

AVALIAÇÃO DA INFLUÊNCIA DOS MODELOS LINEARES GENERALIZADOS NA QUALIDADE DE PREVISÃO DE RESPOSTAS COM RESÍDUOS HETEROCEDÁSTICOS

Mirelli de Castro Cesário – mirellicesario12@gmail.com
Anderson Paulo de Paiva – andersonppaiva@unifei.edu.br

Palavras-chave: Otimização multiobjetivo, Modelos lineares generalizados, qualidade preditiva.

1. INTRODUÇÃO

Os Modelos Lineares Generalizados (*Generalized Linear Model* - GLM) representam uma classe de modelos estatísticos que ampliam a abordagem clássica dos Modelos Lineares Ordinários (*Ordinary Least Squares* - OLS). Introduzidos por Nelder e Wedderburn (1972), permitem a modelagem de variáveis que seguem diferentes distribuições de probabilidade. Antes de seu desenvolvimento, os modelos lineares eram restritos a situações em que a variável dependente atendia a premissas rigorosas, como normalidade dos dados e homocedasticidade dos resíduos. No entanto, a normalidade e a homogeneidade da variância constituem casos específicos, enquanto, na prática, erros não normais e variâncias não constantes ocorrem com frequência (Myers; Montgomery, 1997). Em processos industriais, essas violações são comuns e resultam em modelos com baixa capacidade preditiva e interpretação limitada.

Para superar essas limitações, o GLM configura-se como uma alternativa mais flexível, especialmente adequada para variáveis pseudocontínuas, como a vida útil de ferramentas de corte e as dimensões de peças usinadas, que frequentemente apresentam distribuições assimétricas e resíduos heterocedásticos.

Neste trabalho, a modelagem é conduzida tanto por OLS quanto por GLM, sendo fundamental, após a construção dos modelos, avaliar sua adequação aos dados experimentais. Para isso, utilizam-se métricas como o coeficiente de determinação (R^2), os intervalos de confiança das previsões (IC) e o erro padrão de

previsão (EPP). O R^2 indica a proporção da variabilidade explicada pelo modelo, constituindo um indicador clássico da qualidade do ajuste. Os intervalos de confiança expressam a incerteza associada às estimativas, permitindo avaliar a precisão das previsões em diferentes regiões do espaço experimental. Já o EPP quantifica a discrepância esperada entre valores observados e previstos, sendo particularmente relevante para verificar a robustez do modelo diante de novos dados. A análise integrada dessas métricas assegura não apenas o ajuste aos dados experimentais, mas também a capacidade preditiva do modelo para aplicações futuras.

Uma vez verificada a adequação do modelo, podem-se empregar técnicas de otimização. Em processos industriais, a otimização multiobjetivo desempenha papel central ao equilibrar variáveis conflitantes como custo, produtividade e qualidade do produto (Shahsavari-Pour *et al.*, 2024). Nesse contexto, métodos como análise de componentes principais e análise fatorial contribuem para a redução da dimensionalidade dos dados, facilitando a interpretação dos fatores subjacentes e permitindo a construção de funções objetivo e restrições linearmente independentes (Almeida *et al.*, 2019; Melo *et al.*, 2022; Naves *et al.*, 2017; Toledo *et al.*, 2023).

Diante desse contexto, o objetivo deste trabalho é avaliar a capacidade preditiva dos modelos OLS e GLM na modelagem da vida da ferramenta em um processo de usinagem, utilizando modelos de superfície de resposta obtidos por meio de um arranjo composto central. Na modelagem via GLM, são empregados pesos para estabilizar a variância de previsão e reduzir a heterocedasticidade dos resíduos, aumentando a precisão e a confiabilidade das estimativas.

2. DESENVOLVIMENTO

2.1 Referencial teórico

Os Modelos Lineares Generalizados (GLM), introduzidos por Nelder e Wedderburn (1972), representam uma extensão dos modelos lineares clássicos ao permitirem a utilização de diferentes distribuições de probabilidade para a variável resposta. Diferentemente do modelo linear tradicional, que pressupõe normalidade dos resíduos e relação estritamente linear entre variáveis, os GLMs flexibilizam essas restrições por meio da introdução de funções de ligação e da possibilidade de

considerar distribuições como binomial e Poisson, além de distribuições contínuas assimétricas, como lognormal, gamma, exponencial e Weibull. Dessa forma, tornam-se adequados para uma ampla gama de problemas em que as premissas da regressão linear não são atendidas.

