

## Uso de inteligência artificial para simulação didática de dados

Rafael Sergio Gonçalves<sup>1</sup>; 0009-0009-4477-1361  
Rômulo Moura Dos Santos<sup>1</sup>; 0009-0002-5093-5069  
Leonardo Simal Moreira <sup>1</sup>; 0009-0004-1565-2201  
Italo Pinto Rodrigues<sup>1</sup>; 0000-0002-6832-8358

1 – UniFOA, Centro Universitário de Volta Redonda, Volta Redonda, RJ.  
[italoprodriques@gmail.com](mailto:italoprodriques@gmail.com) (contato principal)

**Resumo:** A integração de metodologias ativas ao ensino de Estatística e Data Science tem impulsionado a busca por recursos que conectem modelos estatísticos a contextos reais. Nesse cenário, o ChatGPT desponta como ferramenta capaz de gerar dados customizados para práticas didáticas. Este trabalho teve como objetivo desenvolver e avaliar um assistente personalizado no ChatGPT, estruturado por meio de prompt engineering, para produzir conjuntos de dados simulados completos, contextualizados e aderentes a distribuições estatísticas definidas pelo usuário. O estudo foi conduzido em caráter exploratório-computacional, sendo testado no caso “tempo de processamento até o computador travar”, modelado pela distribuição de Weibull. O assistente gerou 121 amostras acompanhadas de variáveis contextuais (carga de CPU e uso de memória), que foram analisadas por meio de tabelas de frequência, histogramas, ajuste da curva teórica e cálculo do erro médio quadrático. Os resultados evidenciaram ajuste consistente à Weibull, com parâmetros plausíveis e correlações negativas entre as variáveis contextuais e o tempo até a falha, conferindo realismo ao conjunto. A pesquisa demonstra que, quando guiado por um prompt estruturado, o ChatGPT é capaz de produzir dados estatisticamente consistentes e pedagogicamente úteis, favorecendo metodologias ativas e práticas aplicadas no ensino de Estatística.

**Palavras-chave:** ChatGPT. Geração de dados simulados. Ensino de estatística. Distribuição de Weibull. Inteligência artificial aplicada à educação.

## INTRODUÇÃO

A formação em Ciência de Dados exige não apenas domínio de técnicas de análise, mas também a habilidade de conectar modelos estatísticos a contextos reais. Nesse cenário, a possibilidade de customizar dados para fins didáticos torna-se estratégica para promover aprendizagem ativa e significativa em Estatística e Data Science. O movimento das metodologias ativas — como aprendizagem baseada em projetos, gamificação e sala de aula invertida — já vem redesenhando práticas docentes ao deslocar o protagonismo para o estudante e ampliar o engajamento nas atividades práticas (Gomez-del Rio; Rodriguez, 2022; Prince; Felder, 2006).

A convergência entre essas metodologias e a Inteligência Artificial ampliou o espaço para experiências pedagógicas com modelos de linguagem. Em particular, o ChatGPT tem sido destacado como um recurso capaz de gerar dados customizados para simulações e experimentos, apoiando professores e alunos com conjuntos aderentes a contextos reais de ensino (Brown *et al.*, 2020; Javaid *et al.*, 2023). Nessa direção, estudos de base já propõem abordagens dirigidas por prompt para produzir dados tabelados, com controle de tamanho de amostra, variáveis adicionais e pressupostos de distribuição, favorecendo reprodutibilidade e padronização da atividade didática.

Apesar desses avanços, persistem duas lacunas para o uso docente: (i) assegurar que o ChatGPT siga integralmente requisitos operacionais, e (ii) verificar, por meio de diagnóstico estatístico, se a variável de interesse realmente segue a distribuição especificada ao ponto de reproduzir a curva de ajuste no histograma — etapa já ilustrada no seu material com cálculo de erro médio quadrático (RMS) e comparação empírica-teórica (Rodrigues, 2024; Rodrigues; Campos, 2023).

Diante disso, este trabalho tem como objetivo: (1) apresentar o desenvolvimento de um assistente no ChatGPT (um GPT personalizado) que gera dados customizados para docentes de Data Science e Estatística a partir de um prompt-base com regras operacionais claras; (2) investigar se o ChatGPT é capaz de criar dados relevantes e contextualizados para uma aplicação definida pelo usuário; e (3) demonstrar, por meio de um estudo de caso, que é possível obter a curva de alguma distribuição com os dados simulados.

