

3-Statistics 3

أكاديمية طويق
Tuwaiq Academy



By: Eng. Esraa Madhi

Statistical Concepts for Data Scientists - Data Normalization/Scaling

Normalization is just a **feature scaling** technique that is needed for our features while feeding to the model of machine learning.

Z - Score:

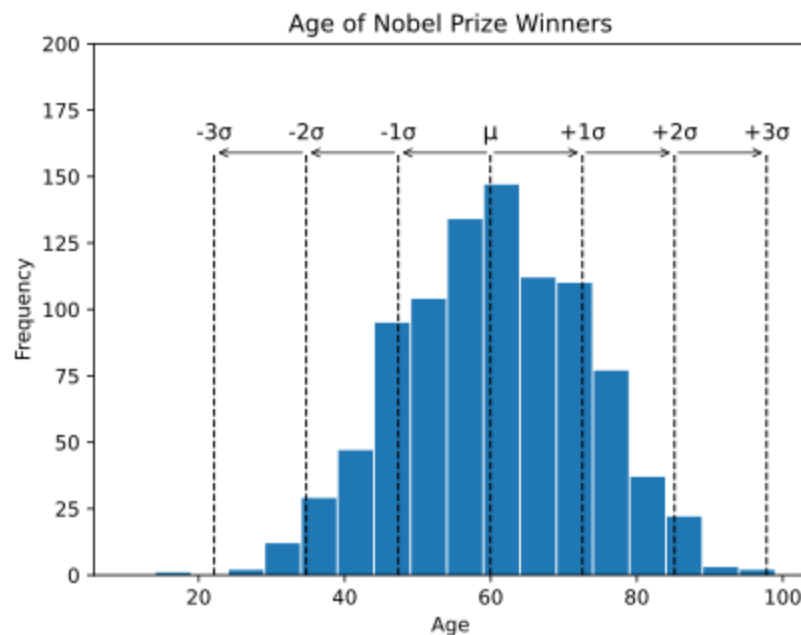
The Z-score is the signed number of standard deviations by which the value of an observation or data point is above the mean value of what is being observed or measured.

Z-score normalization refers to the process of normalizing every value in a dataset such that the mean of all of the values is 0 and the standard deviation is 1. It normalizes the data into a simpler form.

We use the following formula to perform a z-score normalization on every value in a dataset:

$$\text{New value} = (x - \mu) / \sigma$$

This formula allows the conversion of the distance of any x value from its mean into standard deviation units.



User-uploaded image: img_histogram_std.svg

Here are some important facts about z-scores:

- A positive z-score says the data point is above average.
- A negative z-score says the data point is below average.
- A z-score close to 0 says the data point is close to average.
- A data point can be considered unusual if its z-score is above 3 or below -3.

Example:

<https://www.statology.org/z-score-normalization/>



Z-Score Normalization: Definition & Examples - Statology • www.statology.org