

Planejamento e Experimento Estatístico sobre Dimensão da Moagem Utilizando Análise Full Factorial 2^3 e Modelagem por Método de Regressão Linear Múltipla

Décio Miranda Filho - ra: 236087

13 de junho de 2023

1 Introdução

O Processo de moagem CNC (Computer Numerical Control) é um processo de usinagem popular na indústria moderna devido à sua capacidade de remover materiais com uma ferramenta de corte de vários pontos em uma taxa mais rápida e com uma qualidade de superfície razoavelmente boa. Objetivando obter a granularidade superficial desejada, é necessário escolher os parâmetros de controle de maneira apropriada. Nessa moagem CNC, existem múltiplos parâmetros que controlam a qualidade de tal superfície. A análise da rugosidade superficial no processo de moagem CNC é um processo desafiador para a área de pesquisa. Muitos fatores envolvidos no processo de usinagem precisam ser otimizados para obter a melhor qualidade superficial possível. Neste estudo, três parâmetros de usinagem são considerados: "namely spindle", "speed feed rate" e "depth of cut". Além disso a variável resposta será a chamada "D" que representa o efeito sobre a dimensão da moagem. Este trabalho contempla um delineamento estatístico feito em "full factorial design"²³ e aplicação a um modelo de regressão linear múltipla a fim de entender o efeito dos parâmetros citados sobre a variável resposta.

2 Coleta dos Dados e Análises Estatísticas

2.1 Informações sobre o Experimento e Planejamento

O tipo de delineamento utilizado na tabela de "delineamento fatorial completo de experimentos". Nesse tipo de delineamento, todas as combinações possíveis dos fatores de interesse são testadas. Cada um desses fatores foi testado em dois níveis, resultando em um modelo fatorial 2^3 . No caso desse estudo, os fatores de interesse são a velocidade do eixo, a taxa de avanço e a profundidade de corte. Para cada fator, diferentes níveis ou valores são selecionados, e todas as combinações possíveis desses níveis são testadas. Isso resulta em um total de 125 experimentos, em que cada experimento representa uma combinação específica de valores para os fatores. O objetivo desse tipo de delineamento é obter informações detalhadas sobre o efeito de cada fator, bem como possíveis interações entre os fatores. Ao testar todas as combinações possíveis, é possível analisar o impacto individual de cada fator e como eles interagem entre si. Os dados obtidos desses experimentos são então analisados estatisticamente para identificar padrões, relações e tendências.

Portanto, o delineamento fatorial completo de experimentos permite uma investigação abrangente e sistemática dos fatores de interesse, fornecendo informações valiosas para o desenvolvimento do modelo preditivo e para o entendimento das relações entre os parâmetros de corte e a dimensão fractal obtida nas operações de fresagem CNC.

Vale ressaltar que os dados foram obtidos a partir de Sahoo et al. (2012).

2.2 Análise Descritiva e Exploratória dos Dados

Como citado anteriormente, o dataset do experimento possui um total de 4 variáveis principais e 1 variável índice: (*Depth_of_cut*, *Spindle_speed* e *Feed_rate*) e a variável dependente (*D1*). Enquanto as 3 primeiras são todas numéricas inteiras, a última é numérica contínua. Conforme a figura abaixo, pode-se notar a distribuição da variável. Com média de 1.37944, mediana 1.38 e desvio padrão 0.03258498.