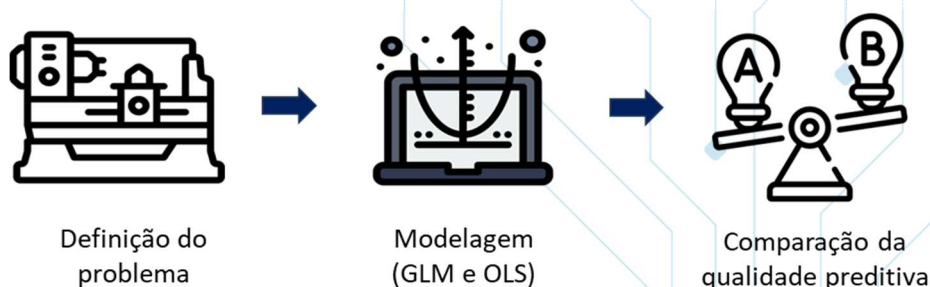
O uso de GLMs é especialmente indicado em situações em que a variável resposta representa contagens ou proporções, mas também se aplica a contextos associados à análise de confiabilidade, como funções de risco e tempo de falha (Myers *et al.*, 2010). Essa versatilidade faz com que os GLMs sejam amplamente utilizados em diferentes áreas do conhecimento, desde ciências biológicas até engenharia de processos, oferecendo uma estrutura estatística mais robusta para modelagem e previsão.

Diversos trabalhos têm explorado o uso de GLM em diferentes contextos. Li e Deng (2021) desenvolveram um algoritmo para projetos experimentais ótimos em GLM, enquanto Perumal *et al.* (2024) e Venkatesh *et al.* (2024) aplicaram GLM em estudos de usinagem, associando-o a métodos como ANOVA e RSM para otimização de parâmetros. Zubair *et al.* (2023) utilizaram GLM para avaliar revestimentos em insertos de metal duro, obtendo ganhos significativos em rugosidade e vida útil da ferramenta. Em áreas distintas, Hayat e Higgins (2014) destacaram a relevância do GLM para dados de contagem na enfermagem, e Navarro *et al.* (2017) aplicaram o modelo para identificar variáveis que influenciam o desgaste de ferramentas na usinagem de compósitos.

2.2 Metodologia

A condução deste trabalho segue o fluxograma da Figura 1, que organiza de forma estruturada todas as etapas, desde a definição do processo até a avaliação dos resultados preditivos.

Figura 1 - Fluxograma metodológico



A metodologia é organizada em três etapas principais. Na primeira, define-se o processo a ser analisado com base no experimento de Salgado Júnior (2010). Em seguida, realiza-se a modelagem do problema utilizando os métodos OLS e GLM, sendo o OLS empregado para obter os resíduos que servirão como pesos na modelagem por GLM. Por fim, calculam-se métricas como R^2 , R^2 ajustado, R^2 de previsão, intervalos de confiança e erro padrão de previsão, e os resultados são comparados para avaliar a qualidade preditiva dos modelos.

3. RESULTADOS E DISCUSSÃO

O experimento analisado consiste em um processo de torneamento AISI 52100 com as seguintes variáveis de entrada: velocidade de corte, taxa de avanço e profundidade de corte - para mais detalhes consultar Paiva, Ferreira e Balestrassi (2007). A variável de interesse nesse estudo é o tempo de vida da ferramenta, conforme ilustrado na Tabela 1.

Tabela 1 - Matrix experimental

Vc	f	ap	Vida
200,000	0,050	0,150	16,750
240,000	0,050	0,150	11,500
200,000	0,100	0,150	9,850
240,000	0,100	0,150	8,500
200,000	0,050	0,300	11,500
240,000	0,050	0,300	7,450
200,000	0,100	0,300	8,200
240,000	0,100	0,300	6,250
220,000	0,075	0,225	8,600
220,000	0,075	0,225	6,800
187,340	0,075	0,225	10,100
252,660	0,075	0,225	7,600
220,000	0,034	0,225	17,500
220,000	0,116	0,225	7,200
220,000	0,075	0,103	12,000
220,000	0,075	0,348	6,700
220,000	0,075	0,225	7,200
220,000	0,075	0,225	9,100

