

Standardization of data in statistics involves transforming variables to have a mean of zero and a standard deviation of one. This is typically achieved by subtracting the mean of the variable and dividing by its standard deviation:

$$z = \frac{x - \mu}{\sigma}$$

where:

- x is the original value,
- μ is the mean of the variable,
- σ is the standard deviation of the variable,
- z is the standardized value.

Why standardize data?

1. **Comparison:** It allows for comparison between variables with different units or scales by bringing them to a common scale.
2. **Improved Performance:** Many machine learning algorithms, like gradient descent and algorithms using distances (e.g., k-nearest neighbors), perform better or converge faster with standardized data.
3. **Assumption of Algorithms:** Some algorithms, such as linear regression and support vector machines, assume data is centered around zero and scaled uniformly.

Standardization helps in ensuring that each feature contributes equally to the model and improves the overall effectiveness of the training process.

