



Data Science Academy

www.datascienceacademy.com.br

Microsoft Power BI Para Data Science

Escore Z

Uma das medidas mais comuns em Estatística é o **Escore z**. O Escore z identifica o número de desvios padrões que um determinado valor possui de distância em relação à média do conjunto de dados ao qual faz parte. Pense no Escore z como uma simples conversão de um valor para uma outra unidade específica, a fim de fazer conclusões. O Escore z em si, não possui unidade. É apenas um número. Vamos a um exemplo. A tabela abaixo, mostra o número de calorias de sanduíches:

Hamburger	Restaurante	Calorias
Cheeseburger	McDonalds's	300
Big Mac	McDonalds's	430
Whopper	Burger King	540
Double Cheeseburger	Bob's	670
Chicken Burger	McDonalds's	780
Bacon Burger	Bob's	840
Quarteirão	McDonalds's	1.230
Mega Burger	Burger King	1.420
Média da amostra		776,30
Desvio padrão da amostra		385,10

Vamos calcular o Escore z para o **Mega Burger** do Burguer King

$$\text{Escore}_z = (x - x_1) / s = (1.420 - 776,30) / 385,10 = 1.67$$

Onde x é o valor individual, x1 é a média da amostra e s é o desvio padrão. O que significa o resultado 1.67? Significa que as calorias do **Mega Burger**, são **1.67 desvio padrão** acima da média. Agora vamos calcular o Escore z do **Big Mac**:

$$\text{Escore}_z = (x - x_1) / s = (430 - 776,30) / 385,10 = -0.90$$

O valor negativo, indica que as calorias do **Big Mac**, estão abaixo da média, neste caso, exatamente 0.90 abaixo da média. O que nos leva a considerar que de acordo com os padrões calóricos o ideal é optar pelo Big Mac uma vez que ele está abaixo do valor médio de calorias da média de todos os sanduíches que participaram deste experimento. O **Escore z** terá sempre um dos 3 atributos:

- **Positivo** – valores acima da média
- **Negativo** – valores abaixo da média
- **Zero** – valores iguais a média

O **Escore z** também possui uma característica importante: ele nos ajuda a identificar os outliers (valores extremos) dos nossos dados. Valores de dados que possuem **Escore z** acima de **+ 3** ou abaixo de **-3** são classificados como outliers.