



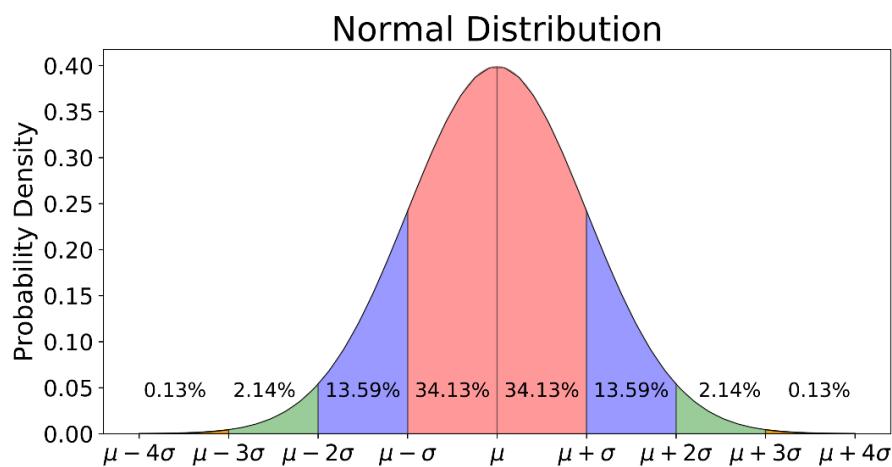
Big Data Analytics com R e Microsoft Azure Machine Learning 3.0

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

Estatística na Prática 2 Definindo o Teste de Shapiro-Wilk



O Teste de Shapiro-Wilk tem como objetivo avaliar se uma distribuição é semelhante a uma distribuição normal. A distribuição normal também pode ser chamada de gaussiana e sua forma assemelha-se a de um sino. Esse tipo de distribuição é muito importante, por ser frequentemente usada para modelar fenômenos naturais. A imagem abaixo ilustra a distribuição normal de uma variável aleatória:



Na prática, podemos querer saber, por exemplo, se o peso dos participantes da nossa amostra segue ou não uma distribuição normal. Para isso, podemos usar o Teste de Shapiro-Wilk.

Como resultado, o teste retornará a estatística W, que terá um valor de significância associada, o valor-p. Para dizer que uma distribuição é normal, o valor-p precisa ser maior do que 0,05.

Precisamos do Teste de Shapiro-Wilk para este estudo de caso pois uma das suposições do Teste t é que os resíduos seguem uma distribuição normal (simplificando: os dados seguem uma distribuição normal). Logo, validamos a suposição com o Teste de Shapiro-Wilk antes de aplicar o Teste t.

Assim como a maioria dos testes estatísticos, o Teste de Shapiro-Wilk formula duas hipóteses:

- H0 (Hipótese Nula) = Os dados seguem uma distribuição normal.
- Ha (Hipótese Alternativa) = Os dados não seguem uma distribuição normal.

Aplicamos o teste, interpretamos o resultado e então decidimos se rejeitamos ou não a H0. É exatamente o que faremos a seguir.