como o p-valor. O p-valor define o menor nível de significância que conduz à rejeição da hipótese nula (H_0) com os dados fornecidos (MONTGOMERY, 2021). O p-valor não indica a probabilidade da hipótese nula ser verdadeira ou falsa, pois seu cálculo se baseia na suposição de que essa hipótese nula seja verdadeira. Indica, então, quão provável é obter, em um estudo com hipótese nula aceita, uma variância tão grande quanto a obtida nos dados experimentais por efeito de erros aleatórios de amostragem (MINITAB, 2014).

Em uma análise de p-valor, é avaliado conforme o intervalo de confiança. Um intervalo de confiança de 95%, modelo mais comum a ser utilizado, possui um nível de significância α de 5%. Então, neste caso, para p-valores menores do que 5%, há evidências de que a variância

não é somente por um erro aleatório de amostragem, indicando que a hipótese nula possa ser rejeitada.

Na tabela ANOVA, analisa-se dois testes F: a similaridade da regressão com o resíduo; e a similaridade da falta de ajuste com o erro puro. No primeiro, representado na equação (3.4.1), tem-se a comparação da MQ de Regressão e de Resíduo, onde n representa a quantidade de experimentos e p a quantidade de parâmetros da equação. Pensando no experimento, a aceitação de H_0 indicaria que o efeito de regressão se confundiria com o resíduo, ou seja, não seria conhecido se a diferença do resultado teórico com o experimental se dá pela qualidade da regressão ou por algum efeito aleatório presente no resíduo, situação que representaria um modelo ruim. A expectativa, então, é que a hipótese nula de similaridade entre as variâncias seja rejeitada, e seja aceita a hipótese alternativa (H_1) de que as variâncias são estatisticamente diferentes.

$$F = \frac{MQR}{MQr} = \frac{SQR}{1} * \frac{n-p}{SQR} \quad (3.4.1)$$

Para saber se o valor obtido na equação (3.4.1) é um valor ideal, é necessário avaliar conforme o F tabelado e avaliar a proporção. O F tabelado deve ser avaliado conforme os GLs das MQs e o nível de significância α (comumente definido como 5%, indicando um intervalo de confiança de 95%), $F_{1,n-p,0,05}$. Para rejeitar H_0 , a razão entre F calculado e F tabelado deve ser maior que 1, e quanto maior for, menos provável será a similaridade estatística das variâncias.

Neste caso, também é possível avaliar pelo p-valor. Para isso, então, procura-se nos valores tabelados, qual valor de significância seria necessário para alcançar o F calculado, com base nos GLs das MQs. Para rejeitar H_0 , o p-valor esperado em um intervalo de confiança de 95%, deve ser menor do que 0,05, e, novamente, quanto menor o valor obtido, menos provável seria de que um estudo apresentasse as variâncias obtidas somente por causa de erros aleatórios.

Agora, para o segundo teste de similaridade de variâncias, representado pela equação (3.4.2), tem-se a comparação da MQ de Falta de Ajuste e de Erro Puro, onde n representa a quantidade de experimentos, p a quantidade de parâmetros da equação e m a quantidade de experimentos únicos (sem considerar as réplicas). Pensando no experimento, a aceitação da hipótese nula indicaria que o efeito de falta de ajuste se confundiria com o erro puro, ou seja, o

resíduo do modelo poderia ser considerado somente como a aleatoriedade dos experimentos (situação desejável). A expectativa, então, é que a hipótese nula de similaridade estatística entre as variâncias seja aceita.

$$F = \frac{MQFaJ}{MQEP} = \frac{SQFaJ}{m-p} * \frac{n-m}{SQEP} \quad (3.4.2)$$

Ao avaliar conforme a distribuição F, é esperado que a razão seja menor que 1, pois a probabilidade de apresentar uma similaridade estatística seria maior. Já para o p-valor, é esperado que seja um valor maior do que 0,05, que indicaria que há mais do que 5% de probabilidade de um experimento com hipótese nula aceita apresentar uma distribuição de dados igual à obtida experimentalmente e uma confiança de 95% não seria suficiente para rejeitar H_0 .

Como apresentado anteriormente, os testes devem ser feitos em conjunto, para que uma interpretação corrobore com a outra. Nenhum dos testes indica, por si só, se o modelo é adequado ou não para os experimentos.

3.4.3.5. Coeficiente de determinação e coeficiente de correlação

O coeficiente de determinação (R^2) é uma razão de soma dos quadrados, que pode ser definida pela equação (3.5.1), e é frequentemente usado para avaliar a adequação do modelo de regressão.

$$R^2 = \frac{SQR}{SQT} = 1 - \frac{SQr}{SQT} \quad (3.5.1)$$

Este é um coeficiente que tem seu valor máximo em 1 e seu resultado é interpretado como a porcentagem da variabilidade dos dados que é explicada pelo modelo, ou seja, um R^2 próximo de 1 indica que o modelo explica quase 100% da variabilidade dos dados.

Porém, é necessário analisar este coeficiente com cautela. O valor do coeficiente pode ser artificialmente aumentado conforme são adicionados termos de maior ordem, mesmo que o modelo se torne pior do que o anterior. Ou seja, embora o R^2 possa ser muito próximo de 1, isso não implica que o modelo de regressão forneça uma boa previsão dos dados experimentais futuros (MONTGOMERY, 2021).

