

Aplicação da Regressão Polinomial na Previsão da Participação Feminina em Bolsas de Produtividade STEM

Gabriel A. A. Bezerra¹, Clarice D. de Albuquerque²
UFCA, Juazeiro do Norte, CE

O direito à educação nem sempre foi dado às mulheres. Em tempos passados existia a crença que as mulheres eram inferiores aos homens e consequentemente acreditavam que elas não tinham espaço no mundo acadêmico.

Apesar do espaço das mulheres vem aumentando no mundo acadêmico, quando comparamos a presença delas nas áreas de STEM (sigla em inglês para Ciência, Tecnologia, Engenharia e Matemática), percebemos uma grande discrepância entre a quantidade de homens e mulheres. “No nosso país, apesar de 47% de todos os trabalhadores serem mulheres, apenas 24% trabalham nas carreiras de STEM” [1].

Este trabalho tem como objetivo fazer uma análise dessa diferença por meio do método de regressão polinomial e tentar prever futuros resultados, usando como base um banco de dados de bolsas de produtividade do CNPq.

As bolsas de produtividade do CNPq são concedidas a pesquisadores de destaque, com objetivo de valorizar sua produção científica. O banco de dados consiste em uma tabela, na qual cada linha representa um projeto desenvolvido por um pesquisador entre os anos de 2013 e 2023.

A Tabela 1 descreve as quantidades normalizadas de bolsas de produtividade feita por homens e mulheres.

Tabela 1: Proporção de bolsas de produtividade ao longo dos anos.

Sexo	2013	2015	2017	2019	2021	2023	Total
F	0.37698	0.386347	0.376705	0.386485	0.385207	0.400228	0.385923
M	0.62302	0.613653	0.623295	0.613515	0.614793	0.599772	0.614077

Na Tabela 2 é apresentado um recorte comparando com as bolsas de produtividade apenas nas áreas STEM.

Tabela 2: Proporção de bolsas de produtividade em STEM ao longo dos anos.

Sexo	2013	2015	2017	2019	2021	2023	Total
F	0.218211	0.225241	0.209895	0.226643	0.217064	0.239758	0.224522
M	0.781789	0.774759	0.790105	0.773357	0.782936	0.760242	0.775478

Pode-se perceber uma diferença de 16,10% a menos comparando a média da participação feminina nas áreas STEM.

A regressão linear [3] é um dos métodos mais utilizados em aprendizado de máquina para a obtenção de estimadores. Esse método busca determinar uma função linear que melhor se ajuste a um conjunto de pontos, possibilitando a modelagem da tendência observada.

¹gabriel.bezerra@aluno.ufca.edu.br

²clarice.albuquerque@ufca.edu.br

A regressão polinomial, [2], é um aprimoramento do método de regressão linear que obtém melhores resultados quando lidamos com conjuntos de dados não lineares, ou seja, quando os dados não seguem um padrão funcional aparente. Usando-a obtemos uma função polinomial que descreve melhor os dados deste trabalho.

Para implementar essa abordagem, usamos `PolynomialFeatures` para gerar o polinômio e `LinearRegression` para ajustar os dados. Primeiro, testamos a eficácia do modelo removendo os dados de 2023. O valor previsto para 2023 foi 0.22137, enquanto o valor real era 0.23976, resultando em um erro relativo de 7.67%.

Com base nessa validação preliminar, aplica-se agora a regressão polinomial ao conjunto completo de dados referentes às bolsas de produtividade para mulheres nas áreas STEM, estendendo a análise para os anos subsequentes e projetando possíveis tendências futuras.

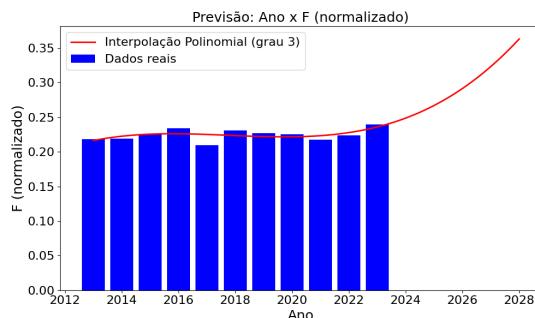


Figura 1: Gráficos da função adquirida por regressão polinomial. Fonte: Produzido pelo autor.

O polinômio escolhido foi de grau 3, pois os de grau 1 e 2 apresentaram menor precisão. Embora o aumento do grau pudesse melhorar o ajuste, as melhorias a partir do grau 3 foram insignificantes ou, em graus muito elevados, o modelo deixava de representar adequadamente o comportamento dos dados.

A função identificada oferece uma visão geral do comportamento da tendência ao longo do tempo. Ao analisar o gráfico, observa-se um crescimento na participação feminina nas bolsas de produtividade. Esse padrão sugere que, nos próximos anos, a presença das mulheres nessas bolsas continuará aumentando, o que representa um avanço importante para a equidade de gênero no meio acadêmico e científico.

Em conclusão, apesar dos desafios ainda existentes, há sinais positivos de mudança. A tendência observada aponta para uma maior inclusão e representatividade feminina nas áreas STEM. Com o fortalecimento de políticas de incentivo e ações afirmativas, esse crescimento pode se tornar ainda mais expressivo, contribuindo para um ambiente acadêmico e profissional mais diverso e igualitário.

Referências

- [1] A. Fernandes. “Brazilian women underrepresented in STEM fields”. Em: **Valor International** (2021). Acessado em: 21/05/2025, <https://valorinternational.globo.com/business/news/2021/09/15/brazilian-women-underrepresented-in-stem-fields.ghtml>.
- [2] A. Geron. **Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems**. 2nd. O'Reilly Media, Inc., 2019. ISBN: 1492032646.
- [3] P. A. Morettin e W. O. Bussab. **Estatística básica**. Saraiva, 2004. ISBN: 9788502081772.