

Aplicação da Elipse de Confiança como apoio à análise de precisão na medição de concentração de emulsões de água em óleo (W/O)

Ronã Rinston Amaury Mendes (IFSULDEMINAS)

Matheus Costa Pereira (Unifei)

Michele de Santana Carmelossi (Unifei)

Vinícius Montgomery de Miranda (Unifei)



Este trabalho objetiva apresentar uma aplicação da estatística multivariada mediante o uso da elipse de confiança. Seu conceito e aplicação é de grande importância e tem como proposição manter os valores esperados das variáveis que envolvem um complexo conjunto de dados próximo ao desejável e com a mínima variabilidade. Baseada na matriz de variância-covariância, a elipse de confiança é uma representação gráfica bidimensional dessa variabilidade.

As elipses de confiança apresentadas neste trabalho baseiam-se na pesquisa de Paiva (2004) que analisa a precisão na medição de concentração de emulsões de água em óleo (W/O), processo muito utilizado principalmente na indústria do petróleo. Foram utilizadas respostas altamente correlacionadas de um arranjo real para a primeira elipse e resultados simulados, mas dentro dos parâmetros do primeiro arranjo para a segunda elipse. Demonstrou-se com essas elipses a dispersão das emulsões estudadas com o objetivo de facilitar a análise e a tomada de decisão diante de um processo produtivo.

Evidencia-se, assim, a importante contribuição que o uso de técnicas estatísticas avançadas aliadas à simulação computacional pode proporcionar para o desenvolvimento e a otimização de processos produtivos dos mais diversos.

Palavras-chave: Elipse de confiança, análise estatística multivariada, água em óleo, processos produtivos.

1. Introdução

Conhecer, entender e controlar a variabilidade nos processos é essencial em setores manufatureiros e coletar dados confiáveis para essa tratativa e análise é uma tarefa demasiadamente desafiadora. Desvios em processos podem se dar por inúmeros fatores como ambientais, limitações operacionais ou até mesmo ligados a variações presentes nas matérias-primas. Aplicações estatísticas sistematizadas que aprimorem a apresentação de regiões de confiança para a análise dessa variabilidade desses processos são sempre desejáveis.

Para Silva e Santos (2020) o impacto dessa variação não tratada devidamente poderá ser observado internamente no comprometimento da qualidade do produto e aumento dos custos de produção e externamente na diminuição do nível de satisfação do cliente e na perda de competitividade no mercado.

Importante papel é reservado ao estudo e análise da região de confiança dos dados multivariados provenientes desses processos. Trata-se este de um intervalo estimado que indica uma faixa de valores com certa probabilidade de confiança para um determinado conjunto de parâmetros.

Em aplicações multivariadas e para processos com duas dimensões, sua representação visual se dá através de uma elipse de confiança formada com certo conjunto de pontos, gerados com certa medida de incerteza associada às estimativas dadas, mas propensa a conter os verdadeiros valores dos parâmetros.

Coube a Saporta e Hatabian (1986) e Saporta (1990) descrever como a plotagem da elipse de confiança é feita, de acordo com categorias na análise fatorial. Sua construção leva em consideração as correlações entre as variáveis e a variabilidade dentro do grupo de dados e permite uma representação mais precisa da relação entre esses pontos e uma interpretação mais clara dos resultados. Padrões incomuns nos dados podem fazer com que alguns pontos extremos se localizem fora da região da elipse. Análises complementares podem ajudar a identificar sua discrepância e tratar esses pontos.

Conforme Schelp (2018), usa-se a elipse de confiança para se obter a geometria correta entre duas variáveis com base em suas médias, desvio padrão e no coeficiente de correlação de Pearson. Ela tem uma ampla aplicação prática com a vantagem de simplificar e deixar mais intuitiva uma análise multivariada, ao mesmo tempo que fornece uma abordagem estatística robusta para a compreensão e interpretação de dados complexos.

Além de ser uma ferramenta eficiente e flexível na representação visual da região de confiança, também possui propriedades matemáticas que a habilitam estimar intervalos de confiança, realizar testes de hipóteses e fazer comparações entre grupos de dados. Através do uso da

estatística multivariada, diversas aplicações podem ser feitas como controle da qualidade de produtos manufaturados, monitoramento da variabilidade de características importantes ao longo do tempo, como também a identificação de variações em processos produtivos, auxiliando no direcionamento de ações corretivas para melhoria da qualidade do produto.