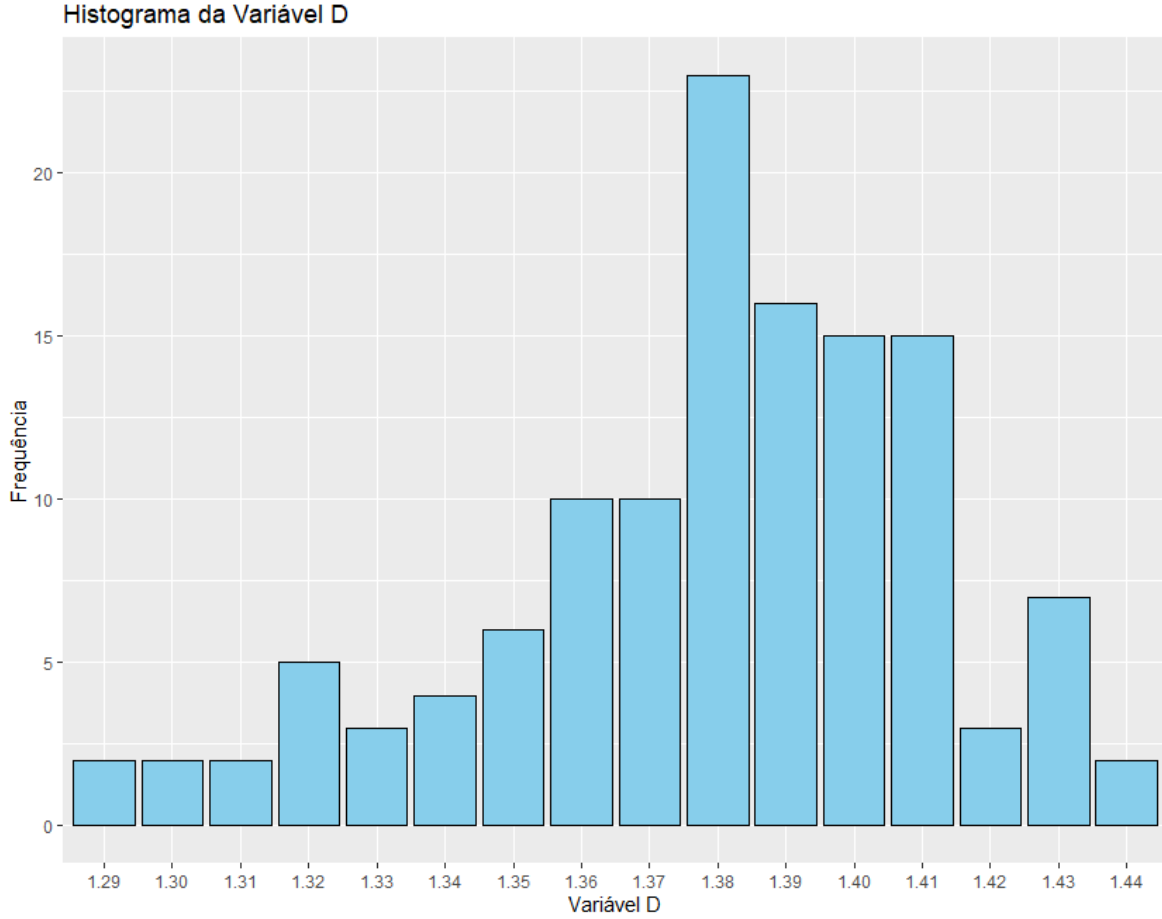


Figura 1: Distribuição da Variável D feita em R.

3 Modelagem

3.1 Descrição da Técnica

A técnica de modelagem utilizada neste estudo foi a regressão linear Múltipla. O objetivo foi investigar a relação entre as variáveis independentes (e a variável dependente (D1) no processo de fresamento CNC. A regressão linear é uma técnica estatística que permite estimar a relação entre uma variável dependente e uma ou mais variáveis independentes.

Para ajustar o modelo de regressão linear, foram utilizados os dados experimentais obtidos a partir do delineamento fatorial completo 2^k realizado. O delineamento fatorial completo permite explorar o efeito das variáveis independentes em diferentes níveis ou combinações. Neste caso, foram considerados os fatores, cada um com níveis entre (-1,+1). Assim, foram realizados experimentos em que esses fatores foram variados e os valores da variável dependente D1 foram registrados. A análise de regressão linear foi realizada para estimar os coeficientes de regressão, que representam o efeito médio de cada variável independente em D1, mantendo as outras variáveis constantes. Além disso, foram calculados os valores p para avaliar a significância estatística dos coeficientes estimados. A estimação foi feita por Mínimos Quadrados, cuja equação

$$D1 = \hat{\beta}_0 + \hat{\beta}_1 \cdot \text{Depth_of_cut} + \hat{\beta}_2 \cdot \text{Spindle_speed} + \hat{\beta}_3 \cdot \text{Feed_rate} \quad (1)$$

3.2 Avaliação dos Resultados do Modelo e Inferência do Modelo

A análise de regressão linear foi realizada utilizando um delineamento fatorial completo 2^k para investigar a relação entre as variáveis independentes (*Depth_of_cut*, *Spindle_speed* e *Feed_rate*) e a variável

dependente ($D1$) no processo de fresamento CNC. Os resultados estatísticos obtidos fornecem informações valiosas sobre o efeito das variáveis independentes em $D1$, bem como a qualidade do ajuste do modelo.

