Reading: Define and Demonstrate Descriptive Statistics

Bookmark this page

After completing this topic, you will be having knowledge about:

1. **Mean**
2. **Median**
3. **Mode**
4. **Standard Deviation**

1. Mean:

- The mean is the simple mathematical **average** of a set of two or more numbers
- The mean is the most common measure of the central tendency of a set of points. However, the mean is very **sensitive to outliers**.
- Mean can only be used with **numeric data**.

In Excel -> It can be computed by **Average ()**.

In R and Python **->mean ()**.

Example 1

Let's continue with example 1:

Mean =17.41

(19+4+33+2+51+32+2+41+18+2+4+1)/12

Example 2

Python offers more than one way to calculate the mean of values. In the examples below, you will see two of the most commonly used ways of calculating mean in Python.

In [1]:
```
from statistics import mean

list1 = [19,4,33,2,51,32,2,41,18,2,4,1]

mean(list1)
```

Out[1]: 17.416666666666668

In [2]:
```
import numpy as np

list1 = [19,4,33,2,51,32,2,41,18,2,4,1]

np.mean(list1)
```

Out[2]: 17.416666666666668

2. Median:

- The middle number of a set of numbers is called the **median**. It is found by ordering all data points and picking out the one in the middle (or if there are two middle numbers, taking the mean of those two numbers).
- It may be thought of as the "middle" value of a data set.

Let's continue with **example 1**:

Arrange data in increasing order 1,2,2,2,4,4,18,19,32,33,41,51

As m is even, take an average of 2 middle numbers

Median = 11 (Calculate i.e. (4+18)/2)

In the examples below, you will see the two of the packages we saw earlier for calculating mean, can be used to calculate median, as well, in Python.

In [1]:
```
from statistics import median

list1 = [19,4,33,2,51,32,2,41,18,2,4,1,100]

median(list1)
```

Out[1]: 18

In [2]:
```
import numpy as np

list1 = [19,4,33,2,51,32,2,41,18,2,4,1,100]

np.median(list1)
```

Out[2]: 18.0

3. Mode:

- The frequency of an attribute value is the number of times the value occurs in the data set.
- It is found by collecting and organizing the data in order to count the frequency of each result.
- The **mode** is the most frequent number—that is, the number that occurs the highest number of times.
- The notions of frequency and mode are typically used with categorical data, but it can be used on any data type.

Example 1

Let's continue with example 1:

The dataset is 19,4,33,**2**,51,32,**2**,41,18,2,4,1 and mode is 2 as 2 occurs the most number of times in this set.

Example 2

Let's look at an example of finding the mode for both numeric and categorical data using Python.

In [1]:
```
from statistics import mode

list1 = [19,4,33,2,51,32,2,41,18,2,4,1,100]

mode(list1)
```

Out[1]: 2

In [2]:
```
list1 = ["Bangalore","Gurgaon","Bangalore","Chennai","Chennai","Mumbai",
         "Delhi","Cochin","Kolkata","Chennai"]

mode(list1)
```

Out[2]: 'Chennai'

4. Standard Deviation:

- It is a measure of the dispersion of data.

- It measures how much the members of a collection of numbers vary from the mean of the collection.
- The formula to calculate the standard deviation is :

$$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{N}}$$

σ is the standard symbol used for calculating standard deviation. Σ represents the summation. x represents each element in the collection. μ represents the mean of the numbers. N is the count of numbers in the collection. Sometimes, it is also calculated with the denominator as N-1.

Example 1

Let's continue with the dataset from example 1:

The dataset is 19,4,33,2,51,32,2,41,18,2,4,1.

The mean is 17.41

The standard deviation is calculated as below.

$$\sqrt{\frac{(19-17.41)^2 + (4-17.41)^2 + (33-17.41)^2 + (2-17.41)^2 + (51-17.41)^2 + (32-17.41)^2 + (2-17.41)^2 + (41-17.41)^2 + (18-17.41)^2 + (2-17.41)^2 + (4-17.41)^2 + (1-17.41)^2}{12}}$$

This returns 17.04.

Example 2

Let's look at an example of finding the standard deviation using Python.

```
In [1]: from statistics import stdev
        import numpy as np

In [2]: list1 = [19,4,33,2,51,32,2,41,18,2,4,1]
        stdev(list1)

Out[2]: 17.799174308394363

In [3]: np.std(list1)

Out[3]: 17.041411978536935
```

In today's modern age of disruption, SkillUp Online is your ideal learning platform that enables you to upskill to the most in-demand technology skills like Data Science, Big Data, Artificial Intelligence, Cloud, Front-End Development, DevOps & many more. In your journey of evolution as a technologist, SkillUp Online helps you work smarter, get to your career goals faster and create an exciting technology led future.

Corporate

- Home
- About Us

- Enterprise
- Blog
- Press

Support

- Contact us
- Terms of Service
- Privacy Policy