O presente trabalho tem como intuito contribuir com um método de geração de uma elipse de confiança usando-se o software Microsoft Excel[®]. Utiliza como objeto de estudo a pesquisa feita por Paiva (2004) que analisa a precisão na medição de concentração de emulsões de água em óleo (W/O) com base nos fatores altamente correlacionados média e coeficiente de variação.

2. Equacionamento da elipse de confiança

Uma elipse é uma figura plana, bidimensional, obtida da secção de um plano que corta uma superfície cônica. Na Análise Estatística Multivariada ela representa a incerteza contida nas estimativas, como médias, desvios padrão e outras medidas de tendência central e variabilidade, onde sua região representa certa proporção de observações com base na distribuição de determinado conjunto de dados. Ela pode ser inicialmente definida pela equação (2.1):

$$\left(\frac{x_1-h}{a}\right)^2 + \left(\frac{x_2-k}{b}\right)^2 = 1 \quad (2.1)$$

Com o desenvolvimento de equacionamentos, tem-se sua definição pela equação (2.2)

$$\mathbf{x}^T \Sigma \mathbf{x} = [x_1 \quad x_2] \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1^2 \sigma_1^2 + x_2^2 \sigma_2^2 + 2\sigma_{12} x_1 x_2 \quad (2.2)$$

A construção da elipse de confiança se dá pela matriz de variância-covariância e utiliza a relação entre o coeficiente de correlação e a covariância. Ela possibilita visualizar e interpretar as relações existentes entre duas variáveis de determinado conjunto de dados, representadas em eixos, e fazer complexas avaliações comparando e buscando padrões entre grupos ou fazendo o tratamento estatístico de estudos experimentais. Seu eixo maior e mais influente é chamado principal e o eixo secundário será seu ortogonal.

A elipse representa a região em que um certo percentual de distribuição de probabilidade está contido, de acordo com certo nível de confiança. Sua área varia de acordo do nível de confiança e da quantidade de dados disponíveis para a análise e indica, a relação entre as variâncias e covariâncias das variáveis em questão e a função de densidade de probabilidade multivariada dada pela equação (2.3).

$$(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) = \frac{1}{1-\rho_{12}^2} \left[\left(\frac{x_1-\mu_1}{\sigma_1} \right)^2 - 2\rho_{12} \left(\frac{x_1-\mu_1}{\sigma_1} \right) \left(\frac{x_2-\mu_2}{\sigma_2} \right) + \left(\frac{x_2-\mu_2}{\sigma_2} \right)^2 \right] \quad (2.3)$$

As informações sobre as variáveis e suas relações em um conjunto de dados multivariados constam da matriz de variância-covariância (Σ), uma matriz simétrica e quadrada, onde o sinal da sua covariância sinaliza sua positividade ou negatividade, os elementos de sua diagonal são as variâncias de cada uma das variáveis e os elementos fora de sua diagonal são as covariâncias. Essa matriz 2×2 Σ é importante para a compreensão da relação entre as variáveis e para a identificação de padrões e tendências nos dados multivariados, descrita pela equação (2.4).

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} \quad (2.4)$$

Os graus de liberdade (p) da matriz de variância-covariância são calculados a partir do número de observações (n) e de variáveis (a) do conjunto de dados e representam a quantidade de restrições ou limitações das variáveis e determina um número independente de direções, apresentada pela equação (2.5), de acordo com Montgomery e Runger. (2010).

$$p = n - a \quad (2.5)$$

Já o coeficiente de correlação de Pearson (ρ) indica a força e a direção da relação entre duas variáveis quantitativas, assumindo valores em um intervalo fechado entre -1 (indicando uma forte correlação negativa) a +1 (indicando uma forte correlação positiva) entre as variáveis.

Uma correlação igual a 1 é considerada perfeita. Suas variâncias são iguais e a rotação máxima é de 45° . De forma oposta, seu mínimo é 0 quando a correlação também tem também esse valor. Ela pode ser definida como a covariância padronizada quando dividida pelo produto dos desvios padrão das variáveis em questão, conforme dado pela equação (2.6).

$$\rho = \frac{\sigma_{x_1 x_2}}{\sqrt{\sigma_{x_1}^2 \sigma_{x_2}^2}} \quad (2.6)$$

O ângulo de rotação (θ), dado em radianos, é algo que serve para diagonalizar a matriz fazendo com que ele esteja alinhado com os seus respectivos eixos e é definido pela equação (2.7).

$$\theta = \frac{1}{2} \arctan \left[\frac{2\rho_{12} \sqrt{\sigma_1^2 \sigma_2^2}}{(\sigma_1^2 - \sigma_2^2)} \right] \quad (2.7)$$