Já o coeficiente de correlação mede a associação entre os valores de duas variáveis (x e y). Seus valores podem assumir entre -1 e $+1$. Em um sistema qualquer, um valor negativo é representante de uma correlação inversa (um valor diminui enquanto outro aumenta) e um valor positivo é representante de uma correlação direta (ambos aumentam ou diminuem juntos).

Para equações lineares, a correlação é feita pelo R de Pearson, enquanto para equações não-lineares, é possível determinar por meio do ρ de Spearman. Os cálculos envolvem equações mais complexas que não serão descritas neste trabalho.

Esse valor é obtido para cada par de variáveis e indica a direção (positiva e negativa) e força da associação entre elas. Porém, não indica necessariamente a qualidade da regressão linear, nem a existência da causalidade entre as variáveis, e sua função é próxima a de um comparativo, para verificar se a equação segue as mesmas correlações avaliadas com este coeficiente. Então, assim como o R^2 , deve-se ter cautela ao afirmar algo baseando-se somente deste valor.

3.4.3.6. Intervalo de confiança

Ao avaliar o descrito até então, vê-se que a variância cumpre um grande papel estatístico para o sistema. Na regressão, é relevante ter conhecimento da variância aceitável de um coeficiente com um grau de confiança relativo. Essa variância do modelo proposto será tratada como o intervalo de confiança da regressão.

Para encontrar esse intervalo de confiança (IC), um dos métodos utiliza das matrizes propostas anteriormente, por meio dos termos presentes na matriz $(X^tX)^{-1}$. A diagonal principal dessa matriz contém as variâncias de cada coeficiente. Considerando a matriz $(X^tX)^{-1}$ discutida na seção 3.4.3.2, o coeficiente b_i terá sua variância de acordo com a posição $a_{i+1,i+1}$. Ou seja, b_0 terá sua variância na posição 1,1 enquanto b_1 terá na posição 2,2.

Após obter a variância, é feito o cálculo da raiz quadrada do produto da MQ do Resíduo pela variância para obter o erro de um coeficiente ϵ_i . Com esse valor do erro, multiplica-se pelo valor de t de Student com graus de liberdade do resíduo, podendo ser subtraído de 1 (MONTGOMERY, 2021). As etapas de cálculo estão detalhadas na Figura 19.

Figura 19 - Demonstração dos cálculos para obter os valores do intervalo de confiança dos coeficientes de uma regressão linear.

$$(\mathbf{X}^t\mathbf{X})^{-1} = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{bmatrix}$$

$$\begin{aligned}
 \varepsilon_0 &= \sqrt{MQ_r * a_{1,1}} & IC_0 &= \varepsilon_0 * t_{\frac{\alpha}{2}, n-p-1} \\
 \varepsilon_1 &= \sqrt{MQ_r * a_{2,2}} & IC_1 &= \varepsilon_1 * t_{\frac{\alpha}{2}, n-p-1} \\
 \varepsilon_2 &= \sqrt{MQ_r * a_{3,3}} & IC_2 &= \varepsilon_2 * t_{\frac{\alpha}{2}, n-p-1}
 \end{aligned}$$

$$y = (b_0 \pm IC_0) + (b_1 \pm IC_1)x_1 + (b_2 \pm IC_2)x_2$$

Fonte: Elaborado pelo autor. Adaptado de PARTE 6, 2024 24:30 e de MONTGOMERY, 2021

Após o cálculo do IC, é avaliado se o intervalo do coeficiente terá interseção com o eixo de valor zero. Caso tenha, esse coeficiente é desconsiderado da equação, pois não será significativo para o modelo de regressão.

3.5. Definição da Estratégia de Controle do Procedimento Analítico

Esta etapa se passa antes da validação, e neste momento é elaborado uma versão inicial da estratégia de controle do procedimento analítico que deverá ser confirmado após a finalização da validação. Esta estratégia não possui um modelo pré-definido e geralmente segue normas internas da empresa que solicitou o desenvolvimento do método.

Esta estratégia lida com todos os controles existentes, sejam eles do próprio procedimento, controles ambientais e o Teste de Adequação do Sistema (SST – *system suitability test*). Todos são realizados para garantir que o procedimento atenda ao ATP. O SST contempla um conjunto de testes desenvolvidos para verificar se o procedimento está adequado para a análise pretendida. Isso aumenta a detectabilidade de certas falhas e é realizado antes de toda análise.

Cada SST é específico para um método e possui critérios pré-definidos para parâmetros específicos. Os componentes do SST devem ser escolhidos de acordo com uma análise de risco, junto dos dados de desenvolvimento. Alguns exemplos para componentes de teste para um método de HPLC seriam: repetibilidade da injeção; relação sinal-ruído; retenção; e fator de cauda.

Os CQAs definidos no ATP também tem seus controles definidos, com as condições, materiais ou critérios aceitáveis bem definidos e explicitamente especificados no procedimento.

São definidos também os pontos de trabalho que serão realizados, determinando cada parâmetro que foi avaliado durante o desenvolvimento e os intervalos ou valores dos parâmetros a serem utilizados na rotina. Estes pontos de trabalho são definidos não somente pelos resultados favoráveis durante a etapa de DoE, mas também por facilidades operacionais, como menor quantidade de solventes, menor custos ou menor tempo.

3.6. Validação

Durante a validação, os analistas seguirão os parâmetros definidos até agora no ATP e na Estratégia de Controle. Os parâmetros serão aplicados em um número mínimo de corridas para garantir a confiabilidade e repetibilidade do método nestas condições, avaliando os equipamentos, parâmetros, materiais e testes.