Como contribuição, o artigo consolida um roteiro prático para docentes: um prompt estruturado que força o ChatGPT a: (i) solicitar contexto e distribuição, (ii) produzir 121

amostras completas, (iii) evitar código externo, (iv) incluir duas variáveis contextuais adicionais, e (v) explicar as colunas e suas relações, de forma didática e reproduzível. Ao acoplar esse roteiro a um pipeline simples de verificação (tabela de frequências → histograma → ajuste à distribuição solicitada), demonstra-se um ciclo completo de geração-validação que dialoga diretamente com as demandas de ensino ativo.

## MÉTODOS

O estudo foi conduzido em caráter exploratório-computacional, com foco no desenvolvimento e na avaliação de um assistente pedagógico baseado no ChatGPT para geração de dados simulados aplicados ao ensino de estatística. Por não envolver seres humanos, não houve necessidade de submissão a comitês de ética em pesquisa. A estratégia metodológica baseou-se na criação de um GPT personalizado por meio de *prompt engineering*, no qual se estabeleceram regras operacionais precisas. O assistente foi configurado para solicitar ao usuário o contexto do conjunto de dados a ser gerado e a distribuição estatística desejada, produzindo em seguida uma tabela completa com 122 linhas, composta por um cabeçalho e 121 amostras. A tabela deveria conter obrigatoriamente uma coluna de tempo, expressa em segundos ou outra unidade equivalente, obedecendo à distribuição especificada, além de pelo menos duas variáveis adicionais relacionadas ao contexto. Foram também estabelecidas orientações para garantir variabilidade realista entre os valores, evitando repetições desnecessárias, e para que, após a geração da tabela, o assistente explicasse de forma didática o significado das colunas e a relação entre as variáveis. Nos casos em que o usuário não fornecesse todas as informações, o assistente deveria solicitar apenas os detalhes essenciais e, se necessário, assumir padrões plausíveis explicitando suas suposições.

A avaliação do desempenho do assistente foi realizada a partir de diferentes cenários contextuais, como falhas de componentes eletrônicos, resistência de materiais e tempo até falha de sistemas mecânicos. Em todos os testes, especificou-se uma distribuição, reconhecida como adequada para estudos de confiabilidade. Cada execução resultou em um conjunto de 121 amostras, que foram exportadas para planilhas eletrônicas com o objetivo de viabilizar análises subsequentes.

A análise estatística seguiu um fluxo composto por quatro etapas principais: elaboração de tabelas de frequências absolutas e relativas, construção de histogramas para a variável tempo, ajuste da curva teórica da distribuição indicada aos dados empíricos e cálculo do erro médio quadrático entre os valores observados e esperados. Esse procedimento permitiu verificar a aderência dos dados gerados à distribuição teórica, fornecendo evidências sobre a capacidade do ChatGPT em produzir dados estatisticamente consistentes e pedagogicamente úteis para aplicações didáticas.

## **RESULTADOS E DISCUSSÃO**

Os resultados obtidos demonstram que o ChatGPT, quando configurado com um prompt estruturado e regras claras de geração de dados, é capaz de produzir amostras realistas, consistentes com distribuições estatísticas específicas e contextualizadas a cenários didáticos. No caso analisado, o comportamento da variável tempo até o travamento seguiu a distribuição de Weibull, evidenciado tanto por parâmetros ajustados próximos aos valores esperados em aplicações de confiabilidade quanto pelos testes de aderência estatística. Esse achado reforça a viabilidade de empregar a ferramenta como gerador de dados simulados para fins de ensino, atendendo a uma lacuna recorrente identificada entre professores: a dificuldade de obter conjuntos de dados completos, variáveis e ajustáveis a diferentes contextos de sala de aula.

A incorporação de variáveis adicionais — carga de CPU e uso de memória — mostrou-se relevante não apenas para enriquecer a análise estatística, mas também para estimular a interpretação integrada de múltiplos fatores, aproximando os estudantes de situações reais de modelagem e confiabilidade de sistemas. A correlação negativa observada entre tempo até a falha e as variáveis de estresse computacional ilustra a capacidade do ChatGPT de produzir dados internamente coerentes, favorecendo atividades pedagógicas que combinam análise exploratória, ajuste de modelos e interpretação prática.