Fonte: Paiva, Ferreira e Balestrassi (2007)

Ao analisar os resultados apresentados na Tabela 2 é possível constatar que o modelo GLM apresenta menor desvio padrão e maiores valores para R^2 , R^2 ajustado e R^2 de previsão. Embora os valores de R^2 sejam elevados nos dois modelos analisados, nota-se que o GLM apresenta melhores resultados nas outras métricas avaliadas. Desse modo, ao obter R^2 ajustado superior ao modelo OLS, constata-se que o GLM está modelando os dados de forma mais eficiente, sem adicionar complexidade desnecessária. De maneira análoga, a obtenção de R^2 de previsão superior ao modelo OLS mostra que o modelo possui maior robustez preditiva, sendo mais confiável para inferências para novos dados.

Tabela 2 - Modelagem da vida da ferramenta

	OLS	GLM
Constante	7,968	7,914
V_c	-1,251	-1,286
f	-2,341	-2,215
a_p	-1,639	-1,662
$V_c * V_c$	0,234	0,225
$f * f$	1,547	1,479
$a_p * a_p$	0,422	0,530
$V_c * f$	0,750	0,640
$V_c * a_p$	0,075	0,083
$f * a_p$	0,675	0,649
S	1,257	1,256
R^2	92,94%	99,31%
R^2 -ajust	84,99%	98,53%
R^2 -prev	57,88%	93,20%

Para avaliar a eficácia do método proposto, foram elaborados dois gráficos: o primeiro representa o erro padrão de previsão dos valores ajustados (EPP) e o segundo mostra os intervalos de confiança associados aos pontos experimentais (ICs). Nas Figuras 2 e 3 é possível observar que o modelo GLM apresenta EPP consistentemente inferior ao do modelo OLS, evidenciando maior precisão nas estimativas, especialmente diante da heterocedasticidade presente nos dados. Além disso, os ICs obtidos com o GLM são visivelmente mais estreitos, refletindo uma menor amplitude e, conseqüentemente, maior confiança nas previsões.

Figura 2 - Erro padrão de previsão

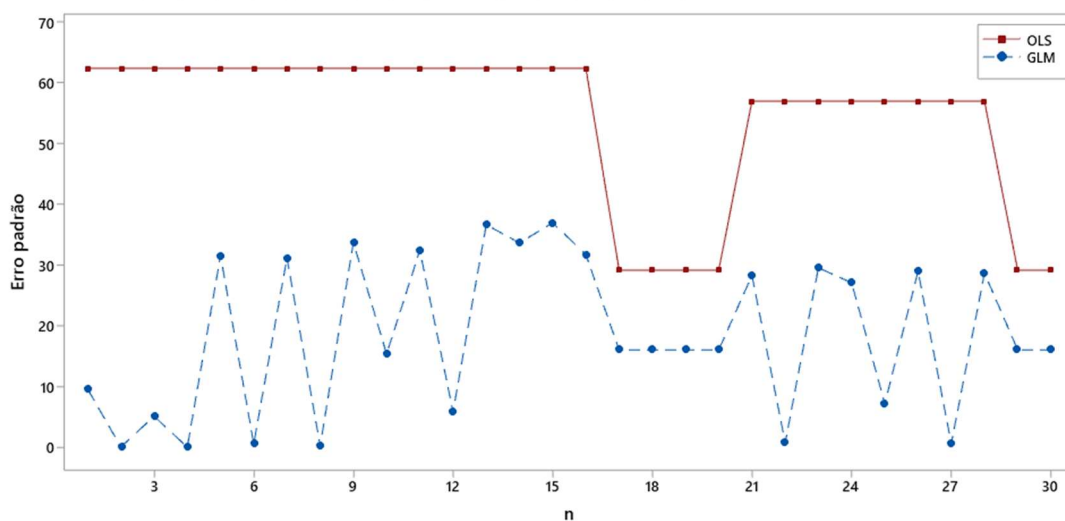
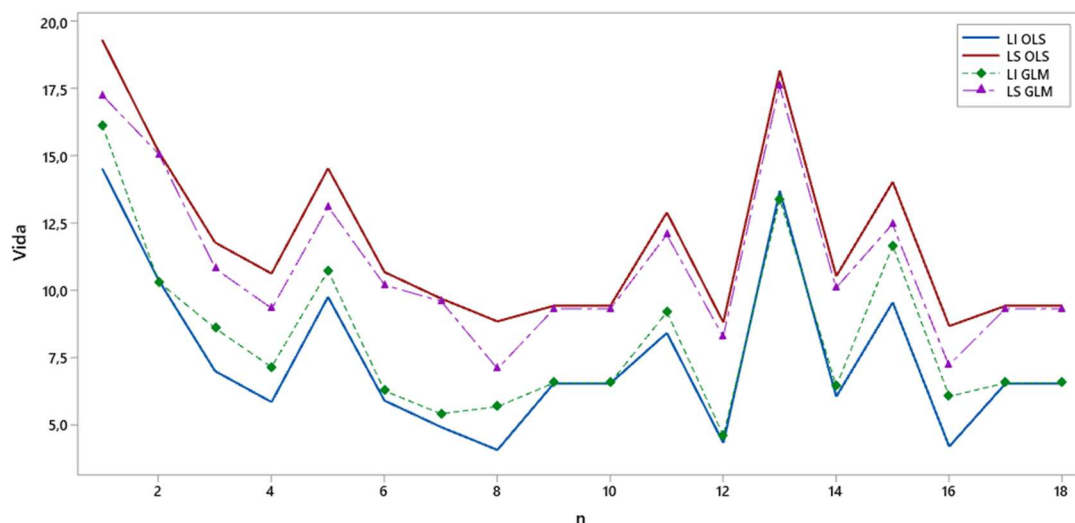