A partir da aprovação do método pelo estudo de validação, este está apto a ser utilizado na rotina do laboratório. A partir disso, qualquer alteração realizada para melhoria que esteja fora do escopo analisado nesta etapa de validação, necessitará de uma nova submissão para órgão regulatório.

O uso da ferramenta DoE ocasiona uma maior necessidade de testes durante a validação, para que sejam contemplados diversos pontos dentro do *design space* para garantir a qualidade dentro da sua área, aumentando o esforço nesta etapa. Porém, é aqui que apresenta sua maior vantagem regulatória, pois após validado o *design space*, espera-se que a quantidade de alterações pós-submissão seja menor quando comparados com procedimentos realizados com base em tentativa e erro ou OVAT.

4. CONSIDERAÇÕES FINAIS

O desenvolvimento e melhoria de fármacos é uma necessidade cada vez mais recorrente na indústria farmacêutica que, em geral, está melhorando seus paradigmas, buscando sempre a qualidade do produto e a segurança dos pacientes, porém de uma forma mais ágil. Com a disseminação da AQBd, os laboratórios buscam ainda mais o conhecimento prévio sobre seus produtos para que, durante o processo de análise, se possa economizar recursos e tempo, além de possuir uma maior garantia de que o produto comercializado está atendendo os critérios de qualidade tão prezados pelos fabricantes.

Com a ajuda dos métodos descritos neste trabalho é possível realizar um estudo mais preciso e eficaz sobre as melhores condições para a fabricação e análise dos fármacos. A

metodologia DoE é uma ferramenta potente para estudar detalhadamente o comportamento dos medicamentos e os parâmetros que afetam a qualidade do produto. Com essa ferramenta, é possível planejar experimentos de forma eficiente para obter um conhecimento minucioso sobre os efeitos que impactam nas características do fármaco, utilizando de experimentos mais enxutos e rápidos.

Esse estilo de lidar com o desenvolvimento analítico é ideal para um mercado cada vez mais rápido e mais demandante, possibilitando realizar melhorias nos medicamentos de forma ágil e construindo um sistema robusto de análise, reduzindo o impacto regulatório de um estudo de melhoria. É uma metodologia que se baseia muito mais no conhecimento e análise crítica do que em ensaios laboratoriais, sendo necessário elaborar uma rotina de trabalho adequada para garantir que está sendo analisado e testado o que é realmente necessário para o estudo.

O uso do DoE requer um maior conhecimento estatístico e o uso correto das ferramentas e testes presentes atualmente, porém é justamente esse requerimento que permite com que o método seja mais eficaz, conhecendo o máximo possível do método analítico com uma quantidade menor de análises e experimentos. Com uma melhora na qualidade dos analistas e o investimento devido às novas possibilidades, é evidente que o sistema da AQbD irá englobar a maioria dos novos estudos de laboratórios industriais em um futuro muito próximo.

REFERÊNCIAS BIBLIOGRÁFICAS

- AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA (ANVISA). **Guia nº 10/2017** para Tratamento Estatístico da Validação Analítica, versão 1. 2017. 19 p.
- INTERNATIONAL COUNCIL FOR HARMONISATION OF TECHNICAL REQUIREMENTS FOR FARMACEUTICALS FOR HUMAN USE. **Validation of Analytical Procedures Q2(R2)**. 2023. 36 p.
- INTERNATIONAL COUNCIL FOR HARMONISATION OF TECHNICAL REQUIREMENTS FOR FARMACEUTICALS FOR HUMAN USE. **Pharmaceutical development Q8(R2)**. 2009. 24 p.
- INTERNATIONAL COUNCIL FOR HARMONISATION OF TECHNICAL REQUIREMENTS FOR FARMACEUTICALS FOR HUMAN USE. **Analytical Procedure Development Q14**. 2023. 40 p.
- ARAÚJO, A. S. et al.; Key Information Related to Quality by Design (QbD) Applications in Analytical Methods Development. **Brazilian Journal of Analytical Chemistry**. São Paulo, v. 8, n. 30, p. 14-28, 2021. DOI: 10.30744/brjac.2179-3425.RV-27-2020.
- BRUNS, R. E.; SCARMINIO, I. S.; NETO, B. B. **Statistical Design – Chemometrics**. Amsterdam: Elsevier, 2006.
- CONCEIÇÃO L. R. V da et al. Ethanolysis Optimisation of Jupati (*Raphia taedigera* Mart.) Oil to Biodiesel Using Response Surface Methodology. **Journal of the Brazilian Chemical Society**. Brazil, v. 26, n. 7, p. 1321-1330, 2015. DOI: 10.5935/0103-5053.20150097.
- ILLOWSKY, B.; DEAN, S. **Introductory Statistics**. Texas: OpenStax, 2013. Disponível em: <<https://openstax.org/books/introductory-statistics/pages/1-introduction>>. Acesso em: 03 mar. 2025.
- MILLER, J. N.; MILLER, J. C. **Statistics and Chemometrics for Analytical Chemistry**, 6ed. Harlow: Pearson, 2010.
- MINITAB**. Entendendo Análise de Variância (ANOVA) e o teste F. 01 abr. 2019. Disponível em: <<https://blog.minitab.com/pt/entendendo-analise-de-variancia-anova-e-o-teste-f>>. Acesso em 29 abr. 2025.

MINITAB. How to Correctly Interpret P Values. 17 abr. 2014. Disponível em: <<https://blog.minitab.com/en/adventures-in-statistics-2/how-to-correctly-interpret-p-values>>.