A matriz quadrática de rotação simétrica e ortogonal (\mathbf{P}) também é conhecida como matriz de autovetores, utilizada para realizar uma rotação no espaço geométrico e descrever a mudança de orientação e de posição. Os autovetores das matrizes podem ser revelados a partir da matriz de rotação. Diante disso, ela possui importância na transformação de dados originais em variáveis secundárias oriundas da matriz. A equação (2.8) demonstra a matriz de rotação baseada no ângulo de rotação dado pela equação (2.7).

$$\mathbf{P} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \quad (2.8)$$

A matriz de autovalores ($\mathbf{\Lambda}$) é dada pela decomposição espectral, também conhecida como decomposição em autovalores. Ela decompõe uma matriz quadrada em autovalores e autovetores correspondentes e, nela, é possível identificar a variação dos dados explicada por cada componente principal. Ainda, tem a função de expressar a matriz como uma combinação linear de seus autovetores e autovalores e sua decomposição é dada pela equação (2.9) e algebricamente representada pela equação (2.10).

$$\mathbf{\Lambda} = \mathbf{P}^T \mathbf{\Sigma} \mathbf{P} \quad (2.9)$$

$$\mathbf{\Sigma} = \sum_{i=1}^p \lambda_i \mathbf{e}_i \mathbf{e}_i^T \quad (2.10)$$

O método matemático de Cholesky, conforme Silva e Santos (2019), é utilizado para decompor uma matriz simétrica e definida positiva em um produto matricial de duas matrizes, uma triangular inferior e uma transposta de sua matriz conjugada. A decomposição de Cholesky simplifica o processo de solução de equações lineares e é essencial para analisar propriedades como a positividade matricial. Ela possui suma importância na análise estatística multivariada, pois permite a análise de componentes principais e a estimação de modelos de regressão linear e a geração de dados multivariados normais.

A matriz de Cholesky (\mathbf{L}^T) é uma triangular superior associada a uma matriz $\mathbf{\Sigma}$ e é obtida pela técnica conhecida como decomposição de Cholesky, exibida na equação (2.11).

$$\mathbf{L}^T = \mathbf{X} \mathbf{\mu}^{-1} \mathbf{Z}^{-1} \quad (2.11)$$

Cabe ao vetor das médias ($\mathbf{\mu}$) em uma elipse de confiança definir o seu centroide, ou seja, descrever seu centro e sua forma de distribuição multivariada. A elipse não terá seu formato alterado pela média, pois ela propicia realizar a translação, mas não a rotação, ou seja, apenas irá mudar sua posição no espaço. Ela é dada pela equação (2.12).

$$\mathbf{\mu}^T = [\mu_x \quad \mu_y] \quad (2.12)$$

O grau de confiança (α) representa a probabilidade de a elipse conter a média amostral e seu valor pode variar entre 0 e 1. Para garantir alta confiabilidade para a elipse é indicado utilizar um grau de confiança de 95%. Conforme este valor é aumentado, maior será a quantidade de pontos contidas dentro da elipse e, também, maior será o tamanho da amostra necessária.

O qui-quadrado (χ^2) é importante na avaliação e adequação dos modelos teóricos e na observação de dados. Utilizado para avaliar a diferença entre as frequências observadas e as frequências esperadas, quanto maior seu valor, maior será a diferença entre essas frequências. Indicando que a distribuição teórica não é a adequada e não se ajusta bem em relação à

observação dos dados. O χ^2 pode ser representado pela equação (2.13) ou também pela equação (2.14) que relaciona o grau de confiança e o grau de liberdade pelo inverso da distribuição do χ^2 .

$$\chi^2 = \sum_{i=1}^k \sum_{j=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (2.13)$$

$$\chi^2 = INV.QUI(\alpha, p) \quad (2.14)$$

A matriz de correlação residual padronizada (Ψ) serve para calcular o tamanho dos semieixos da elipse de confiança. Ela é encontrada pela aplicação da análise de correlação em uma matriz de resíduos padronizados.