Os coeficientes estimados são uma medida importante do efeito médio de cada variável independente em $D1$, mantendo todas as outras variáveis constantes. O coeficiente estimado para $Depth_of_cut -1$ foi de -0.027600 com um erro padrão de 0.007669. Isso indica que um aumento de uma unidade em $Depth_of_cut -1$ está associado a uma diminuição média de 0.027600 em $D1$, considerando todas as outras variáveis constantes. De maneira similar, os coeficientes estimados para as outras variáveis independentes ($Depth_of_cut0$, $Depth_of_cut0.5$, $Depth_of_cut1$, $Spindle_speed-1$, $Spindle_speed0$, $Spindle_speed0.5$, $Spindle_speed1$, $Feed_rate-1$, $Feed_rate0$, $Feed_rate0.5$ e $Feed_rate1$) fornecem informações sobre seus respectivos efeitos médios em $D1$.

Os valores p são fundamentais para avaliar a significância estatística dos coeficientes estimados. No presente estudo, encontramos coeficientes com valores p significativos, como $Depth_of_cut -1$ ($p = 0.000477$), $Depth_of_cut0$ ($p = 0.034654$) e $Spindle_speed1$ ($p = 0.000276$). Esses valores p inferiores a 0.05 indicam que essas variáveis têm efeitos estatisticamente significativos em $D1$, considerando um nível de significância de 5%. Por outro lado, variáveis como $Depth_of_cut0.5$, $Depth_of_cut1$, $Spindle_speed-1$, $Spindle_speed0$, $Spindle_speed0.5$, $Feed_rate-1$, $Feed_rate0$, $Feed_rate0.5$ e $Feed_rate1$ não apresentaram significância estatística com base nos valores p .

O R-quadrado ajustado é uma medida da proporção da variabilidade total de $D1$ explicada pelo modelo de regressão linear. No presente estudo, o valor de R-quadrado ajustado foi de 0.3076, o que indica que aproximadamente 30.76% da variabilidade em $D1$ pode ser explicada pelas variáveis independentes incluídas no modelo considerando também que quanto mais próximo o valor do R-quadrado ajustado estiver de 1, melhor será o ajuste do modelo aos dados.

A estatística F é utilizada para testar a hipótese nula de que todos os coeficientes de regressão são iguais a zero, o que implicaria que as variáveis independentes não têm efeito significativo em $D1$. No nosso caso, a estatística F foi de 5.59, com um valor p extremamente baixo ($1.967e-07$), indicando que o modelo como um todo é estatisticamente significativo. Isso implica que pelo menos uma das variáveis independentes tem um efeito significativo em $D1$.

Tabela 1: Coefficients

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.369120	0.008744	156.575	< 2e-16 ***
Depth_of_cut-1	-0.027600	0.007669	-3.599	0.000477 ***
Depth_of_cut0	0.016400	0.007669	2.138	0.034654 *
Depth_of_cut0.5	0.012000	0.007669	1.565	0.120472
Depth_of_cut1	0.000400	0.007669	0.052	0.958497
Spindle_speed-1	0.004400	0.007669	0.574	0.567303
Spindle_speed0	-0.000800	0.007669	-0.104	0.917107
Spindle_speed0.5	0.008800	0.007669	1.147	0.253640
Spindle_speed1	0.028800	0.007669	3.755	0.000276 ***
Feed_rate-1	-0.003600	0.007669	-0.469	0.639687
Feed_rate0	-0.006000	0.007669	-0.782	0.435657
Feed_rate0.5	0.010400	0.007669	1.356	0.177801
Feed_rate1	0.008400	0.007669	1.095	0.275737

Sobre a Análise dos Resíduos foi feito um teste KS e obtido um valor-p de 0.773, portanto não havendo evidências suficientes para rejeitar a Hipótese nula de que os resíduos possuem distribuição Normal. Segue abaixo o QQ-PLOT corroborando tal resultado:

Em resumo, os resultados estatísticos do modelo de regressão linear baseado no delineamento fatorial indicam que $Depth_of_cut -1$, $Depth_of_cut0$ e $Spindle_speed1$ têm efeitos estatisticamente significativos em $D1$. Essas variáveis independentes contribuem para explicar aproximadamente 30.76% da variabilidade observada em $D1$.