Acesso em 29 abr. 2025.

MONTGOMERY, D. C.; RUNGER, G. C. **Applied Statistics And Probability for Engineers**, 7ed. Arizona: Arizona State University, 2021.

MOURE, R.B. et al. Definições e escopo dos elementos-chave do Analytical Quality by Design (AQbD) para o desenvolvimento de métodos na indústria farmacêutica. **Vigil Sanit Debate**, Rio de Janeiro, v.11, 2023, DOI: 10.22239/2317-269X.02162.

NOVAES, C. G. et al. Otimização de Métodos Analíticos Usando Metodologia de Superfícies de Resposta – Parte I: Variáveis de Processo. **Revista Virtual de Química**. [S. l.], v. 9, n. 3, p. 1184-1215, 2017. DOI: 10.21577/1984-6835.20170070.

PARTE 4: planejamento fatorial fracionário, vídeo 1 (DoE). [S. l.: s. n.], 2024. 1 vídeo (19 min).

Publicado por Edénir Pereira Filho. Disponível em:

<https://www.youtube.com/watch?v=fRfnwFnaS4o&list=PL4CuftF4l_fD0cWHdKY-W-_OP36Hhs4aw&index=17>. Acesso em: 08 fevereiro 2025.

PARTE 6: proposição de modelos, vídeo 1 (DoE). [S. l.: s. n.], 2024. 1 vídeo (34 min). Publicado

por Edénir Pereira Filho. Disponível em: <https://www.youtube.com/watch?v=n6l9N4P-br8&list=PL4CuftF4l_fD0cWHdKY-W-_OP36Hhs4aw&index=23>. Acesso em: 08 fevereiro

2025.

PARTE 8: proposição de modelos em um planejamento fatorial, vídeo 1 (DoE). [S. l.: s. n.],

2024. 1 vídeo (26 min). Publicado por Edénir Pereira Filho. Disponível em:

<https://www.youtube.com/watch?v=piqwmwyevlY&list=PL4CuftF4l_fD0cWHdKY-W-_OP36Hhs4aw&index=31>. Acesso em: 08 fevereiro 2025.

PARTE 11: planejamentos Doehlert e Box-Behnken, vídeo 1 (DoE). [S. l.: s. n.], 2024. 1 vídeo

(12 min). Publicado por Edénir Pereira Filho. Disponível em:

<https://www.youtube.com/watch?v=_VkuELO-GSg&list=PL4CuftF4l_fD0cWHdKY-W-_OP36Hhs4aw&index=38>. Acesso em: 08 fevereiro 2025.

PARTE 12: problemas com múltiplas respostas, vídeo 1 (DoE). [S. l.: s. n.], 2024. 1 vídeo (13

min). Publicado por Edénir Pereira Filho. Disponível em:

<https://www.youtube.com/watch?v=IV8PPoHDdxk&list=PL4CuftF4l_fD0cWHdKY-W-_OP36Hhs4aw&index=40>. Acesso em: 08 fevereiro 2025.

PENNSSTATE. **STAT 501 Regression Methods: Weighted Least Squares**. PennState Eberly College of Science. Disponível em: <<https://online.stat.psu.edu/stat501/lesson/13/13.1>>. Acesso em: 12 mar. 2025.

PEREIRA-FILHO, E. R.; PEREIRA, F. M. V.; LEMOS, S. G. Introducing Factorial Design Concepts Using Experiments with Multi-Colored Chocolates. **The Chemical Educator**. [S. l.], v. 14, p. 200-203, 2009. DOI: 10.1333/s00897092238a.

REZVANI, N. BOLDUC, DL. Monte Carlo Analysis. **Encyclopedia of Toxicology**, [S. l.], v.3, p. 393-396, 2014. DOI: 10.1016/B978-0-12-386454-3.00631-X.

SATHULURI, K. et al. Analytical quality by design (AQbD) in the ICHQ14 guidelines for analytical procedure development. **Accreditation and Quality Assurance**, [S. l.], v.30, p. 1-14, 2025. DOI: 10.1007/s00769-024-01587-w.

THODE, J. What are system suitability tests (SST) of analytical methods? **LOESUNGSFABRIK**. 28 jul. 2018. Disponível em: <<https://mpl.loesungsfabrik.de/en/english-blog/method-validation/sst>>. Acesso em: 24 mar. 2025.

TOME, T. et al. Development and Optimization of Liquid Chromatography Analytical Methods by Using AQbD Principles: Overview and Recent Advances. **Organic Process Research & Development**, [S. l.], v.23, n. 9, p. 1784-1802, 2019. DOI: 10.1021/acs.oprd.9b00238.

VENKATESH et al. Quality by Design: The Tool for Regulatory Compliance. **International Journal of Pharmaceutical Sciences and Research**. [S. l.], v. 10, n. 3, p. 1095-1100, 2019. DOI: 10.13040/IJPSR.0975-8232.10(3).1095-00.

APÊNDICE 01 – ATRIBUTOS DE DESEMPENHO

Os atributos de desempenho são aqueles que permitem avaliar o desempenho do método analítico por meio da avaliação dos resultados obtidos durante os experimentos, para garantir que apresentarão resultados satisfatórios e variabilidade aceitável. Possuem valores, normalmente, com base na tecnologia utilizada, porém com conceitos bem definidos.

A seguir, serão descritos os conceitos, com base em definições do ICH (**ICH Q2(R2), 2023**) (**ICH Q14, 2023**).