Para esse cálculo, além do qui-quadrado, também será utilizada a matriz de autovalores. Essa matriz analisa os efeitos e avalia a forma e a direção das relações entre as variáveis do modelo em questão. Assim, a equação (2.15) representa o comprimento dos semieixos dada pela correlação residual padronizada.

$$\Psi = \sqrt{\chi^2 \Lambda} \quad (2.15)$$

Os raios internos e externos da elipse são dados a partir da avaliação da matriz de correlação residual padronizada, eles estão relacionados com a diagonal principal dessa matriz, o termo a_{11} é o raio interno, enquanto o termo a_{22} representa o raio externo. Os raios são exibidos na equação (2.16) e na equação (2.17).

$$r_i = \left(\sqrt{\chi^2 \Lambda} \right)_{11} \quad (2.16)$$

$$r_e = \left(\sqrt{\chi^2 \Lambda} \right)_{22} \quad (2.17)$$

Os componentes principais (**PC**) presentes em uma elipse de confiança estão intrinsecamente relacionados aos eixos que definem a orientação da elipse. O **PC₁** está associado ao eixo primário, de maior variância, enquanto o **PC₂** está associado ao eixo secundário, de menor variância. Esses componentes principais são obtidos por meio da decomposição espectral da matriz de correlação residual padronizada. Sua análise proporciona importantes informações sobre as relações existentes entre as variáveis.

Encontrados os componentes principais, passa-se a definir os eixos primário, referente às coordenadas do **PC₁**, e o secundário, referente às coordenadas do **PC₂**. Assim, as equações (2.18) a (2.21) trazem os componentes 1 e 2 nos eixos x e y. Eles precisam ser somados ao vetor média para serem plotados de forma correta.

$$PC_{x_1} = \left(\sqrt{\chi^2 \Lambda} \right)_{11} [\cos(\theta)] \quad (2.18)$$

$$PC_{y_1} = \left(\sqrt{\chi^2\Lambda}\right)_{11} [\text{sen}(\theta)] \quad (2.19)$$

$$PC_{x_2} = \left(\sqrt{\chi^2\Lambda}\right)_{22} [-\text{sen}(\theta)] \quad (2.20)$$

$$PC_{y_2} = \left(\sqrt{\chi^2\Lambda}\right)_{22} [\cos(\theta)] \quad (2.21)$$

De grande importância, a medida de distância desenvolvida por Mahalanobis (1936) e que traz seu nome, considera a covariância entre as variáveis em um conjunto de dados multivariados. Seu cômputo se dá por meio da transformação linear dos dados e possibilita avaliar a distância entre o centro e um ponto qualquer de um conjunto de dados. A equação (2.22) representa a distância de Mahalanobis e a equação (2.23) é um modo de obter a matriz de covariância para esta distribuição em um conjunto de dados.

$$d = \sqrt{(\mathbf{x}_0 - \boldsymbol{\mu})^T - \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})} \quad (2.22)$$

$$\boldsymbol{\Psi} = \boldsymbol{\Sigma} - \boldsymbol{\mu}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^T - \boldsymbol{\mu}^T \quad (2.23)$$

A relevância de se apresentar o método de construção de uma elipse se justifica pela forma objetiva com que ela permite visualizar o grau de incerteza que um conjunto de dados possui. Ela pode ser construída com o auxílio da equação (2.24).

$$\begin{bmatrix} x_1^* \\ x_2^* \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} e_{11} & e_{12} \\ e_{21} & e_{22} \end{bmatrix} \begin{bmatrix} \sqrt{x^2\lambda_1} & 0 \\ 0 & \sqrt{x^2\lambda_2} \end{bmatrix} \begin{bmatrix} \cos\beta \\ \sin\beta \end{bmatrix} \quad 0 \leq \beta \leq 2\pi \quad (2.24)$$

A variância é uma medida estatística de dispersão que representa a magnitude da diferença de cada valor de um conjunto de dados e sua média, enquanto a covariância descreve a relação linear entre duas variáveis, ou seja, descreve como duas variáveis se movem juntas em relação às suas médias. Elas são medidas estatísticas que não possuem significância física. Entretanto, a correlação é a medida padronizada da covariância e a partir dela é possível descobrir a força e a direção entre as duas variáveis, sendo mais facilmente compreendida do que a covariância.

Um importante conceito sobre a covariância é que se o valor da relação entre as variáveis for igual a 0 significa que há uma independência entre elas.

Embora nem todas as matrizes os possuem, os autovetores e autovalores são de suma importância nas análises estatísticas multivariadas. Um autovetor é um vetor não nulo associado a uma matriz quadrada que não muda de direção ao ser multiplicado por essa matriz, resultando em um múltiplo escalar do mesmo vetor. Já o autovalor é um fator escalar correspondente ao autovetor, que multiplicado por um vetor não nulo, após a aplicação de uma transformação não linear, produz um novo vetor. Cada autovalor tem associado a ele um conjunto de autovetores. Matrizes diferentes podem possuir os mesmos autovalores, mas com autovetores diferentes.

A decomposição espectral de uma matriz de covariância ou correlação dá origem a um produto de três matrizes: uma primeira matriz de autovetores, uma segunda matriz diagonal de autovalores e uma terceira matriz transposta dos autovetores. Enquanto os autovalores expressam a variabilidade ou importância relativa dos eixos, os autovetores exprimem o sentido ou a direção dos eixos principais do conjunto de dados.