Spindle_speed: O fator $Spindle_speed$ também apresenta um valor p significativo (0.0009671), indicando que ele tem um efeito estatisticamente significativo em $D1$. O valor de F (4.9935) indica que a variabilidade explicada pelo fator é maior do que a variabilidade não explicada, embora em menor grau do que o fator $Depth_of_cut$.

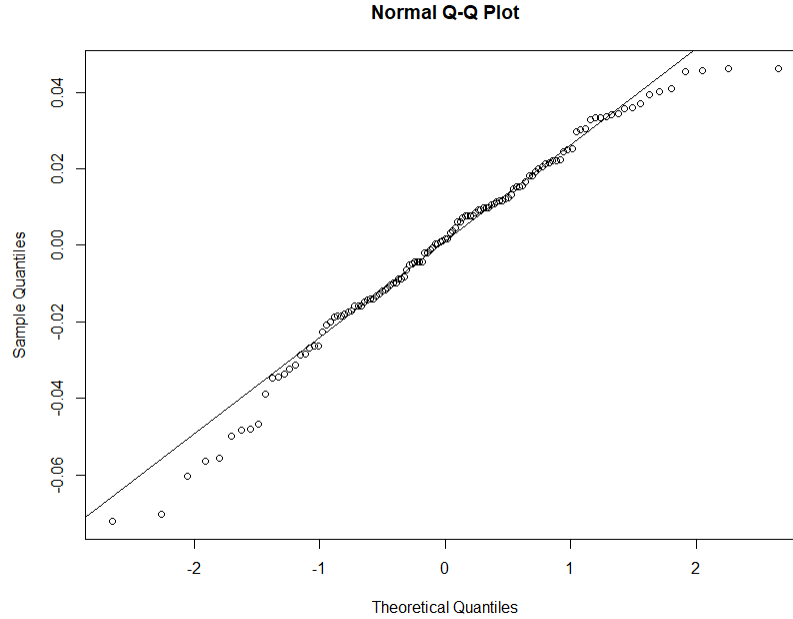


Figura 2: QQ Plot

Feed_rate: O fator *Feed_rate* não apresenta um valor p significativo (0.1354678), indicando que não há evidências estatísticas suficientes para afirmar que esse fator tem um efeito significativo em *D1*. O valor de *F* (1.7916) indica que a variabilidade explicada pelo fator é comparável à variabilidade não explicada.

Essas análises permitem concluir que os fatores *Depth_of_cut* e *Spindle_speed* têm efeitos estatisticamente significativos na variável de resposta *D1*. No entanto, o fator *Feed_rate* não apresenta evidências estatísticas suficientes para afirmar seu efeito significativo em *D1*.

Depth_of_cut: O fator *Depth_of_cut* apresenta um valor p extremamente baixo (6.01×10^{-7}), indicando que há evidências estatísticas significativas de que esse fator tem um efeito significativo na variável de resposta *D1*. O valor de *F* (9.9853) também sugere que a variabilidade explicada pelo fator é maior do que a variabilidade não explicada.

Para mais detalhes a tabela abaixo:

Tabela 2: Tabela de Análise de Variância					
Variável	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Depth_of_cut	4	0.029365	0.0073412	9.9853	6.01×10^{-7} ***
Spindle_speed	4	0.014685	0.0036712	4.9935	0.0009671 ***
Feed_rate	4	0.005269	0.0013172	1.7916	0.1354678
Residuals	112	0.082342	0.0007352		

4 Críticas, Vantagens e Desvantagens

O delineamento fatorial completo utilizado neste estudo apresenta vantagens e desvantagens em relação a outros planejamentos estatísticos. Uma das principais vantagens é que ele permite investigar os efeitos de múltiplas variáveis independentes e suas interações em uma única análise, o que é especialmente útil quando se deseja entender como diferentes fatores influenciam uma resposta. Além disso, o delineamento fatorial completo é eficiente em termos de custo, pois requer um número relativamente pequeno de experimentos em comparação com outros delineamentos.