1. Seletividade/Especificidade

Ambos os termos são usados para descrever o quanto a determinação um analito é afetada por outras substâncias na mesma matriz. Especificidade normalmente define a medida inequívoca de um analito específico enquanto seletividade define a possibilidade de determinar analitos sem a interferência de outros compostos com comportamento similar.

Este parâmetro pode ser considerado com qualidade adequada por alguns métodos nas seguintes situações: (i) ao demonstrar a ausência de interferência pela presença de impurezas na matriz que poderiam estar presentes na amostra, como substâncias relacionadas ou de degradação; (ii) por comparação do resultado de um outro método analítico bem definido e com outro princípio de medida; (iii) ou por justificativa inerente da tecnologia e seus parâmetros técnicos, como em casos da resolução de isótopos em espectrometria de massas ou deslocamento na Espectroscopia de Ressonância Magnética Nuclear (RMN).

Tendo em vista as análises de doseamento feitas com técnicas de separação, a especificidade pode ser demonstrada pela resolução de dois componentes que eluem muito próximos ou pela comparação dos espectros. Para determinações do analito de interesse com impurezas na matriz, é possível realizar a adição seletiva das impurezas e comparar sua resposta à uma análise com baixos níveis de impureza, demonstrando que a resposta do analito não sofre interferência.

Caso não seja possível isolar as impurezas para realizar a adição, é possível fazer a comparação com outro procedimento ortogonal, isto é, um procedimento com diferentes princípios de medição. Isso é necessário para garantir a veracidade do resultado por meio de dois procedimentos independentes.

2. Faixa

Pode ser definido como o intervalo entre o menor e maior resultado contemplado no procedimento analítico com níveis adequados de precisão, exatidão e resposta. Pode ser validado pela avaliação dos resultados reportáveis, estabelecendo uma faixa adequada de análise. Os modelos que são empregados podem ser linear, não-linear ou multivariado.

A faixa reportável é o intervalo entre o menor e maior resultado reportáveis, isto é, um resultado gerado pelo procedimento após cálculos e processamentos aplicáveis, e geralmente variam de acordo com a especificação do produto para o procedimento analítico em questão. Em condições normais, é menor do que a faixa de trabalho, que configura-se como a faixa contemplada pelo instrumento analítico com resultados confiáveis. Ambas as faixas podem ser idênticas. Normalmente é necessário realizar alguma transformação matemática da faixa de trabalho conhecida para chegar aos resultados reportáveis.

Em casos em que não seja possível obter materiais de pureza suficiente para validar o alcance total (até 100% da pureza), é possível extrapolar o resultado.

O Quadro A1 exemplifica alguns alcances reportáveis de acordo com o procedimento.

Quadro A1 – Exemplos de faixas reportáveis para procedimentos analíticos comuns.

Procedimento analítico	Extremidade inferior da faixa reportável	Extremidade superior da faixa reportável
Doseamento	80% do conteúdo declarado ou 80% do limite inferior do critério de aceitação da especificação	120% do conteúdo declarado ou 120% do limite superior do critério de aceitação da especificação
Uniformidade de conteúdo	70% do conteúdo declarado	130% do conteúdo declarado
Impurezas	Limite reportável	120% do critério de aceitação da especificação

Fonte: Elaborado pelo autor. Adaptado de ICH Q2(R2), 2023.

3. Resposta

A resposta de um procedimento analítico é a habilidade de obter um sinal que se relaciona à concentração na amostra com base em alguma função matemática conhecida.

Já a resposta linear é um item avaliado por meio de uma curva analítica, que pode ser preparada por diversos métodos como: padronização externa, padronização interna, superposição de matriz e adição padrão.

4. Limite de Detecção (LOD) e Limite de Quantificação (LOQ)

O LOD (*Limit of Detection*) é definido como a menor quantidade de um analito em uma amostra que pode ser detectada pela análise, mas não necessariamente quantificada com exatidão.

Já o LOQ (*Limit of Quantification*) é a menor quantidade de um analito em uma amostra que pode ser definida quantitativamente com adequada exatidão e precisão. É mais usado em métodos de determinação de substâncias relacionadas/impurezas, pois trabalha com valores ínfimos de concentração na amostra.

5. Exatidão e Precisão

Exatidão pode ser definida como a proximidade da concordância entre um valor de referência aceito externamente e um valor medido.

Já a precisão pode ser definida como a proximidade de um grau de dispersão entre diferentes medidas obtidas de múltiplas amostras da mesma solução homogênea. Pode ser considerada em três níveis: repetibilidade, precisão intermediária e reprodutibilidade.

A repetibilidade contempla a precisão dos resultados de análise realizada em mesmas condições operacionais após um pequeno intervalo de tempo. A precisão intermediária contempla as variações dentro do laboratório, e leva em conta fatores ambientais como diferentes dias ou condições climáticas e atmosféricas, além de condições operacionais como diferença de analista ou diferença de equipamento. E a reprodutibilidade contempla a precisão entre diferentes laboratórios, geralmente aplicados para a padronização do método entre diferentes locais.

6. Robustez

A robustez pode ser definida como uma medida da capacidade da resposta em permanecer inalterada por pequenas, mas deliberadas variações nos parâmetros do método e provém uma indicação da sua confiabilidade durante o uso de rotina.

