Discriptive Statistics

1. Measure of central tendency.
2. Measure of dispersion(spread).

1. Measure of central tendency: mean,median,mode

In [1]:
```
import pandas as pd
import numpy as np
```

In [2]:
```
x=np.random.randint(20,25,10)
print(x)
```

```
[22 24 24 22 20 24 24 24 23 21]
```

In [4]:
```
np.mean(x)
```

Out[4]:
```
22.8
```

In [5]:
```
np.median(x)
```

Out[5]:
```
23.5
```

In [9]:
```
df=pd.DataFrame(x)
df
```

Out[9]:

	0
0	22
1	24
2	24
3	22
4	20
5	24
6	24
7	24
8	23
9	21

In [11]:
```
df.mode()   # mode dosen't work in numpy bt possible in pandas.
```

Out[11]:

	0
0	24

1. Measure of dispersion(spread)

This is the standard deviation or varience.
e.g RCB average score is 185 so measure of dispersion is +/-15, this is the measure of dispersion.

variance=summation of each of x with square of mean(mue) indidually with every x divisible by no of elements.
standard devaition=sqrt of variance.

In [14]:
```
print(np.var(x))   #variance of x
```

```
1.9600000000000002
```

In [15]:
```
np.std(x)         # standard daviation is the sqrt of variance
```

Out[15]:
```
1.4000000000000001
```

In [17]:
```
y=[10,5,30,50,45,65,80,90,40,70]
y_=pd.DataFrame(y)
```

In [18]:
```
y_.describe()
```

Out[18]:

	0
count	10.000000
mean	48.500000
std	28.387204
min	5.000000
25%	32.500000
50%	47.500000
75%	68.750000
max	90.000000

In [19]:
```
# Inter Quartile => Q3-Q1
# Quartile Range => max-min
# IQRange
```

Distribution

Normal Distribution

Skewed Right: lets take marks of the students-if majority of marks in between 30-70 then graph is bell shape more towards left n less towards right.

Symmetric Distribution: majority of marks in 50-75 the bell shape more in center.

Skewed Right: less marks in between 10-30 then bell shape more towards in right n less in left.

PDF: Probability Distribution/Density Function
CDF: Commulative Distribution Function

PDF is like a KDE graph and CDF gives the % i.e.want to find out %students got marks more than 80 by observing the graph on x and y by getting the cross point and finding out the area after the line called as CDF.

- x = Variable of Interest, should be continuos variable
- mu = Mean
- sigma = Std Dev
- Asysmptote = x axis
  ```
  This asysmptote is when there is a graph n its ends never touches to x axis that is called as aymptote.
  ```

Imp Note:
When x is given, use cdf function to find p
When p is given, use pdf function to find x

Inferential Statistics

Sampling data and infer the result to ddescribe entire population.

Central Limit Theorem

This theorm is specifiaclly for non normal distribution

Normal Distribution : Distribution is always normal irrespective of sample size.

Non-Normal Distribution : If sample is adequate (appr > 30 sample), distribution starts looking normal.

```
If there a non-normal distribution which means the graph will not be in bell shape it may be of non-shape so this
kind of data can be normal distribution after takin >30sample each time means 4-5 times so, it'll slowly starts
looking like a normal distribution.
```

Formula:

muxbar=mu(mean of sample mean=population mean)
sigmaxbar=sigma/sqrt(n)(population std/sqrt of sample space)

$$\mu_{\bar{x}} = \mu$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

In [ ]:
```
```