



Big Data Analytics com R e Microsoft Azure Machine Learning 3.0

Big Data Analytics com R e Microsoft Azure Machine Learning Versão 3.0

Estatística na Prática 2 Definindo o Teste t



O Teste t é um teste estatístico frequentemente utilizado para testar hipóteses sobre diferenças entre até duas médias (para mais de duas médias usamos a ANOVA).

Os Testes t são testes de hipótese úteis na Estatística quando é necessário comparar médias. Você pode comparar uma média amostral com um valor hipotético ou com um valor alvo usando um Teste t para uma amostra. Você pode comparar as médias de dois grupos com um Teste t para duas amostras. Se você tiver dois grupos com observações pareadas (por exemplo, antes e depois das medições), use o Teste t pareado.

É possível usar o Teste t para comparar a média de uma amostra com a média populacional (one sample t test), para comparar duas médias amostrais (two sample t test) ou para comparar duas médias de uma mesma amostra que foi investigada em dois momentos do tempo (paired ou matched t test).

Por utilizar dados amostrais da média para estimar a média populacional (parâmetro μ), ele é considerado um teste paramétrico.

O Teste t possui as seguintes suposições:

- 1- Os dados são aleatórios e representativos da população.
- 2- A variável dependente é contínua.
- 3- Ambos os grupos são independentes (ou seja, grupos exaustivos e excludentes).
- 4- Os resíduos do modelo são normalmente distribuídos.
- 5- A variância residual é homogênea (princípio da homocedasticidade).

Para o nosso exemplo neste estudo de caso, iremos considerar como verdadeiras as suposições de 1 a 3 e validaremos as suposições 4 e 5. Para a suposição 4 usaremos o Teste de Shapiro-Wilk e para a suposição 5 usaremos o Teste F.

Eventualmente, quando os pressupostos do modelo são violados, a literatura recomenda que testes não-paramétricos com propostas parecidas sejam usados. Para o nosso exemplo os pressupostos não foram violados e por isso o Teste t foi aplicado.

Hipóteses do Teste t:

- H₀ (Hipótese Nula) = Não há diferença significativa entre as médias dos 2 grupos.
- H_a (Hipótese Alternativa) = Há diferença significativa entre as médias dos 2 grupos.

Aplicamos o Teste t, interpretamos o resultado e decidimos se rejeitamos ou não a H₀. É o que faremos nas aulas a seguir.