Seu teste é dado normalmente por meio de pequenas variações dos parâmetros do procedimento analítico, porém, a partir dos experimentos realizados durante o DoE, é possível construir uma Região de Concepção Operacional do Método (MODR – *Method Operable*

Design Region) ou também conhecido como *design space*. O MODR será explicado depois que seja definido o conceito de superfície de resposta.

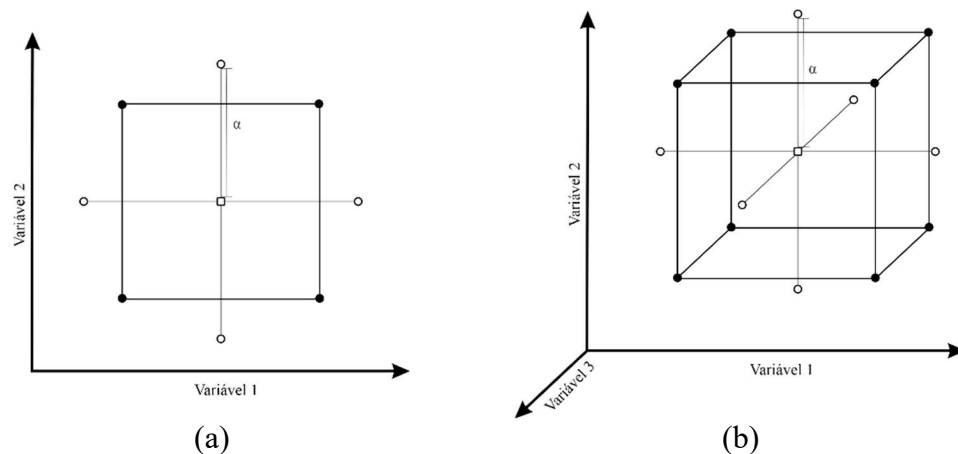
Seguindo as agências regulatórias, é recomendado que a robustez seja demonstrada durante o desenvolvimento do método. Por uma visão regulatória, a análise da robustez é essencial para economia de recursos, pois uma MODR bem construída permite a aplicação do procedimento em toda sua área sem ser considerada como uma mudança do procedimento. Isso permite uma aplicação mais ampla sem a necessidade de revalidações de todo o procedimento.

APÊNDICE 02 – PLANEJAMENTOS DE OTIMIZAÇÃO

1. Planejamento Composto Central (CCD)

O CCD é um modelo próximo dos planejamentos fatoriais descritos na triagem. O detalhamento deste modelo pode ser feito pela divisão em três componentes, iniciando por um planejamento fatorial, representado pelos pontos pretos na Figura A1. É criado um ponto central representando a média entre os níveis das variáveis (0,0), representado pelo quadrado branco na Figura A1, e pontos axiais que passam pelo centro de uma variável e estão a uma distância α do ponto central ($0, p.c.\pm\alpha$) e ($p.c.\pm\alpha, 0$), representado pelos pontos brancos na Figura A1. Esse modelo também é referenciado como fatorial com estrela devido à disposição formada pelos pontos axiais.

Figura A1 – Pontos experimentais para geração do planejamento composto central (a) para duas variáveis e (b) para três variáveis.



Fonte: Elaborado pelo autor. Adaptado de NOVAES, 2017.

O valor de α é calculado de acordo com o número de variáveis, e varia de 1 à \sqrt{k} . Quando $\alpha = \sqrt{k}$, o design se torna esférico, pois todos os pontos (exceto o ponto central) se encontram na mesma circunferência e ocasiona uma quantidade maior de níveis. Já no caso de $\alpha = 1$, os pontos axiais se encontram nas faces do cubo formado pelos pontos derivados do planejamento fatorial completo e ocasiona um estudo com somente três níveis, mas com mais replicatas em cada nível (BRUNS, 2006) (NOVAES, 2017).

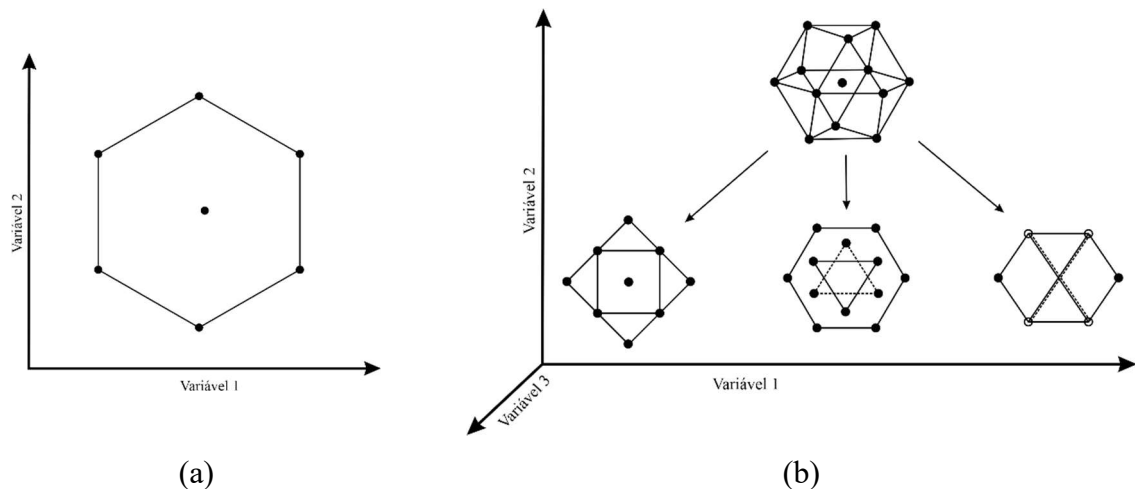
2. Planejamento Doehlert (DD)

Pode-se considerar o DD como uma opção prática e econômica em relação aos outros modelos, como visto no Quadro 6, em que necessita menos experimentos comparado às outras opções. Sua peculiaridade é de que suas matrizes não são rotáveis e os pontos não são ortogonais, isto é, não há uma perpendicularidade entre as posições geométricas de cada experimento. Consequentemente, é necessário saber qual variável será atribuída à cada eixo antes de iniciar os experimentos pois isso influenciará a quantidade de níveis representados para cada variável.