A classificação matricial em relação ao sinal de seus elementos se dá com a obtenção de seu determinante e de seus autovalores. Frisa-se que cada elemento singularmente não tem tanta importância, pois sua análise não depende do valor específico de cada um, mas apenas dos autovalores e do determinante, tal como demonstrado abaixo:

- Matriz positiva definida: seu determinante é positivo ($|A| > 0$) e os seus autovalores também são positivos ($\lambda_j > 0$);
- Matriz positiva semidefinida: seu determinante é positivo ($|A| > 0$) e os seus autovalores são maiores ou iguais a 0 ($\lambda_j \geq 0$);
- Matriz negativa definida: seu determinante é negativo ($|A| < 0$) e os seus autovalores também são negativos ($\lambda_j < 0$);
- Matriz negativa semidefinida: seu determinante é negativo ($|A| < 0$) e os seus autovalores são menores ou iguais a 0 ($\lambda_j \leq 0$);
- Matriz indefinida: seus autovalores (λ_j) possuem sinais contrários.

Um método muito eficaz para gerar dados aleatórios de uma distribuição multivariada preservando a estrutura de correlação original dos dados é a decomposição de Cholesky. Ela pode ser aplicada em qualquer matriz simétrica positiva definida, ideal no estudo em questão, pois com o uso de tal técnica, a matriz original é reescrita em termos de uma nova matriz que é triangular superior.

Para se encontrar o qui-quadrado é necessário realizar a soma quadrática de variáveis normais padronizadas, com média 0 e desvio padrão 1. Esta distribuição de probabilidade contínua possui grande aplicabilidade na construção da elipse de confiança pelo fato que os semieixos possuem proporcionalidade com o qui-quadrado, dada pela equação (2.25).

$$\mathbf{x}^T \Sigma \mathbf{x} = [x_1 \quad x_2] \begin{bmatrix} \frac{1}{\sigma_{x_1}^2} & 0 \\ 0 & \frac{1}{\sigma_{x_2}^2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{x_1^2}{\sigma_{x_1}^2} + \frac{x_2^2}{\sigma_{x_2}^2} = \left(\frac{x_1}{\sigma_{x_1}}\right)^2 + \left(\frac{x_2}{\sigma_{x_2}}\right)^2 \rightarrow c^2 = z_1^2 + z_2^2 \quad (2.25)$$

O componente principal é dado pela direção que explica uma maior variação quando se trata de um conjunto de dados multivariados. Eles são obtidos a partir da combinação linear de atributos originais que buscam encontrar a maior parte de variação das informações. Na análise

dos componentes, é descoberto que uma parte dos componentes carregam o sinal e outra parte o ruído.

3. Construção da elipse de confiança

O presente trabalho busca contribuir para o desenvolvimento de uma sequência de criação para a elipse de confiança usando-se o software Microsoft Excel[®] para que o usuário possa fazer a alteração do nível de confiança, das variâncias, das médias, das correlações e dos pontos e visualize com clareza e rapidez a elipse sendo alterada. Para isso, apresenta-se uma elipse de confiança usada como apoio à análise de precisão na medição de concentração de emulsões de água em óleo (W/O) com base na pesquisa feita por Paiva (2004). Sua construção, apresentada no Capítulo 2, sugere a seguinte sequência:

