**What are Anomalies?**

An anomaly, while having a seemingly complex and negative name, simply means a value’s difference from the average in a set of data.

For example, let’s say we have the following set of numbers:

{ 2, 9, 1, 5, 5, 8 ,9 }

From this set, we can extract a few statistics:

Mode (Most often occurring number) = 9

Median (Number in the center of the data) = 5

Mean (Average of all data) = (Sum of all numbers) / (How many numbers are in the set)

= (1 + 4 + 3 + 5 + 5 + 8 + 9) / 7

= 35 / 7

= 5

Now that we know the average of this set of numbers, it’s relatively easy to find an anomaly for any value in the set. For example, let’s try the anomaly for 9:

Anomaly = The absolute value of (value – average)

= abs(9 – 5)

= abs(4)

= 4

**Why is this Useful? ( + An Explanation of Sensors)**

In any modern data collection scenario, it is crucial to use sensors to assign numeric values to things we wish to observe to achieve the highest accuracy and precision possible. Things like wind speed, pressure, temperature, humidity, etc. all require sensors to conduct mass, continuous measurements. However, due to the relative nature of these sensors, even measurements of the same exact thing with different sensors can yield differing results.

For example, let’s say two people measure the temperature (in Fahrenheit) of a cup of water over a period of 10 minutes using two different sensors, and get the following values:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Temperature Values over Time | | | | | | | | | | |
| **Time (Min)** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| **Sensor 1 (F)** | 74.3 | 74.1 | 73.5 | 73.9 | 74.7 | 75.5 | 76.1 | 76.3 | 76.7 | 76.5 |
| **Sensor 2 (F)** | 75 | 74.8 | 74.2 | 74.6 | 75.4 | 76.2 | 76.8 | 77 | 77.4 | 77.2 |

Despite the fact that the two people were measuring the exact same cup of water, the values are completely different! So, how can we normalize this data in order to compare the trends in each of the measurements? By taking the anomalies of each.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Temperature Values over Time** | | | | | | | | | | | **Avgs** |
| **Time (Min)** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |  |
| **Sensor 1 (F)** | 74.3 | 74.1 | 73.5 | 73.9 | 74.7 | 75.5 | 76.1 | 76.3 | 76.7 | 76.5 | 75.16 |
| **Sensor 2 (F)** | 75 | 74.8 | 74.2 | 74.6 | 75.4 | 76.2 | 76.8 | 77 | 77.4 | 77.2 | 75.86 |
| **Sensor 1 Anomalies** | 0.9886 | 0.9859 | 0.9779 | 0.9832 | 0.9939 | 1.0045 | 1.0125 | 1.0152 | 1.0205 | 1.0178 | 1 |
| **Sensor 2 Anomalies** | 0.9887 | 0.9860 | 0.9781 | 0.9834 | 0.9939 | 1.0045 | 1.0124 | 1.0150 | 1.0203 | 1.0177 | 1 |

Graphing this data, we can see that two line up almost perfectly.

