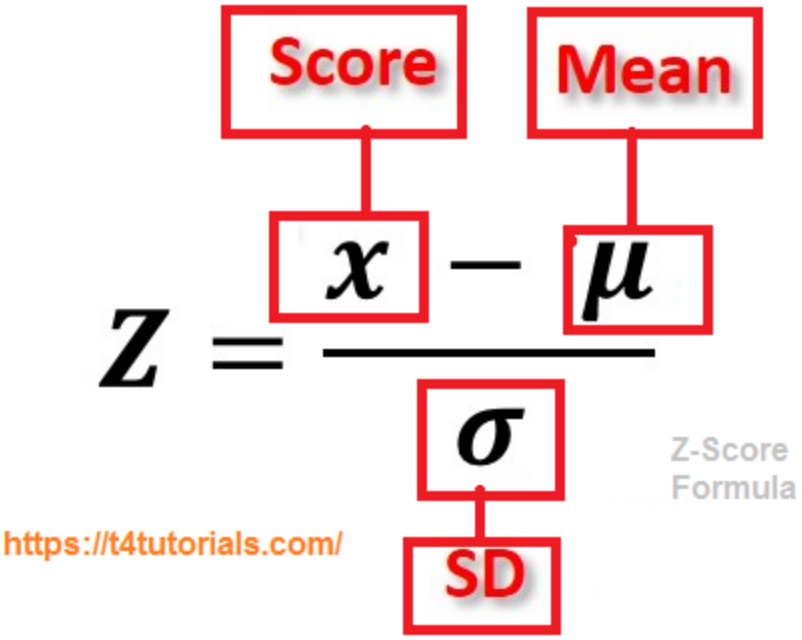
# Z-Score Normalization (Data Mining)



## Z-Score Normalization – (Data Mining)

Z-Score helps in the normalization of data. If we normalize the data into a simpler form with the help of z score normalization, then it's very easy to understand by our brains.

## Z- Scroe Formula



# How to calculate Z-Score of the following data?

| marks |
| --- |
| 8 |
| 10 |
| 15 |
| 20 |

$$s = \sqrt{\frac{\sum(x - \bar{x})^2}{n}}$$



$$\text{Standard deviation} = \sqrt{\frac{\sum(\text{every individual value of marks} - \text{mean of marks})^2}{n}}$$



Mean of marks = 8 + 10 + 15 + 20 /4 = 13.25

$$= \sqrt{\frac{(8 - 13.25)^2 + (10 - 13.25)^2 + (15 - 13.25)^2 + (20 - 13.25)^2}{4}}$$

$$= \sqrt{\frac{(-5.25)^2 + (-3.25)^2 + (1.75)^2 + (6.75)^2}{4}}$$



$$= \sqrt{\frac{27.56 + 10.56 + 3.06 + 45.56}{4}} = \sqrt{\frac{86.74}{4}} = \sqrt{21.6} = 4.6$$

**Figure:** z score normalization standard deviation

**Mean =** 13.25

**Standard deviation =** 5.37

$$ZScore = \frac{x-\mu}{\sigma} = \frac{8-13.25}{4.6} = -1.14$$



$$ZScore = \frac{x-\mu}{\sigma} = \frac{10-13.25}{4.6} = -0.7$$



$$ZScore = \frac{x-\mu}{\sigma} = \frac{15-13.25}{4.6} = 0.3$$

$$ZScore = \frac{x-\mu}{\sigma} = \frac{20-13.25}{4.6} = 1.4$$

**Figure:** z score normalization data mining

| **marks** | marks after z-score normalization |
|---|---|
| 8 | -1.14 |
| 10 | -0.7 |
| 15 | 0.3 |
| 20 | 1.4 |