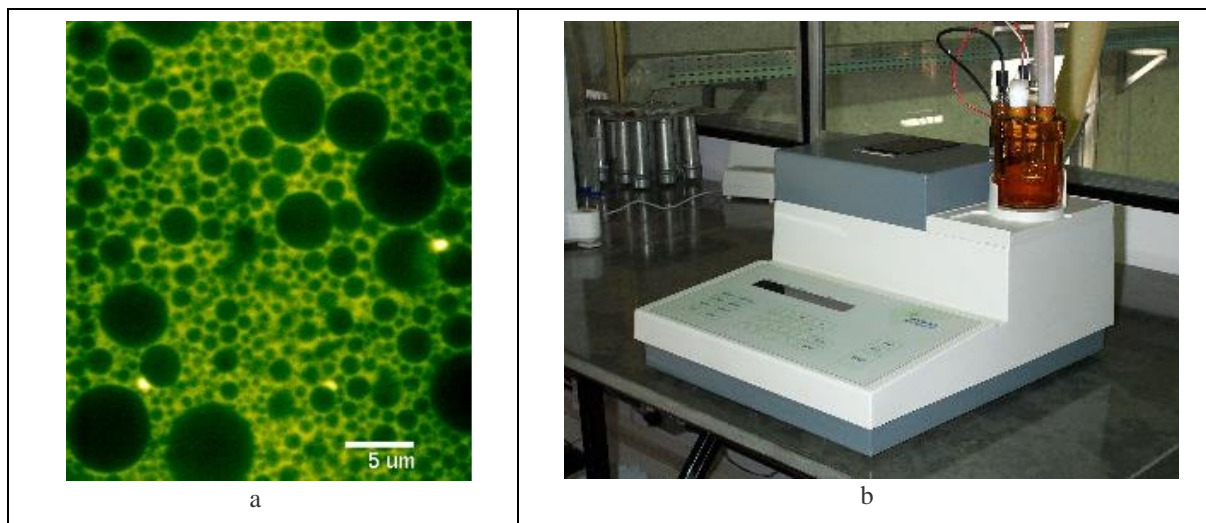
- a) Coleta dos dados;
- b) Estimação dos parâmetros;
- c) Cálculo da matriz de variância-covariância;
- d) Cálculo dos autovalores e autovetores;
- e) Determinação dos eixos e semieixos da elipse;
- f) Definição do nível de confiança;
- g) Determinação do raio; e
- h) Desenho da elipse.

Foi aplicada a Metodologia de Superfície de Resposta na determinação das condições ótimas de homogeneização de emulsões de petróleo do tipo W/O (*Water in Oil*), aspecto crucial na garantia da qualidade de uma indústria de petróleo.

Uma boa homogeneização pressupõe a minimização do coeficiente de variação do diâmetro característico de gota. Minimizar esse coeficiente possibilita tornar a emulsão mais homogênea e evitar excessiva trituração das gotas. Estas, usualmente excedem 0,5 μm de diâmetro e tornam-se visíveis sob microscópio óptico, conforme Figura 1a.

Foram considerados Temperatura, Velocidade de Rotação e Tempo como fatores de influência na homogeneização de emulsões, investigados no equipamento de titulação Karl-Fisher, conforme demonstrado na Figura 1b. A média (\bar{Y}) do diâmetro característico de gota $D(V; 0,5)$ e o coeficiente de variação do tamanho de gota ($C.V.\%$), altamente correlacionados, foram escolhidos para compor o arranjo para a geração da elipse de confiança.

Figura 1 – Exemplo de emulsão W/O (a) e equipamento de titulação Karl-Fischer Cólumétrico (b)



Fonte: Adaptado de Paiva (2004)

Adotou-se um arranjo fatorial completo com esses fatores e respectivos níveis de ensaio, com 21 experimentos, 5 pontos centrais e 2 réplicas, conforme Tabela 1. Como resposta, adotou-se o coeficiente de variação do diâmetro característico de gota D (V; 0,5), medido a partir do analisador de partículas MALVERNTM (PAIVA et al, 2013).

Tabela 1 – Fatorial para o processo experimental de homogeneização

Experimento	Média (Y)	Coeficiente de Variação (C.V.%)
1	26,423	0,080
2	2,850	0,035
3	32,247	0,093
4	3,813	0,030
5	7,605	0,062
6	2,061	0,029
7	14,708	0,064
8	3,440	0,033
9	21,733	0,071
10	2,483	0,043
11	26,706	0,101
12	3,482	0,025
13	7,193	0,059
14	2,069	0,029

15	18,454	0,062
16	2,937	0,025
17	4,498	0,037
18	3,639	0,035
19	3,259	0,034
20	3,520	0,036
21	3,674	0,031

Fonte: Paiva et al. (2004)

Para entendimento do arranjo, quanto menor for o C.V.%, mais homogênea será a emulsão.

A matriz de variância-covariância se deu a partir dos dados da Tabela 1 e permitiu calcular os parâmetros da elipse de confiança da Tabela 2. Para Paiva et al. (2009), a análise da variância-covariância contribui significativamente para a otimização dos resultados, facilita a compreensão dos dados e permite uma melhor tomada de decisão.