Do ponto de vista metodológico, a estratégia de validar os dados gerados por meio de histogramas, testes de aderência e cálculo do erro médio quadrático se mostrou fundamental para assegurar credibilidade ao material produzido. Esse procedimento, ao ser incorporado em atividades didáticas, pode estimular os alunos a desenvolverem senso crítico quanto à

qualidade de dados simulados ou empíricos, além de destacar a importância da verificação estatística em qualquer estudo.

Em termos de contribuição para o ensino, a abordagem proposta permite integrar práticas de Estatística com a utilização de inteligência artificial de forma didática, transparente e reproduzível. Mais do que fornecer dados prontos, o GPT personalizado atua como um recurso interativo que instiga o estudante a compreender pressupostos, parâmetros e relações entre variáveis. Tal potencial contribui para o alinhamento com metodologias ativas de ensino, nas quais o aprendizado ocorre de modo aplicado, colaborativo e centrado na resolução de problemas.

A aplicação do assistente ao estudo de caso “tempo de processamento até o computador travar” resultou em uma base de 121 amostras completas, contemplando três variáveis: tempo até travar (s), carga média de CPU (%) e uso de memória (GB). Os valores mostraram-se variáveis e não repetitivos, preservando plausibilidade com o contexto computacional. O tempo até a falha apresentou distribuição assimétrica à direita, com mínimo de 158,70 s, máximo de 867,50 s, mediana de 455,70 s e média de 478,59 s (DP = 184,36 s). A análise descritiva apontou que 52% dos travamentos ocorreram entre 344 s e 599 s, enquanto aproximadamente 10% ultrapassaram 800 s, sugerindo a presença de cauda longa — característica compatível com a distribuição de Weibull, conforme Tabela 1 e Figuras 1 e 2. Na Tabela 1, lê-se as seguintes informações:

- $n_i$  – Frequência;  $N_i$  – Frequência acumulada;  $f_i$  – Frequência relativa;  $F_i$  – Frequência relativa acumulada;  $f(t)$  - Função Densidade de Probabilidade (f.d.p)

Em que a  $f(t)$  de Weibull é dada por:

$$f(t) = \frac{\beta}{\delta} \left( \frac{t - \gamma}{\delta} \right)^{\beta-1} * e \left[ - \left( \frac{t - \gamma}{\delta} \right)^\beta \right]$$

Onde

- $\beta$  = Parâmetro da forma;  $\delta$  = Parâmetro da escala;  $\gamma$  = Parâmetro de localidade da distribuição

O  $E_{RMS}$  é calculado considerando quantidade total de amostras  $K$ .



$$E_{RMS} = \sqrt{\frac{1}{K} \sum_{i=1}^K (f(t_h) - f(t_w))^2}$$

Tabela 1 – Tabela de frequências.

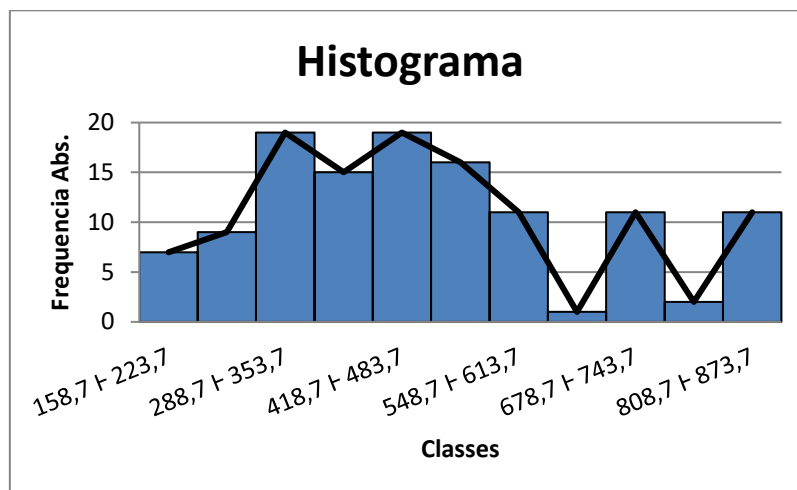
i	Classes	$n_i$	Ponto Médio	$N_i$	$f_i$	$F_i$	$f(t_h)$ Histograma	$f(t_w)$ Weibull	Quadrado do Desvio ( $f_i$ e $f(t)$ )
1	158,7 - 223,7	7	191,2	7	0,058	0,058	0,000890019	0,00101332824	0,00000001521
2	223,7 - 288,7	9	256,2	16	0,074	0,132	0,00114431	0,00139887415	0,00000006480
3	288,7 - 353,7	19	321,2	35	0,157	0,289	0,002415766	0,00169143522	0,00000052466
4	353,7 - 418,7	15	386,2	50	0,124	0,413	0,001907184	0,00184784555	0,00000000352
5	418,7 - 483,7	19	451,2	69	0,157	0,570	0,002415766	0,00185165656	0,00000031822
6	483,7 - 548,7	16	516,2	85	0,132	0,702	0,002034329	0,00171545080	0,00000010168
7	548,7 - 613,7	11	581,2	96	0,091	0,793	0,001398601	0,00147559624	0,00000000593
8	613,7 - 678,7	1	646,2	97	0,008	0,802	0,000127146	0,00118110704	0,00000111083
9	678,7 - 743,7	11	711,2	108	0,091	0,893	0,001398601	0,00088059661	0,00000026833
10	743,7 - 808,7	2	776,2	110	0,017	0,909	0,000254291	0,00061169657	0,00000012774
11	808,7 - 873,7	11	841,2	121	0,091	1,000	0,001398601	0,00039578572	0,00000100564
					1			$E_{RMS}$	0,000567815

