

Modelo t-Student e suas Características

Milene V. Figueira¹

PPGBEA/UFRPE, Recife, PE e UABJ/UFRPE, Belo Jardim, PE

Luana G. Torres², Josimar M. Vasconcelos³, Guilherme R. Moreira⁴, Paulo J. Duarte Neto⁵

PPGBEA/UFRPE, Recife, PE

Nas ciências aplicadas é comum utilizar as distribuições de probabilidade para modelar determinado conjunto de dados, sem questionar as características da distribuição envolvida, se é possível de fato aplicar o modelo ou como este foi formulado. Diante dessa falta de compreensão ou uma falta de solidez científica nessa natureza, foi vista a necessidade de elaborar uma revisão conceitual com respeito a distribuição t-Student, para poder compreender melhor o modelo e como utilizá-lo de maneira mais consistente.

Nessa revisão, focou-se tanto a parte teórica quanto a aplicada. No contexto teórico, ilustrou-se as principais propriedades e na parte prática algumas aplicações com dados reais. Essa distribuição se caracteriza por ser paramétrica, semelhante a distribuição normal, ter caudas pesadas, ser simétrica e trabalhar com graus de liberdade.

O modelo t-Student foi desenvolvido por William Sealy Gosset, um estatístico inglês que trabalhava para a cervejaria Guinness, no início do século XX [3]. Ele tinha a limitação do número de amostras e até então não tinha nenhuma distribuição que contemplasse essa situação, por isso desenvolveu esse modelo. A maior motivação em aplicar essa distribuição de probabilidade é quando estamos trabalhando com pequenas amostras, principalmente em estudos experimentais [1].

O material teórico dessa distribuição foi focado nas principais propriedades da distribuição, ou seja: simetria, os graus de liberdade, média, variância, convergência para função normal e o intervalo de confiança. Os cálculos partiram da função densidade de probabilidade:

$$f\left(\frac{x}{\nu}\right) = \frac{1}{\sqrt{\nu}\beta\left(\frac{1}{2}, \frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad (1)$$

onde ν são os graus de liberdade (que representa o número de observações independentes), x é um valor na reta real ($x \in \mathbb{R}$) e β é a função Beta. Essa função é utilizada para normalizar a distribuição t-Student, que é definida da seguinte forma:

$$\beta(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \quad (2)$$

onde Γ é a função Gamma.

Para entender melhor a função, começou-se pela verificação se a distribuição atendia as propriedades para ser uma função densidade de probabilidade, ou seja, a função ser não negativa para todos os valores de x e a integral da função ter valor um em todo o seu domínio. Foram

¹milene.figueira@ufrpe.br

²luana.goncalvestorres@ufrpe.br

³josimar.vasconcelos@ufrpe.br

⁴guilherme.rochamoreira@ufrpe.br

⁵pjduarteneto@gmail.com

demonstrados todos os cálculos necessários para atender esses requisitos. Além disso, foi feita uma simulação utilizando a linguagem de programação R, fazendo o gráfico da função com alguns graus de liberdade e comparando as curvas geradas entre si. Foi verificado que quanto menor o grau de liberdade, maior é a cauda e menor é o ponto de máximo da função, conforme já era esperado, para manter a relação de área igual a um.

Após essa verificação, foram analisadas as principais propriedades da distribuição que são: simetria e graus de liberdade, média e variância, as caudas pesadas, convergência para função normal e o intervalo de confiança. Para essa verificação foram feitos cálculos de integrais que exigiram algumas substituições e que requeriam um domínio mais avançado de cálculo.

Também calculou-se a função característica da distribuição e sua função geradora de momentos. Foi feito o histograma e graficado junto com a função densidade de probabilidade. Foram feitas medidas de estatística descritiva da distribuição e foi testado no R o tamanho da amostra para que a função se aproximasse da distribuição normal. Assim como, a convergência com a distribuição normal, F-Snedecor e demais modelos que têm afinidades.

Na parte prática, gerou-se o histograma para mostrar a proximidade com a normal conjuntamente com a sua função de densidade de probabilidade. Os cálculos e histogramas foram feitos utilizando pacotes da linguagem R [2]. Foram testados alguns bancos de dados que eram adequados ao modelo, os cálculos foram feitos com a função do modelo e com o R. Os resultados foram compatíveis utilizando os dois métodos.

Por fim, aplicou-se testes estatísticos para mostrar a eficiência nas aplicações desse modelo em dados reais. Ainda foram feitos testes para verificar a normalidade da distribuição, como a quantidade de dados da distribuição t-Student deve ser pequena, nesse caso os testes mais adequados são Shapiro-Wilk e Shapiro-Francia. Também foram utilizados os teste Lilie e Kolmogorov-Smirnov. Em todos os testes a hipótese nula foi aceita.

O principal ponto desse estudo foi entender as particularidades da distribuição e compreender o seu funcionamento. Além de entender as funções envolvidas em sua definição.

Referências

- [1] H. A. Miot. “Avaliação da normalidade dos dados em estudos clínicos e experimentais”. Em: **Jornal Vascular Brasileiro** 16.2 (abr. de 2017). Acessado em 19/03/2024, pp. 88–91. ISSN: 1677-5449. DOI: 10.1590/1677-5449.041117. URL: <https://doi.org/10.1590/1677-5449.041117>.
- [2] T. Olivoto e B. Giacomini Sari. **Software R para avaliação de dados experimentais**. Online. Acessado em 19/03/2024, <https://tiagoolivoto.github.io/e-bookr/index.html>.
- [3] T. N. K. Raju. “William Sealy Gosset and William A. Silverman: Two “Students” of Science”. Em: **Pediatrics** 116.3 (set. de 2005). Acessado em 19/03/2024, pp. 732–735. ISSN: 0031-4005. DOI: 10.1542/peds.2005-1134. URL: <https://doi.org/10.1542/peds.2005-1134>.