Tabela 2 – Dados para a construção da elipse de confiança

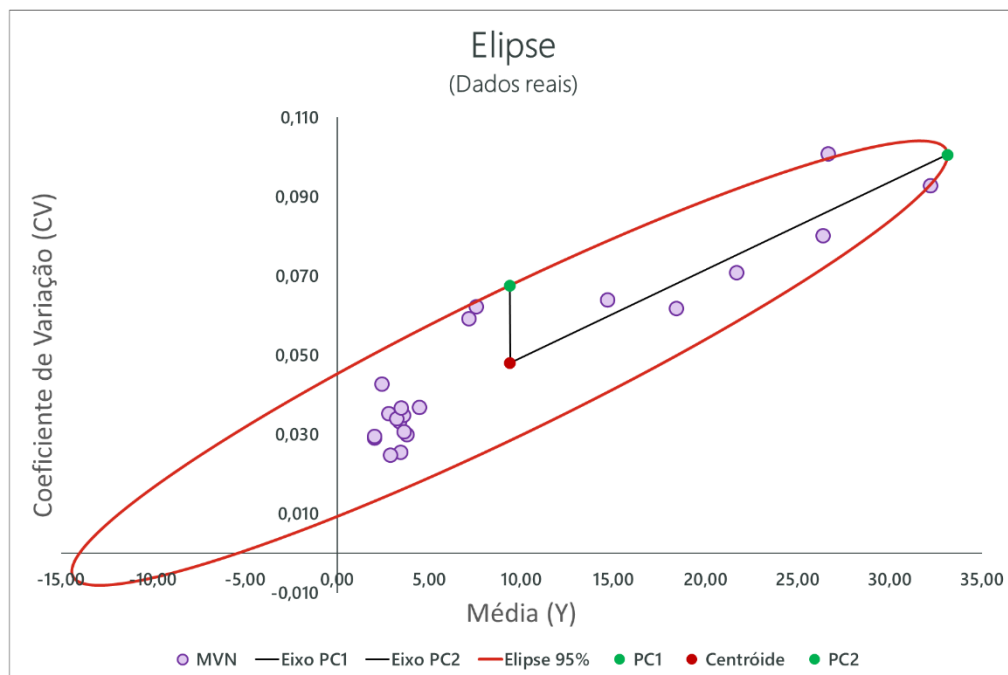
Σ	94,288	0,209	P	1,000	-0,002	μ_1^T	μ_2^T	Pontos	x	y
	0,209	0,001		0,002	1,000	9,371	0,048	PC ₁	23,768	0,053
								PC ₁ + μ	33,139	0,101
p	2,000		Λ	94,289	0,000	χ^2	5,991	PC ₂	0,000	0,020
ρ	0,937			0,000	0,000			PC ₂ + μ	9,371	0,068
θ	0,002					Raios				
x	5,000%		L^T	9,710	0,021	Externo	23,768	$(\chi^2 \Lambda)^{1/2}$	23,768	0,000
				0,000	0,008	Interno	0,020		0,000	0,020

Fonte: Elaborada pelos autores (2023)

Obtidos os resultados necessários contidos no passo a passo apresentado anteriormente, tornou-se possível a construção de duas elipses.

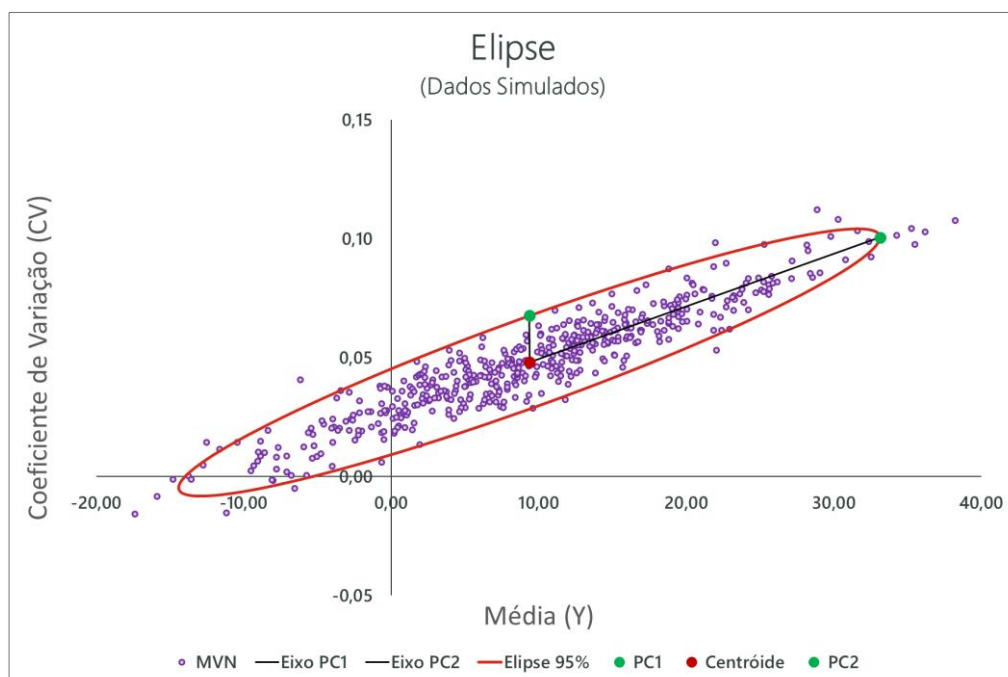
A Figura 2 apresenta a elipse relação de dispersão dos dados experimentais reais entre a média do diâmetro característico das gotas do experimento e o seu coeficiente de variação. Já a Figura 3 apresenta a elipse elaborada com dados simulados, utilizando-se a estrutura matricial dada pelo arranjo real dos dados.