No entanto, uma desvantagem do delineamento fatorial completo é que ele pode exigir um número significativo de experimentos quando o número de variáveis independentes e níveis é grande, o que

pode ser trabalhoso e dispendioso. Além disso, a interpretação dos resultados pode se tornar complexa devido às possíveis interações entre as variáveis independentes.

No contexto da modelagem com regressão múltipla, as vantagens incluem a capacidade de ajustar um modelo que leve em consideração múltiplas variáveis independentes e suas relações lineares com a variável dependente. A regressão múltipla permite quantificar o efeito individual de cada variável independente, controlando o efeito das outras variáveis presentes no modelo. Além disso, a regressão múltipla fornece informações sobre a significância estatística das variáveis independentes e a qualidade global do modelo por meio do coeficiente de determinação (R^2) e do teste F.

Porém, é importante considerar as limitações da regressão múltipla. Uma limitação é a suposição de que a relação entre as variáveis é linear. Se essa suposição não for válida, os resultados da regressão podem ser distorcidos. Além disso, a regressão múltipla pode ser sensível a outliers, multicolinearidade e violações de outras suposições, o que pode comprometer a interpretação e a precisão dos resultados.

Para melhorar o planejamento estatístico e a modelagem com regressão múltipla, algumas sugestões podem ser consideradas. Uma delas é explorar outras técnicas de amostragem e delineamento, como o delineamento fatorial fracionado, que permite reduzir o número de experimentos sem perder informações importantes. Além disso, é importante avaliar com cautela as suposições da regressão múltipla e, se necessário, aplicar transformações nas variáveis para atender a essas suposições. A detecção e tratamento de outliers também são etapas importantes na melhoria da precisão dos resultados.

Outra sugestão é realizar uma validação cruzada do modelo para avaliar sua capacidade de generalização para novos dados. Isso ajuda a verificar se o modelo é robusto e se pode ser aplicado a outras amostras.

5 Conclusão

No contexto deste estudo, a modelagem por regressão linear permitiu investigar a relação entre as variáveis independentes e a variável dependente no processo de fresamento CNC. Os coeficientes estimados e os valores p forneceram informações sobre o efeito e a significância estatística de cada variável independente em relação a D1. Além disso, o ajuste do modelo foi avaliado por meio do coeficiente de determinação e da estatística F. Essas análises contribuíram para uma melhor compreensão do processo de moagem CNC e forneceram informações valiosas para a otimização do mesmo.

Consoante, o delineamento fatorial completo e a modelagem com regressão múltipla oferecem benefícios na investigação dos efeitos de múltiplas variáveis independentes em uma resposta. No entanto, é importante considerar as limitações e explorar técnicas alternativas, além de realizar uma análise cuidadosa das suposições e uma validação adequada do modelo para melhorar a precisão e a generalização dos resultados.

6 Referências

Box, G.E.P., Hunter, J. S., Hunter, W.G. (2005). *Statistics for Experimenters: Design, Innovation, and Discovery*. 2ª Edição. John Wiley Sons, Nova Iorque.

Dean, A.M. and Voss, D. (1999). *Design and Analysis of Experiments*. Springer Verlag, Nova Iorque.

Montgomery, D.C., *Design and analysis of experiments*, John Wiley & Sons, New York, 2ª ed., 537p., 1984.

Neter, J., Wasserman, W. Kutner, M. H. *Applied Linear Statistical Models*. Terceira Edição. Irwin, Boston, 1992.

SAHOO, Prasanta; BARMAN, Tapan Kr. ANN modelling of fractal dimension in machining. In: DAVIM, J. Paulo (Ed.). *Woodhead Publishing Reviews: Mechanical Engineering Series. Mechatronics and Manufacturing Engineering*. Woodhead Publishing, 2012. p. 159-226. ISBN 9780857091505. Disponível em: <https://www.sciencedirect.com/science/article/pii/B9780857091505500054>

SAHOO, P., BARMAN, T.K. and Routara, B.C. (2008) Fractal dimension modeling of surface profile

le and optimization in CNC end milling using response surface method, International Journal of Manufacturing Research, 3(3): 360–77