

# NORMALIZAÇÃO E PADRONIZAÇÃO DE DADOS

## MATHEUS HENRIQUE BUTKOSKI SILVA - 4 SEMESTRE

- 1) Redimensionar os dados se baseia em transformar os dados (sem afetar o valor), para que todos de uma amostra fiquem na mesma escala, é importante para facilitar a compreensão dos resultados e agilizar os processos de comparação.
- 2) **Padronização** de dados é o processo de colocar variáveis diferentes na mesma escala. Isso permite comparar pontuações entre diferentes tipos de variáveis. Normalmente, para padronizar variáveis, é necessário calcular a média e o desvio padrão para uma variável. Já a **Normalização** possui o objetivo de alterar os valores das colunas numéricas no conjunto de dados para uma escala comum, sem distorcer as diferenças nos intervalos de valores. É frequentemente utilizado quando os parâmetros possuem intervalos muito diferentes. A principal diferença é que **padronizar** as variáveis irá resultar em uma média igual a 0 e um desvio padrão igual a 1. Já **normalizar** tem como objetivo colocar as variáveis dentro do intervalo de 0 e 1, caso tenha resultado negativo -1 e 1. A Padronização de dados normalmente é feita utilizando a fórmula **z-score**.
- 3) Como dito na última resposta, a normalização é frequentemente utilizada quando os parâmetros possuem intervalos muito diferentes. Além disso, a normalização também é utilizada quando não se sabe a distribuição dos dados, ou quando se sabe que não é uma gaussiana. Já a padronização é utilizada quando sabe-se que a distribuição dos dados tem uma distribuição gaussiana, ou parecida com a curva da gaussiana.

### REFERÊNCIAS:

<https://fazerpergunta.com/biblioteca/artigo/read/378583-qual-a-diferenca-entre-padronizacao-e-normalizacao>  
<https://cursos.alura.com.br/forum/topico-normalizar-ou-padronizar-148644>