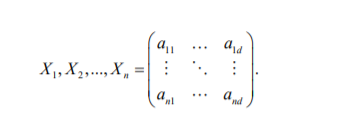
# Z-Score Definição e Exemplo

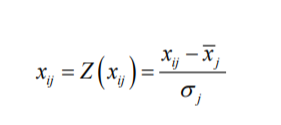
## Definição

Given Y = {X1, X2, …, Xn} denote the dimensional raw data set.

Then the data matrix is an n×d matrix given by:



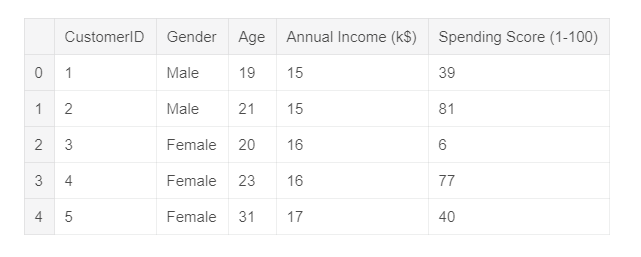
The Z-score is a form of standardization used for transforming normal variants to standard score form. Given a set of raw data Y, the Z-score standardization formula is defined as:

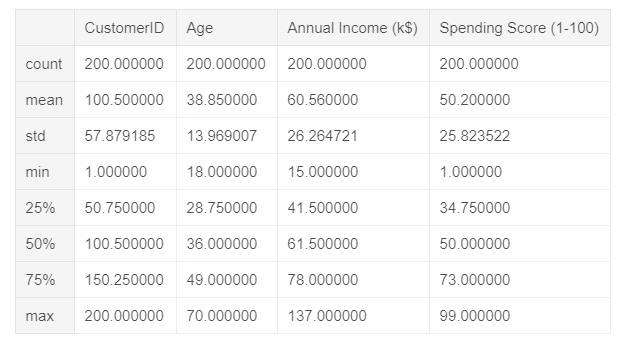


where, and are the sample mean and standard deviation of the jth attribute, respectively. The transformed variable will have a mean of 0 and a variance of 1. The location and scale information of the original variable has been lost (Jain and Dubes, 1988). One important restriction of the Z-score standardization is that it must be applied in global standardization and not in within-cluster standardization (Milligan and Cooper, 1988).

## Exemplo

Pegando um conjunto de dados Mall Customer Segmentation Data podemos observar os dados:





Ao fazer o gráfico de dispersão da idade pela renda anual podemos ver que devido a diferença de dispersão o gráfico fica com os pontos mais próximos.

Aplica-se o z-score em cada ponto os dos dados e pode-se observar uma melhor distribuição dos pontos. Em que a média é próxima de zero e a variância passa a ser próxima de 1.

Fontes

Mall Customer Segmentation Data – link:https://www.kaggle.com/vjchoudhary7/customer-segmentation-tutorial-in-python

Standardization and Its Effects on K-Means Clustering Algorithm – link: <https://pdfs.semanticscholar.org/1d35/2dd5f030589ecfe8910ab1cc0dd320bf600d.pdf>