

Title: Estimates of Variability

Introduction

In this assignment, there are written down concepts of estimates of variability. How to find it (formulas) and the examples of calculation with dataset weather_20.csv file. To implement calculations there is a chosen Temperature feature.

Methods

Mean Absolute Deviation. We need firstly find simple mean and find the absolute ones of all elements in our feature by subtracting mean from each element. then just need to divide sum of all absolute values to length of dataframe.

Variance and Standart Deviation. Variance measures the average squared deviations from the mean when standart deviation represents in the same units as the data. if in MAD we finding absolute values of each elements by subtracting the mean, there we need to find the squared ones and sum them all. So, In next step there are two ways of calculation: for All data and for sample one of it. In first case we just dividing it by the length of dataframe but with second one we need to subtract one from length and then divide it.

Median Absolute Deviation. This variability of data is more robust to outliers rather than others and with this it is ideal for skewed distributions. The calculation of it starts with finding median of dataframe and calculating the absolute values by subtracting the median from each element. And last stage is just sort the absolute values and find median from it.

Findings

The simple mean now is 19.26 where mean AD is 11.8. As everything is clear the difference between them is almost 8 units. And the variance of Temperature feature is 467.06 when the standart deviation is 21.6. The difference is std is square root of variance. It means that as mentioned before std helps us to see the output with the same unit type of our data. The last calculation is finding median AD and its value is 7 when simple median is 16. It is clear that the median AD is more than two times smaller than simple median. With this we can say that the median AD is more robust to outliers than simple one.

Conclusion

Summarizing all calculations it is clear that the standard deviation is greater than the deviation mean absolute, which itself is greater than the median absolute deviation.