Pela forma geométrica de hexágono vista na Figura A2a, referente a um planejamento de duas variáveis, uma das variáveis está sendo analisada em 5 níveis, com distância codificada de 0,5 entre eles, enquanto a outra em 3 níveis, com distância codificada de 0,866. Isso é uma característica ideal para o estudo de variáveis com diferentes relevâncias para o sistema, ou até mesmo que não possuam tantos níveis quanto à outra variável.

Já ao analisar a Figura A2b, com formato de cubooctaedro, nota-se que o planejamento possa ser projetado de três diferentes formas em um plano, podendo gerar diferentes matrizes experimentais durante a otimização. Nestas combinações, as variáveis podem possuir uma quantidade de níveis de 3, 5 ou 7.

Figura A2 - Pontos experimentais para geração do planejamento Doehlert (a) para duas variáveis e (b) para três variáveis, com as possíveis projeções do sólido no plano.

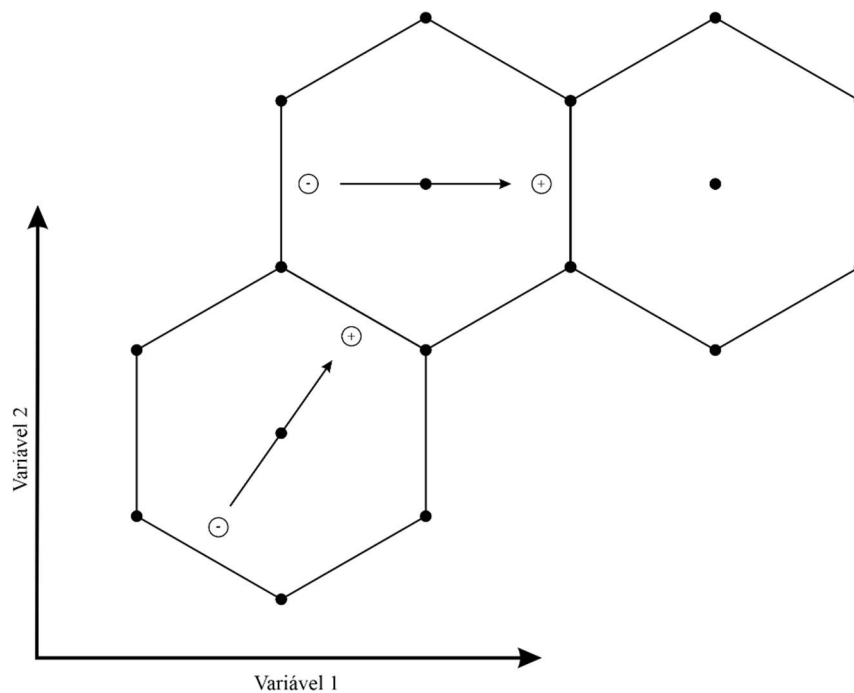


Fonte: Elaborado pelo autor. Adaptado de NOVAES, 2017.

Este planejamento é montado a partir de um modelo Simplex, com uma rotação da forma geométrica ao redor do ponto central. Na Figura A2a, o hexágono é formado da rotação

uniforme de um triângulo equilátero ao redor do ponto central (BRUNS, 2006). O modelo Simplex se baseia em analisar as formas geométricas e, a depender de qual vértice obteve o pior resultado, deslocar os pontos experimentais por entre a aresta oposta para analisar um novo ponto que possa estar mais próximo do resultado ótimo (MILLER, 2010). Por ser derivado de um modelo Simplex, o DD permite realizar esse deslocamento com as estruturas formadas em seu planejamento, como pode ser visto na Figura A3.

Figura A3 – Exemplo de movimentação de níveis em um planejamento Doehlert de duas variáveis.



Fonte: Elaborado pelo autor.

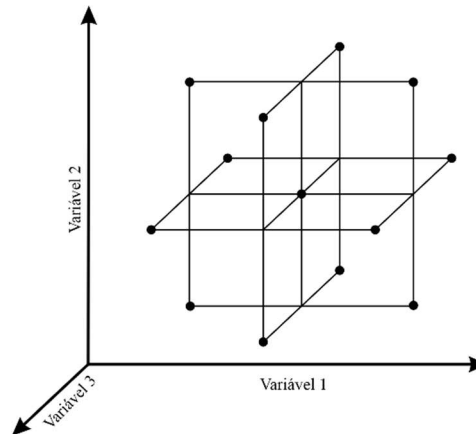
3. Planejamento Box-Behnken (BBD)

De forma geral, o BBD é um planejamento com menos combinações de níveis entre as variáveis pois só necessita de 3 níveis para cada (-1, 0, +1), sem requerer de um ponto mais distante do que outros, como os pontos axiais do CCD, nem uma quantidade diferente de níveis para cada variável, como o DD.