Fonte: Elaborada pelos autores.

Os resultados mostram que o GPT personalizado atendeu às regras operacionais estabelecidas, gerando uma base completa e estatisticamente consistente com a distribuição Weibull. Além disso, as variáveis contextuais se relacionaram de modo lógico ao fenômeno de interesse, o que confere realismo pedagógico e potencial didático ao conjunto.

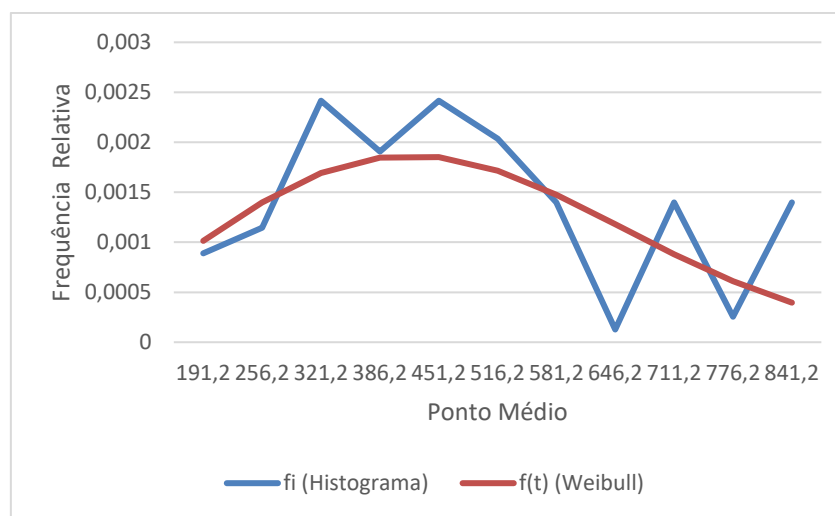


Figura 1 – Histograma dos dados.



Fonte: Elaborada pelos autores.

Figura 2 – Comparação entre os dados gerados pelo ChatGPT e a Distribuição Weibull.



Fonte: Elaborada pelos autores.

## CONCLUSÕES

O estudo demonstrou que a personalização do ChatGPT por meio de um prompt estruturado possibilita a geração de bases de dados simuladas adequadas ao ensino de Estatística e Data Science. O caso analisado, referente ao tempo de processamento até o travamento de computadores, evidenciou que os dados produzidos foram não apenas completos e variados, mas também consistentes com a distribuição de Weibull, validada por análises

estatísticas de aderência e por métricas de erro. Além disso, a inclusão de variáveis contextuais, como carga de CPU e uso de memória, reforçou a plausibilidade do conjunto e ampliou seu potencial pedagógico ao permitir interpretações multifatoriais.

