- ○ xi is each data point in the dataset.
- ○ Median is the median value of the dataset.

5. **Find the Median** of the absolute deviations calculated in step 2.

So, the MAD can be expressed as:

$$\text{MAD} = \text{Median}(|x_1 - \text{Median}|, |x_2 - \text{Median}|, \ldots, |x_n - \text{Median}|)$$

Where:

7. X1,x2,…,xn are the data points in the dataset.

Why Use MAD?

- **Robustness to Outliers**: MAD is more robust than measures like the standard deviation because it uses the median rather than the mean, making it less sensitive to extreme outliers. This is particularly useful in datasets where outliers could distort the spread of the data.
- **Measure of Spread**: MAD provides a measure of how much the data deviates from the median, indicating the variability or consistency of the data.
- **Simplicity**: It is easy to compute and interpret, making it a practical choice in exploratory data analysis or when dealing with non-normal distributions.

Example of Calculating MAD

Let’s consider the following dataset:

$$x = \{2, 4, 6, 8, 100\}$$

1. **Step 1 - Find the Median:** The median of the dataset is **6** (since 6 is the middle value when the data points are arranged in increasing order).
2. **Step 2 - Calculate the Absolute Deviations:**
   - Absolute deviation for 2: $|2-6|=4$
   - Absolute deviation for 4: $|4-6|=2$
   - Absolute deviation for 6: $|6-6|=0$
   - Absolute deviation for 8: $|8-6|=2$
   - Absolute deviation for 100: $|100-6|=94$

   The absolute deviations are: $4, 2, 0, 2, 94$
3. **Step 3 - Find the Median of the Absolute Deviations:** Arrange the absolute deviations in order: $0, 2, 2, 4, 94$.

   The median of these values is **2**.

So, the **MAD** of the dataset is **2**.

Interpretation of MAD

- A higher MAD indicates that the data points are more spread out or have greater variability from the median.
- A lower MAD indicates that the data points are more concentrated around the median.

MAD vs. Standard Deviation

- **Standard Deviation** is sensitive to outliers because it is based on the mean and squares of deviations, which amplify the effect of large deviations.
- **MAD**, on the other hand, is based on the median and absolute deviations, which makes it less sensitive to extreme outliers.

MAD is preferred over standard deviation when you want a more **robust measure** of spread that isn't affected by outliers or skewed distributions.

Applications of MAD

5. **Outlier Detection**: MAD is often used to identify outliers. If a data point's deviation from the median is significantly larger than the MAD, it can be considered an outlier.
6. **Robust Data Analysis**: In situations where the data contains extreme outliers, MAD provides a more reliable measure of central tendency and variability.
7. **Statistical Modeling**: In robust regression or other robust statistical models, MAD is often used to down-weight outliers and reduce their influence on the model's parameters.

MAD for Standard Normal Distribution

For a standard normal distribution (mean = 0, standard deviation = 1), the MAD is approximately **0.6745** times the standard deviation.

This property makes MAD a useful tool for comparing the spread of datasets with different distributions.

Distribution of Errors in Machine Learning

The **distribution of errors** is a key concept in evaluating the performance of a machine learning model. Errors refer to the difference between the predicted values and the actual values (or ground truth) from a dataset. Analyzing the distribution of errors helps in understanding the behavior of the model and can provide insights into whether the model is performing well or if there are issues like bias, variance, or outliers.

Types of Errors

5. **Bias**: This refers to the error introduced by the model's assumptions. High bias means the model is too simplistic and underfits the data. Underfitting typically occurs when the model cannot capture the underlying patterns in the data.
6. **Variance**: This refers to the error introduced by the model's sensitivity to small fluctuations in the training data. High variance means the model is too complex and overfits the data, capturing noise and fluctuations as if they were real patterns.
7. **Residuals**: The residuals are the differences between the actual values and the predicted values, i.e., the errors made by the model.

$$e_i = y_i - \hat{y}_i$$

Where: