Application Of Mean,median And Mode

- Basically We Use This for find Skewness and kurtosis Of data

1. Skewness

- Skewness is a statistical numerical method to measure the asymmetry of the distribution or data set. It tells about the position of the majority of data values in the distribution around the mean value.

Formula For Skewness Calculate:

$$\gamma_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^{3/2}}$$

OR

$$\textbf{Skewness} = \frac{3\,(\textit{Mean} - \textit{Median})}{\textit{Standard Deviation}}$$

γ1 represents coefficient of skewness

xi represents ith value in data vector

xbar represents mean of data vector

n represents total number of observations

There exist 3 types of skewness values on the basis of which asymmetry of the graph is decided. These are as follows:

1. Positive Skew(Right Skew)
2. Negative Skew(Left Skew)
3. Normal DIstribution

```
skewness = 0 : normally distributed.
skewness > 0 : more weight in the left tail of the distribution.
skewness < 0 : more weight in the right tail of the distribution.
```

1. Positive Skew (Right Skew)

- If the coefficient of skewness is greater than 0 i.e. γ1>0, then the graph is said to be positively skewed with the majority of data values less than mean. Most of the values are concentrated on the left side of the graph.

Here Mainly Mean > Median

- when Average of data is greate than middle Value then This occure

Example

In [1]:

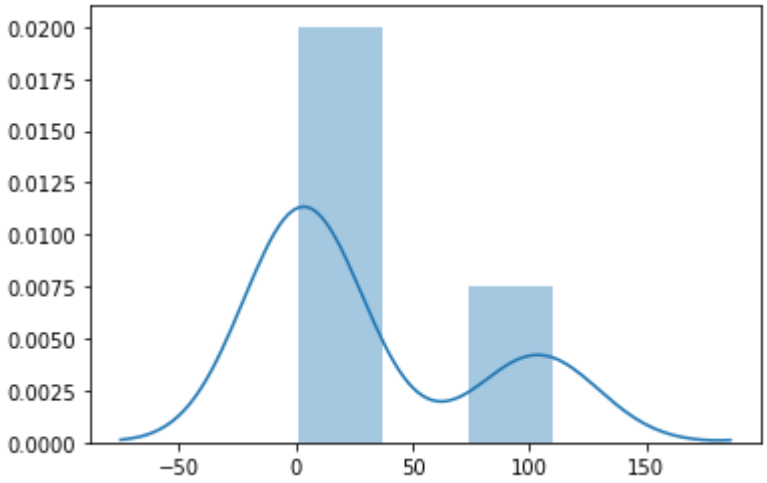
```
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import skew
from statistics import mean,median,mode,stdev
```

In [2]:

```
a = [1,1,2,2,3,4,8,6,100,101,110]
sns.distplot(a)
```

Out[2]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x295ebb714c8>
```



In [3]:

```
print("mean is: ",mean(a))
print("Median is: ",median(a))
```

```
mean is:  30.727272727272727
Median is:  4
```

In The above you clearly see that Mean > Median So it is Positive Skew Or Right Skew

In [4]:

```
print(skew(a))
```

```
1.0234453599856792
```

Here You see the Value of skew is +1 so it is positive skew

2. Negative Skew (Left Skew)

- If the coefficient of skewness is less than 0 i.e. γ1<0, then the graph is said to be negatively skewed with the majority of data values greater than mean. Most of the values are concentrated on the right side of the graph.

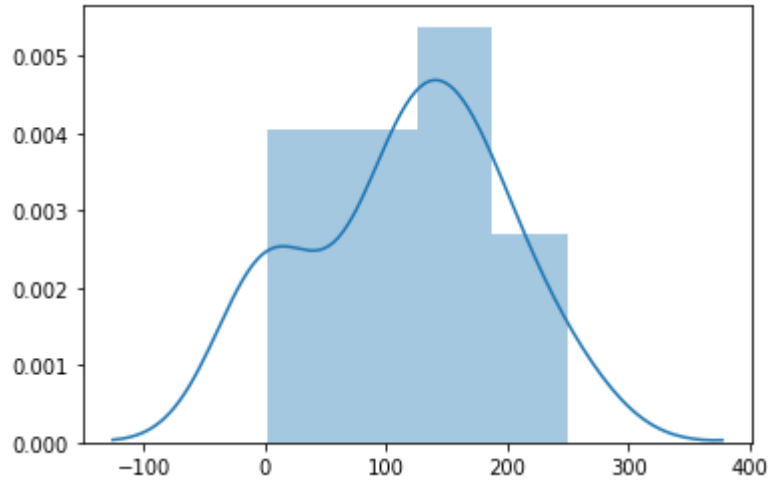
Here Mainly Median > Mean

In [5]:

```
b = [2,3,4,100,101,110,200,250,160,140,150,180]
sns.distplot(b)
```

Out[5]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x295ec2c5e08>
```



In [6]:

```
print("mean is: ",mean(b))
print("Median is: ",median(b))
```

```
mean is:  116.66666666666667
Median is:  125.0
```

Here you Clearly see Median > Mean So it is Negative skew

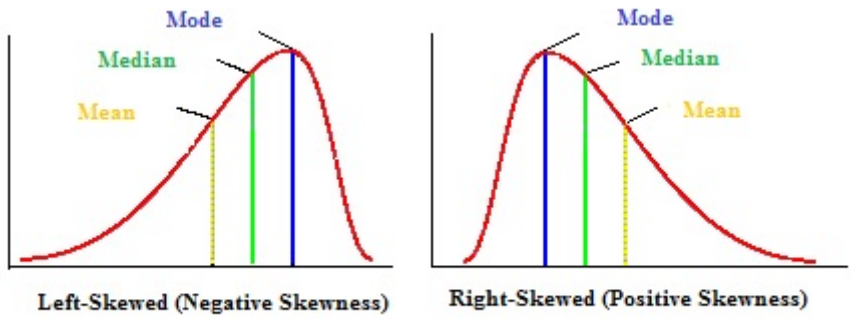
In [7]:

```
print(skew(b))
```

```
-0.1954666401997703
```

Here skew value -ve so it is negative

Both Skew



Now a question ariese why this type of skewness we see in graph ?

Ans - This skewness is cause bacause of outlier. so when your data contain outlier value then this type of skew is present. so after that we nedd to resolve the outlier by using various outlier treatment mechanism like apply Log value or aaply z-score and many more.

3.Normal Distribution (No Skew)

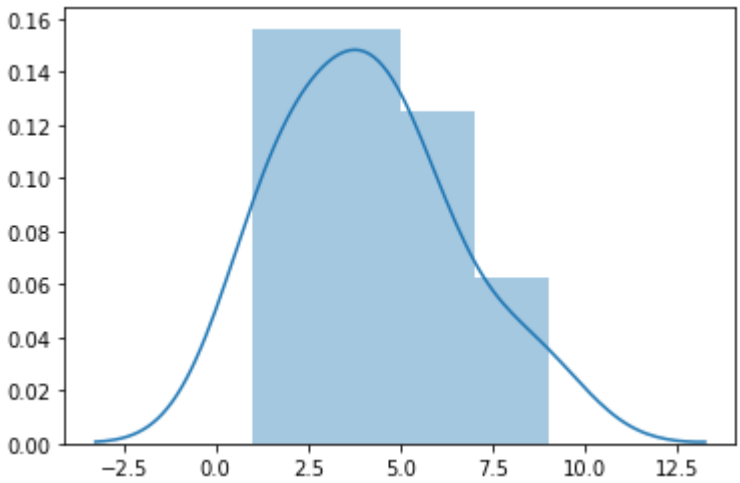
- See Here No outlier present so this graph look like gaussian distribution or normal distribution. we nedd to always make graph normal for machine learning model
- If the coefficient of skewness is equal to 0 or approximately close to 0 i.e. γ1=0, then the graph is said to be symmetric and data is normally distributed.

In [8]:

```
c=[1,1,2,3,4,5,2,4,6,9,8,2,4,5,6,4]
sns.distplot(c)
```

Out[8]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x295ec395788>
```



In [9]:

```
print("mean is: ",mean(c))
print("Median is: ",median(c))
```

```
mean is:  4.125
Median is:  4.0
```

See here Mean and Median Value Almost Same

In [10]:

```
print(skew(c))
```

```
0.5266195519509043
```

And Skew Value is 0 So This is the graph we nedd for further model build