Figura 3 - Intervalo de confiança de previsão



4. CONSIDERAÇÕES FINAIS

Os resultados demonstram que, ao comparar o modelo tradicionalmente utilizado (OLS) com o GLM, este último apresenta desempenho superior em todas as métricas avaliadas, com destaque para o R^2 de previsão. Esse achado evidencia a capacidade do GLM de gerar previsões mais precisas para novos dados, não utilizados no processo de ajuste, ressaltando sua maior generalização.

No que se refere aos intervalos de confiança, observa-se redução na amplitude, sugerindo maior precisão nas estimativas. Contudo, não se identifica um padrão consistente para a média desses intervalos, o que impede concluir que o GLM desloque sistematicamente a média das previsões em relação ao OLS.

A análise do erro padrão de previsão indica que o GLM descreve com maior precisão o comportamento dos resíduos, corroborando a hipótese de heterocedasticidade no problema em estudo. Em termos absolutos, o erro padrão do GLM é inferior ao do OLS, sinalizando um ajuste mais eficiente dos dados.

Esses achados ressaltam a robustez do GLM como alternativa à modelagem tradicional baseada em mínimos quadrados ordinários, sobretudo em contextos em que as premissas do OLS não são plenamente atendidas. Assim, a pesquisa contribui para a modelagem de processos de fabricação em cenários complexos, ampliando o entendimento sobre o uso de GLM em problemas industriais.

AGRADECIMENTOS

Os autores agradecem à CAPES, à UNFEI e à FAPEMIG.

REFERÊNCIAS

ALMEIDA, F. A.; GOMES, G. F.; GAUDÊNCIO, J. H. D.; GOMES, J. H. F.; PAIVA, A. P. A new multivariate approach based on weighted factor scores and confidence ellipses to precision evaluation of textured fiber bobbins measurement system. **Precision Engineering**, [s. l.], vol. 60, p. 520–534, 2019. <https://doi.org/10.1016/j.precisioneng.2019.09.010>.

HAYAT, M. J.; HIGGINS, M. Understanding poisson regression. **Journal of Nursing Education**, [s. l.], vol. 53, no. 4, p. 207–215, 2014. <https://doi.org/10.3928/01484834-20140325-04>.

LI, Y.; DENG, X. An efficient algorithm for Elastic I-optimal design of generalized linear models. **Canadian Journal of Statistics**, [s. l.], vol. 49, no. 2, p. 438–470, 2021. <https://doi.org/10.1002/cjs.11571>.