## Download Excel File Calculations

**Click Here**

F63

| | A | B | C | D | E | F | G | H | I | J |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | | | https://T4Tutorials.com | | | | | | |
| 2 | | | | | | after z-score normalization | | | | |
| 3 | Id | Depend | Sal | Eucledean | Id | Depend | Norm-Dep | Salary | Norm-Sal | Eucledean |
| 4 | E101 | 3 | 50000 | 0 | E101 | 3 | -0.5287788 | 50000 | 4.2954 | 0 |
| 5 | E105 | 5 | 50000 | 49999.37304 | E110 | 3 | -0.5287788 | 45000 | 3.86583 | 0.4295716 |
| 6 | E110 | 3 | 45000 | 5000 | E113 | 3 | -0.5287788 | 57000 | 4.8968 | 0.6014003 |
| 7 | E113 | 3 | 57000 | 7000 | E114 | 3 | -0.5287788 | 42000 | 3.60809 | 0.6873146 |
| 8 | E111 | 6 | 43000 | 7000.000643 | E112 | 4 | 0.293766 | 39000 | 3.35035 | 1.2528822 |
| 9 | E114 | 3 | 42000 | 8000 | E107 | 3 | -0.5287788 | 35000 | 3.00669 | 1.2887149 |
| 10 | E109 | 5 | 40000 | 10000.0002 | E108 | 4 | 0.293766 | 38000 | 3.26443 | 1.3188946 |
| 11 | E112 | 4 | 39000 | 11000.00005 | E102 | 4 | 0.293766 | 65000 | 5.58412 | 1.5288446 |
| 12 | E108 | 4 | 38000 | 12000.00004 | E104 | 4 | 0.293766 | 35000 | 3.00669 | 1.5288446 |
| 13 | E107 | 3 | 35000 | 15000 | E105 | 5 | 1.1163108 | 50000 | 4.2954 | 1.6450896 |
| 14 | E102 | 4 | 65000 | 15000.00003 | E103 | 3 | -0.5287788 | 70000 | 6.01369 | 1.7182865 |
| 15 | E104 | 4 | 35000 | 15000.00003 | E109 | 5 | 1.1163108 | 40000 | 3.43626 | 1.8559221 |
| 16 | E103 | 3 | 70000 | 20000 | E106 | 1 | -2.1738684 | 30000 | 2.57712 | 2.3788291 |
| 17 | E106 | 1 | 30000 | 20000.0001 | E111 | 6 | 1.9388556 | 43000 | 3.694 | 2.5398625 |

**Figure:** Z Score normalization Excel File Calculations

## How do you use a z score table?

1. We can find a specific area under the normal distribution curve.

2. We can find the z-score of the data value and use a Z-Score Table.

Z-Score Table is used to find the area.

A Z-Score Table shows the area percentage to the left of a given z-score on a standard normal distribution.

## Advantages of the z score

The z-score is very useful when we are understanding the data. Some of the useful facts are mentioned below;

The z-score is a very useful statistic of the data due to the following facts;

It allows a data administrator to understand the probability of a score occurring within the normal distribution of the data.

The z-score enables a data administrator to compare two different scores that are from different normal distributions of the data.

## Is a higher or lower Z score better?

Suppose we have data from two persons. Person A has a high Z score value and person B have low Z Score value. In this case, the higher Z-score indicates that Person A is far away from person B.

## What does a negative and a positive z score mean?

A negative z-score indicates that the data point is below the mean.
A positive z-score indicates that the data point is above the mean.

## Why is the mean of Z scores is 0?

The standard deviation of the z-scores is always 1 and similarly, the mean of the z-scores is always 1.
Z-scores values above the 0 represent that sample values are above the mean.
z-scores values below the 0 represent that sample values are below the mean.
In the case of squared z-scores, the sum of the squared z-scores is always equal to the number of z-score values.

## What is the meaning of the high Z score and low Z score?

- Suppose we have a  high z-score value then it means a very low probability of data above this z-score.
- Suppose we have a low z-score value then it means a very low probability of data below this z-score.

## Related Posts:

1. how to normalize the data with min max normalization...
2. Min Max Normalization of data in data mining
3. Standard Deviation normalization of data in data mining
4. Data Normalization before Data Mining
5. Data Stream Mining - Data Mining
6. What is data mining? What is not data mining?
7. Min Max Normalization Python and Matlab - Data Mining
8. Normalization with decimal scaling in data mining - Examples
9. Frequent pattern Mining, Closed frequent itemset, ...
10. Variance and standard deviation of data in data mining