Ao analisar a distribuição geométrica de um modelo de três variáveis do BBD, como visto na Figura A4, nota-se nitidamente que se trata de 3 planos cruzados, como um cubo sem os vértices. Isso se deve pois é feito a partir de uma combinação de planejamentos fatoriais de dois níveis de duas variáveis enquanto a terceira variável é definida no nível intermediário (± 1 , ± 1 , 0). Conseqüentemente, este planejamento não trabalha com os extremos das variáveis (± 1 , ± 1 , ± 1). Essa configuração dos pontos experimentais limita o modelo, porém forma seu nicho

de uso, sendo ideal para analisar sistemas em que não é possível sustentar um experimento com todas as variáveis em nível alto ou baixo. Utilizando as variáveis citadas na triagem como exemplo, para um sistema em que a vazão de solvente em nível alto com uma pressão em nível alto não seja adequado para o uso de uma composição de fase móvel determinada como nível alto ou baixo, o BBD é ideal (**BRUNS, 2006**).

Figura A4 - Pontos experimentais para geração do planejamento Box-Behnken de três variáveis.



Fonte: Elaborado pelo autor. Adaptado de NOVAES, 2017.

Neste planejamento, o menor número de variáveis possíveis de serem testadas é três. Isso é consequente da remoção dos extremos, pois em um sistema de duas variáveis, após a remoção dos extremos não restará pontos suficientes para garantir uma distribuição homogênea e garantir uma estimativa confiável dos coeficientes.

APÊNDICE 03 – MÉTODO DOS MÍNIMOS QUADRADOS PARA CASOS COM DADOS HETEROCEDÁSTICOS

Para definir qual dos métodos entre o MMQO e MMQP será utilizado, é feito um teste de uniformidade das variâncias, avaliando se as variâncias dos pontos experimentais são iguais para qualquer nível da variável analisada, situação denominada homoscedasticidade, ou diferentes conforme se altera o nível da variável, situação denominada heterocedasticidade. Esse teste é chamado de Teste de Cochran (**MILLER, 2010**).

O Teste de Cochran se baseia na comparação estatística entre a maior variância com a soma de todas as variâncias da regressão por meio da proporção entre elas, como visto na equação (3.2.6). Para isso, é primeiro calculado a variância de cada ponto experimental por meio da equação (3.2.7), onde y representa o valor obtido experimentalmente, i para a quantidade de pontos experimentais e j para a quantidade de réplicas, e m a quantidade total de réplicas no ponto experimental i .

$$C = \frac{S_{y,max}^2}{\sum_i^n S_{y_i}^2} \tag{A3.1}$$

$$S_{y_i}^2 = \frac{1}{m-1} \sum_{j=1}^m (y_{ij} - \bar{y}_i)^2 \tag{A3.2}$$

O valor de C obtido será, então, comparado com o valor tabelado para um nível de significância de 5%. Considerando que o C seja menor que o valor de $C_{tabelado}$, aceita-se a hipótese nula (H_0), de que as variâncias são todas estatisticamente iguais e os dados são homocedásticos. Caso o C seja maior ou igual ao $C_{tabelado}$, rejeita-se a hipótese nula, e considera-se que pelo menos uma das variâncias é estatisticamente diferente e os dados são heterocedásticos. Os valores de C tabelados podem ser obtidos do produto do número de pontos experimentais pela quantidade de réplicas para cada ponto.

Com a definição da cedasticidade dos dados, temos que definir o método. Com um conjunto de dados homocedásticos, não há necessidade de pesos para cada variância pois todas são estatisticamente iguais, e por isso é possível utilizar o MMQO, como foi demonstrado anteriormente.

Agora, considerando um caso de dados heterocedásticos, temos que levar em consideração a diferença da variância dos pontos. Para isso, é necessário atribuir pesos para a variância s_i^2 de cada ponto experimental i , por meio da equação (3.2.8).

$$w_i = \frac{s_i^{-2}}{\sum_i \frac{s_i^{-2}}{n}} \quad (\text{A3.3})$$

Para uma regressão linear, é necessário considerar o peso dos coeficientes na equação que define o sistema. Adaptando a equação (3.2.1), a equação (3.2.9) define o sistema com os pesos dos dados.

$$y_n = \beta_0 + \sum_{i=1}^k w_i \beta_i x_{in} + \varepsilon_n \quad (\text{A3.4})$$

Para uma visão matricial do MMQP, é possível assumir uma matriz de significâncias tal qual na Figura A5a, que pode ser transformada em uma matriz de pesos por meio da inversão das variâncias, tal qual na Figura A5b, onde $w_i = 1/\sigma_i^2$. **(PENNSTATE)**.

Figura A5 – Exemplo de matriz para o Método dos Mínimos Quadrados Ponderados contendo (a) as variâncias de cada ponto experimental e (b) os pesos que cada ponto representa.

$$\begin{array}{c} \left[\begin{array}{cccc} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \sigma_n^2 \end{array} \right] \\ \text{(a)} \end{array} \quad W = \quad \begin{array}{c} \left[\begin{array}{cccc} w_1^2 & 0 & \dots & 0 \\ 0 & w_2^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & w_n^2 \end{array} \right] \\ \text{(b)} \end{array}$$

Fonte: Elaborado pelo autor. Adaptado de PENNSTATE.

Com as matrizes da Figura A5, é possível definir que o cálculo da matriz de coeficientes \hat{B} seja definida pela equação (3.2.10).

$$\hat{B} = (X^t W X)^{-1} X^t W Y \quad (\text{A3.5})$$