Figura 2 – Elipse de confiança com os dados experimentais reais



Fonte: Elaborada pelos autores (2023).

Figura 3 – Elipse de confiança com os dados simulados



Fonte: Elaborada pelos autores (2023).

4. Conclusões

A utilização de elipses de confiança é de grande relevância na estatística multivariada. Fornecem uma robusta e fácil representação gráfica da variabilidade e das incertezas associadas

aos parâmetros de determinado experimento ou de complexos conjuntos de dados. Este trabalho apresentou breve conceito da elipse de confiança, seu equacionamento e sua aplicação como apoio à análise da precisão na medição de concentração de emulsões de água em óleo (W/O). A aplicação utilizando os dados de Paiva (2004) demonstrou que a relação média (\bar{Y}) e Coeficiente de Variação (C.V.%) observada na elipse da Figura 2 apresenta boa parte dos pontos próximos ao mínimo dos eixos, com certa dispersão que faz com que essas duas variáveis tenham comportamento semelhante, denotando emulsões mais homogêneas quanto mais próximas estão ao ponto 0 dos eixos. Esse mesmo comportamento é melhor observado na Figura 3, onde a nuvem de dados aleatórios gerada com os mesmos parâmetros do arranjo real enfatiza essa alta correlação e esparsividade dos dados.

Portanto, é possível através da abordagem de elipses de confiança identificar padrões, outliers e outros dados que auxiliam a tomada de decisão com maior facilidade. Esta característica demonstra como se faz importante e desejável seu uso para tornar mais assertivas e confiáveis as tomadas de decisão nos mais diversos processos produtivos.

REFERÊNCIAS

MAHALANOBIS, P. On the Generalized Distance in Statistics. **National Institute of Sciences of India**, v. 2, n. 1, p. 49–55, 1936.

MONTGOMERY, Douglas C.; RUNGER, George C. **Applied statistics and probability for engineers**. John Wiley & Sons, 2010.

PAIVA, A. P. de. **Estudo da minimização de erro nas medições de concentração de emulsões por titulação Karl-Fisher utilizando-se projeto de experimentos**. Orientador: Pedro Paulo Balestrassi. 2004. Dissertação (Mestrado) – Curso do Programa de Pós-Graduação em Engenharia de Produção, Instituto de Engenharia de Produção e Gestão, Universidade de Itajubá, Itajubá, 2004. Disponível em: <https://repositorio.unifei.edu.br/jspui/handle/123456789/3418>. Acesso em: 20/05/2023.

PAIVA, A. P.; Gomes, J.H.F.; Balestrassi, P.P.; PAPANDREA, P. J.; MOREIRA, N. C. **Estudo da minimização de erro nas medições de concentração de emulsões por titulação Karl-Fischer utilizando-se projeto de experimentos**. Revista Científica e-Locução, v. 04, p. 79-95, 2013. DOI: <https://doi.org/10.57209/e-locucao.v1i04.73>.

SAPORTA, Gilbert. **Probabilités, analyse des données et statistique**. Editions Technip, 2006.

SAPORTA, Gilbert; HATABIAN, Gérard. **Régions de confiance en analyse factorielle**. In: Fourth International Symposium on Data Analysis and Informatics. Elsevier Science Publishers, 1986. p. 499-508.

SCHELP, C. (2018) An Alternative Way to Plot the Covariance Ellipse. GitHub Pages.
https://carstenschelp.github.io/2018/09/14/Plot_Confidence_Ellipse_001.html. Acesso em: 23/05/2023.

SILVA, A. B.; SANTOS, C. D. (2019). Decomposição de Cholesky: Um estudo sobre o método e suas aplicações. **Revista Brasileira de Matemática Aplicada**, 41(2), 327-342.

SILVA, A. B.; SANTOS, C. D. (2020). Aplicação do conceito da região de confiança para a otimização de processos produtivos. **Revista Brasileira de Engenharia de Produção**, 2(1), 50-61.