

Variability:

- Variability (also known as Statistical Dispersion) is another feature of descriptive statistics. Measures of central tendency and variability together comprise of descriptive statistics. Variability shows the spread of a data set around a point.

Measures of Variability

Following are some of the measures of variability that R offers to differentiate between data sets:

- Variance
- Standard Deviation
- Range
- Mean Deviation
- Interquartile Range

1. Variance(σ^2)

- Variance is a measure that shows how far is each value from a particular point, preferably mean value. Mathematically, it is defined as the average of squared differences from the mean value.

Formula For Variance:

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n}$$

Where

- σ^2 specifies variance of the data set
- x_i specifies i^{th} value in data set
- μ specifies the mean of data set
- n specifies total number of observations

Implement Variance Using Python

In [1]:

```
# Python code to demonstrate the working of
# variance() function of Statistics Module

# Importing Statistics module
import statistics

# Creating a sample of data
sample = [2.74, 1.23, 2.63, 2.22, 3, 1.98]

# Prints variance of the sample set

# Function will automatically calculate
print("Variance of sample set is % s"%(statistics.variance(sample)))
```

Variance of sample set is 0.40924

Applications :

- Variance is a very important tool in Statistics and handling huge amounts of data. Like, when the sample mean is unknown then variance is used as biased estimator.
- Real world observations like the value of increase and decrease of all shares of a company throughout the day cannot be all sets of possible observations. As such, variance is calculated from a finite set of data, although it won't match when calculated taking the whole population into consideration, but still it will give the user an estimate which is enough to chalk out other calculations.

2. Standard Deviation (σ)

- Standard deviation in statistics measures the spreadness of data values with respect to mean and mathematically, is calculated as square root of variance.
- Basically It is the Root Over of Variance.
- Standard Deviation is a measure of spread in Statistics. It is used to quantify the measure of spread, variation of a set of data values. It is very much similar to variance, gives the measure of deviation whereas variance provides the squared value.

Formula :

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n}}$$

Where

- σ specifies standard deviation of the data set
- x_i specifies i^{th} value in data set
- μ specifies the mean of data set
- n specifies total number of observations

Implement Standard Deviation Using Python

In [2]:

```
# Python code to demonstrate stdev() function

# importing Statistics module
import statistics

# creating a simple data - set
sample = [1, 2, 3, 4, 5]

# Prints standard deviation
# xbar is set to default value of 1
print("Standard Deviation of sample is % s" % (statistics.stdev(sample)))
```

Standard Deviation of sample is 1.5811388300841898

Applications :

- Standard Deviation is highly essential in the field of statistical maths and statistical study.
- It is commonly used to measure confidence in statistical calculations. For example, the margin of error in calculating marks of an exam is determined by calculating the expected standard deviation in the results if the same exam were to be conducted multiple times.
- It is very useful in the field of financial studies as well as it helps to determine the margin of profit and loss. The standard deviation is also important, where the standard deviation on the rate of return on an investment is a measure of the volatility of the investment.

3. Range

- Range is the difference between maximum and minimum value of a data set.
- range() is commonly used in for looping hence, knowledge of same is key aspect when dealing with any kind of Python code. Most common use of range() function in Python is to iterate sequence type (List, string etc..) with for and while loop.

Python range() Basics :

- In simple terms, range() allows user to generate a series of numbers within a given range. Depending on how many arguments user is passing to the function, user can decide where that series of numbers will begin and end as well as how big the difference will be between one number and the next. range() takes mainly three arguments.
 1. start: integer starting from which the sequence of integers is to be returned
 2. stop: integer before which the sequence of integers is to be returned. The range of integers end at stop – 1.
 3. step: integer value which determines the increment between each integer in the sequence

For Know More Refer : <https://www.geeksforgeeks.org/python-range-function/>
(<https://www.geeksforgeeks.org/python-range-function/>)

4. Mean Deviation (MD)

- Mean deviation is a measure calculated by taking an average of the arithmetic mean of the absolute difference of each value from the central value. Central value can be mean, median, or mode.
- Pandas provide a method to make Calculation of MAD (Mean Absolute Deviation) very easy. MAD is defined as average distance between each value and mean.

Formula :

$$MD \equiv \frac{1}{n} \sum_{i=1}^n |x_i - \mu|$$

Where

- x_i specifies i th value in data set
- μ specifies the mean of data set
- n specifies total number of observations

Implementation Using Python

In [3]:

```
# importing pandas module
import pandas as pd

# importing numpy module
import numpy as np

# creating list
list = [5, 5, 8, 12, 15, 16]

# creating series
series = pd.Series(list)

# calling .mad() method
print("Mean Deviation Of This List Is: ", series.mad())
```

Mean Deviation Of This List Is: 4.166666666666667

5. Interquartile Range (IQR)

- Before going to IQR first we have to know what is a quartile :

Quartiles :

- A quartile is a type of quantile.
- The first quartile (Q1)- is defined as the middle number between the smallest number and the median of the data set
- the second quartile (Q2) – median of the given data set
- the third quartile (Q3)- is the middle number between the median and the largest value of the data set.

Interquartile Range :

- The interquartile range (IQR) is the difference between the 75th and 25th percentile of the data. It is a measure of the dispersion similar to standard deviation or variance, but is much more robust against outliers
- The interquartile range (IQR), also called as midspread or middle 50%, or technically H-spread is the difference between the third quartile (Q3) and the first quartile (Q1).
- It covers the center of the distribution and contains 50% of the observations.
- Mathematically, the interquartile range is depicted as: $IQR = Q3 - Q1$

Where

Q3 specifies the median of n largest values

Q1 specifies the median of n smallest values

Implementation Using Python

- Stats Module provide to find IQR Of a data set.

In [4]:

```
from scipy.stats import iqr  
  
data = [5, 5, 8, 12, 15, 16]  
  
iqr(data)
```

Out[4]:

8.5