Ao articular a geração automática de dados com procedimentos de verificação estatística, a proposta contribui para reduzir a dependência de bases externas e amplia as possibilidades de prática em sala de aula, favorecendo metodologias ativas e o aprendizado aplicado. Nesse sentido, a pesquisa evidencia que modelos de linguagem podem desempenhar papel estratégico na construção de recursos educacionais inovadores, desde que acompanhados de critérios claros de validação e de uso didático.

## AGRADECIMENTOS

Os autores agradecem ao Centro Universitário de Volta Redonda (UniFOA) pelo apoio institucional e financeiro, por meio do Programa Institucional de Bolsas de Iniciação Científica (PIBIC) (95056/17/RPE).

## REFERÊNCIAS

BROWN, Tom B.; MANN, Benjamin; RYDER, Nick; SUBBIAH, Melanie; KAPLAN, Jared; DHARIWAL, Prafulla; NEELAKANTAN, Arvind; SHYAM, Pranav; SASTRY, Girish; ASKELL, Amanda; AGARWAL, Sandhini; HERBERT-VOSS, Ariel; KRUEGER, Gretchen; HENIGHAN, Tom; CHILD, Rewon; RAMESH, Aditya; ZIEGLER, Daniel M.; WU, Jeffrey; WINTER, Clemens; HESSE, Christopher; CHEN, Mark; SIGLER, Eric; LITWIN, Mateusz; GRAY, Scott; CHESS, Benjamin; CLARK, Jack; BERNER, Christopher; MCCANDLISH, Sam; RADFORD, Alec; SUTSKEVER, Ilya; AMODEI, Dario. Language Models are Few-Shot Learners. [s. l.], 2020. DOI 10.48550/ARXIV.2005.14165. Disponível em: <https://arxiv.org/abs/2005.14165>. Acesso em: 22 jun. 2025.

GOMEZ-DEL RIO, T.; RODRIGUEZ, J. Design and assessment of a project-based learning in a laboratory for integrating knowledge and improving engineering design skills. **Education for Chemical Engineers**, [s. l.], v. 40, p. 17–28, jul. 2022. <https://doi.org/10.1016/j.ece.2022.04.002>.

JAVOID, Mohd; HALEEM, Abid; SINGH, Ravi Pratap; KHAN, Shahbaz; KHAN, Ibrahim Haleem. Unlocking the opportunities through ChatGPT Tool towards ameliorating the education system. **BenchCouncil Transactions on Benchmarks, Standards and Evaluations**, [s. l.], v. 3, n. 2, p. 100115, jun. 2023. <https://doi.org/10.1016/j.tbench.2023.100115>.

PRINCE, Michael J.; FELDER, Richard M. Inductive Teaching and Learning Methods: Definitions, Comparisons, and Research Bases. **Journal of Engineering Education**, [s. l.], v. 95, n. 2, p. 123–138, abr. 2006. <https://doi.org/10.1002/j.2168-9830.2006.tb00884.x>.



4º Congresso Brasileiro  
de Ciência e Saberes  
Multidisciplinares

tudo é  
ciência

11º Encontro de Extensão  
Universitária do UNIFOA

23 a 25  
de outubro

Submissões abertas até 07/09

RODRIGUES, Italo Pinto. O ChatGPT na customização de dados para ensino de data science. *In: BRASILIAN CONGRESS OF ENGINEERING EDUCATION, 2024. Proceedings of the LII Brazilian Congress of Engineering Education [...]. [S. l.]: Associação Brasileira de Educação em Engenharia, 2024. DOI 10.37702/2175-957X.COBENGE.2024.5158. Disponível em: [https://abenge.org.br/sis\\_artigo\\_com\\_capa.php/?cod\\_trab=5158](https://abenge.org.br/sis_artigo_com_capa.php/?cod_trab=5158). Acesso em: 16 set. 2025.*

RODRIGUES, Italo Pinto; CAMPOS, Alexandra Matos. ChatGPT como ferramenta de geração de dados para modelagem estatística: uma abordagem prática na disciplina de Data Science. *In: TUDO É CIÊNCIA: CONGRESSO BRASILEIRO DE CIÊNCIAS E SABERES MULTIDISCIPLINARES, 20 dez. 2023. [S. l.: s. n.], 20 dez. 2023. DOI 10.47385/tudoeciencia.975.2023. Disponível em: <https://conferenciasunifoa.emnuvens.com.br/tc/article/view/975>. Acesso em: 16 set. 2025.*