



# Data Science Academy

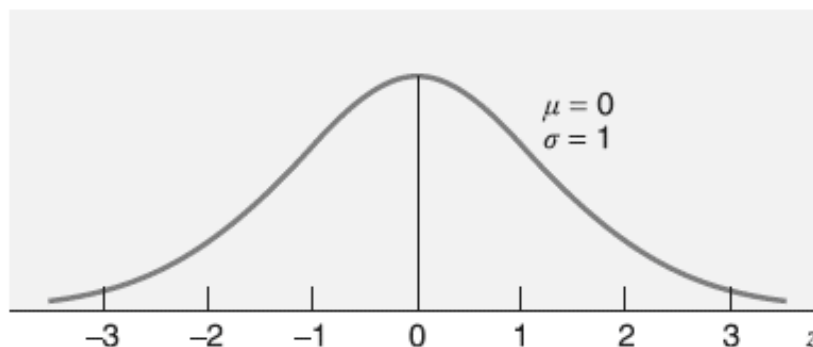
[www.datascienceacademy.com.br](http://www.datascienceacademy.com.br)

## Microsoft Power BI Para Data Science

### Distribuição Normal Padronizada e Escore Z

A distribuição normal padronizada representa um caso especial da distribuição normal. Para a distribuição normal padronizada, o valor da média aritmética é igual a zero e o valor do desvio-padrão é igual a 1.

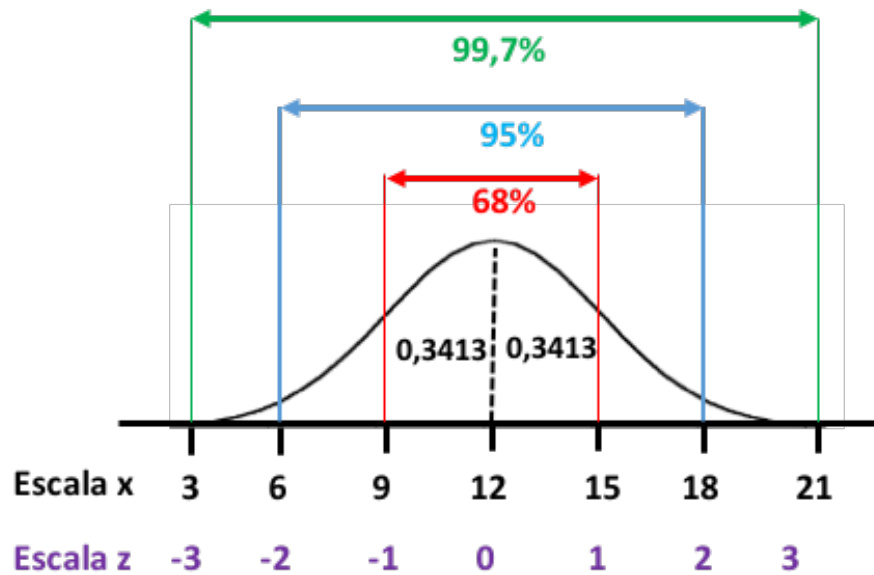
A figura abaixo exibe a curva da distribuição normal padronizada. A variável aleatória que possui a distribuição normal padronizada é representada por  $z$ . Em outras palavras, as unidades correspondentes à curva da distribuição normal padronizada são representadas por  $z$ , e são chamadas de valores de  $z$ , resultados de  $z$ , escore  $z$  ou em inglês,  $z$ -score. Essas unidades são chamadas também de unidades-padrão ou resultados-padrão.



Na figura acima, o eixo horizontal é rotulado como  $z$ . Os valores de  $z$  no lado direito da média aritmética são positivos, enquanto aqueles no lado esquerdo são negativos. O valor de  $z$  para um ponto no eixo horizontal fornece a distância entre a média aritmética e aquele ponto específico, em termos do desvio-padrão. Por exemplo, um ponto com um valor de  $z = 2$  encontra-se duas unidades de desvio-padrão à direita da média aritmética. De maneira similar, um ponto com um valor de  $z = -2$  encontra-se duas unidades de desvio-padrão à esquerda da média aritmética. Embora os valores de  $z$  no lado esquerdo da média aritmética sejam negativos, a área sob a curva é sempre positiva.

A área sob a curva da normal padronizada, entre quaisquer dois pontos, pode ser interpretada como a probabilidade de que  $z$  venha a assumir um determinado valor dentro dos limites daquele intervalo.

Dentre as muitas propriedades da distribuição normal, uma merece destaque especial, a regra empírica. A Regra Empírica define o seguinte: se uma distribuição é simétrica e em formato de sino, aproximadamente 68%, 95% e 99% dos dados desta distribuição estarão em 1, 2 e 3, desvio padrões acima e abaixo da média, respectivamente:



Considerando o exemplo do vídeo anterior, de acordo com a regra empírica, esperamos que **68%** das ligações fiquem entre **9** e **15** minutos, **95%** entre **6** e **18** minutos e **99%** entre **3** e **21** minutos. A partir daí várias inferências podem ser feitas, quando usamos uma distribuição normal.