z-score-in-statistics

April 10, 2024

[1]:
```
import numpy as np
```

Mean of the dataset, the Standard Deviation of tha dataset and the z score of data points

[4]:
```
def calculate_z_score(data):
  mean = np.mean(data)
  std_dev = np.std(data)
  z_scores = (data - mean) / std_dev
  return z_scores
```

[5]:
```
#dataset
dataset = [3,9,23,43,53,4,5,30,35,50,70,150,6,7,8,9, 10]
z_scores = calculate_z_score(dataset)
print('Z-Score :', z_scores)
```

```
Z-Score : [-0.7574907  -0.59097335 -0.20243286  0.35262498  0.6301539
-0.72973781
 -0.70198492 -0.00816262  0.13060185  0.54689523  1.10195307  3.32218443
 -0.67423202 -0.64647913 -0.61872624 -0.59097335 -0.56322046]
```

Data points which lies outside 3 standard deviatioms are outliers

[6]:
```
outliers = [data_point for data_point, \
          z_score in zip(dataset, z_scores) if z_score > 3]
print(f'\nThe outliers in the dataset is {outliers}')
```

```
The outliers in the dataset is [150]
```