MELO, S. A.; PEREIRA, R. B. D.; DA SILVA REIS, A. F.; LAURO, C. H.; BRANDÃO, L. C. Multi-objective evolutionary optimization of unsupervised latent variables of turning process. **Applied Soft Computing**, [s. l.], vol. 120, 2022. <https://doi.org/10.1016/j.asoc.2022.108713>.

MYERS, R. H.; MONTGOMERY, D. C.; VINING, G. G.; ROBINSON, T. J. **Generalized Linear Models: With Applications in Engineering and the Sciences**. 2nd ed. [S. l.]: Wiley, 2010.

MYERS, Raymond H.; MONTGOMERY, D. C. A tutorial on generalized linear models. **Journal of Quality Technology**, [s. l.], vol. 29, no. 3, p. 274–291, 1997. <https://doi.org/10.1080/00224065.1997.11979769>.

NAVARRO, M. D.; MESEGUER, M. D.; SÁNCHEZ, A. I.; GUTIÉRREZ, S. C. Tool wear study in edge trimming on basalt fibre reinforced plastics. **Procedia Manufacturing**, [s. l.], vol. 13, p. 259–266, 2017. <https://doi.org/10.1016/j.promfg.2017.09.067>.

NAVES, F. L.; DE PAULA, T. I.; BALESTRASSI, P. P.; MOREIRA BRAGA, W. L.; SAWHNEY, R. S.; DE PAIVA, A. P. Multivariate Normal Boundary Intersection based on rotated factor scores: A multiobjective optimization method for methyl orange treatment. **Journal of Cleaner Production**, [s. l.], vol. 143, p. 413–439, 2017. <https://doi.org/10.1016/j.jclepro.2016.12.092>.

NELDER, J.; WEDDERBURN, R. Generalized Linear Models. **Journal of Royal Statistical Society - Series B**, [s. l.], vol. 135, p. 370–384, 1972.

PAIVA, A. P.; FERREIRA, J. R.; BALESTRASSI, P. P. A multivariate hybrid approach applied to AISI 52100 hardened steel turning optimization. **Journal of Materials Processing**



IEPG SUMMIT'25

Pensando o futuro com inteligência
artificial e consciência social

Technology, [s. l.], vol. 189, no. 1–3, p. 26–35, 2007.
<https://doi.org/10.1016/j.jmatprotec.2006.12.047>.

PERUMAL, S.; AMARNATH, M. K.; MARIMUTHU, K.; SUBRAMANIAM, P.; RATHINAVELU, V.; JAGADEESH, D. Titanium carbonitride-coated CBN insert featured turning process parameter optimization during AA359 alloy machining. **International Journal of Advanced Manufacturing Technology**, [s. l.], 2024.
<https://doi.org/10.1007/s00170-024-13583-y>.

SHAHSAVARI-POUR, N.; HEYDARI, A.; FEKIH, A.; ASADI, H. A Novel Pareto-Optimal Algorithm for Flow Shop Scheduling Problem. **Mathematics**, [s. l.], vol. 12, no. 18, 2024.
<https://doi.org/10.3390/math12182951>.

TOLEDO, J. V. R. de; PIRES, T. F.; PEREIRA, R. B. D.; FERREIRA, J. R. A multi-objective robust evolutionary optimization approach applied to the multivariate helical milling process of super duplex steel[Formula presented]. **Applied Soft Computing**, [s. l.], vol. 147, 2023.
<https://doi.org/10.1016/j.asoc.2023.110811>.

VENKATESH, R.; LOGESH, K.; SINGH, S.; SINGH, P. K.; HOSSAIN, I.; MOHANAVEL, V.; SOUDAGAR, M. E. M.; ALHARBI, S. A.; AL OBAID, S. Aluminium alloy nanocomposite made with SiC via ultrasonic stir casting: Behaviour study. **Journal of Mechanical Science and Technology**, [s. l.], 2024. <https://doi.org/10.1007/s12206-024-0711-y>.

ZUBAIR, S. W. H.; ARAFAT, S. M.; KHAN, S. A.; UDDIN, G. M.; HAYAT, N. Dry finishing turning of AA7075 with binary and ternary nitrides and carbides ceramic-coated tools. **International Journal of Advanced Manufacturing Technology**, [s. l.], vol. 129, no. 1–2, p. 65–87, 2023. <https://doi.org/10.1007/s00170-023-12105-6>.