**Inferential Statistics**

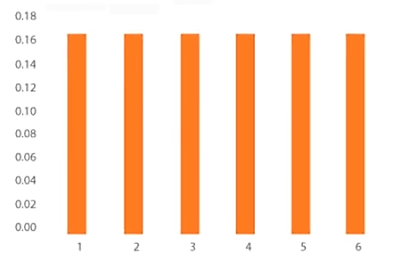
**Open this link and download the file(s)**

**~~https://drive.google.com/open?id=1PEubyOUwogSXWxgEhzp708mJmwBrQmIS~~**

**Distribution:**

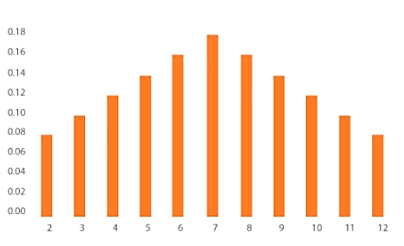
A distribution is a function that shows the possible values for a variable and how often they occur.

->Rolling 1 dice (Probability of 1,2,3,4,5,6 is 1/6 or 0.16 and 0 for every other number)



Discrete Uniform Distribution

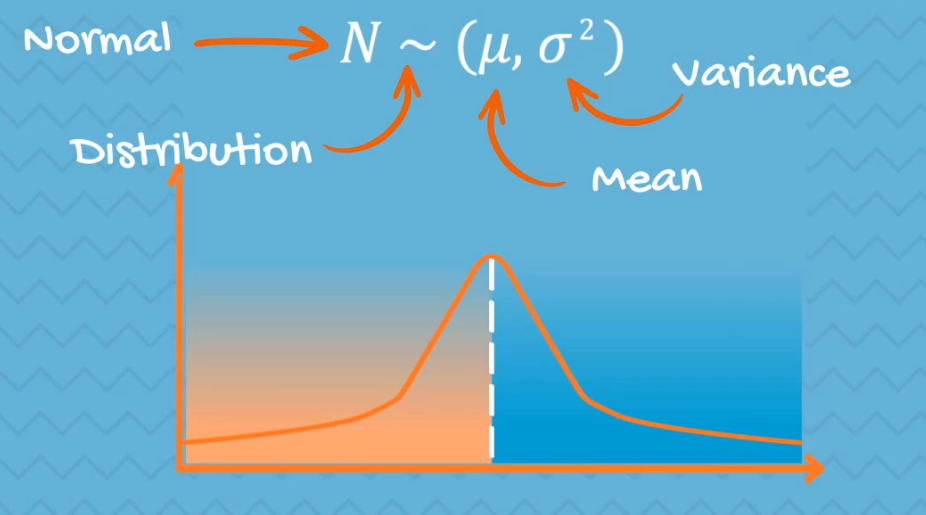
->Rolling 2 dice (Probability of 1 is 0, 2 is 1/36 or 0.003 and so on)



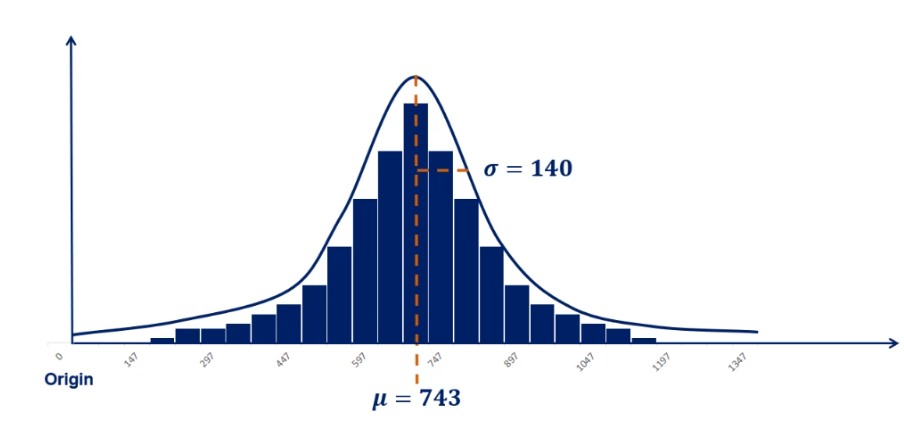
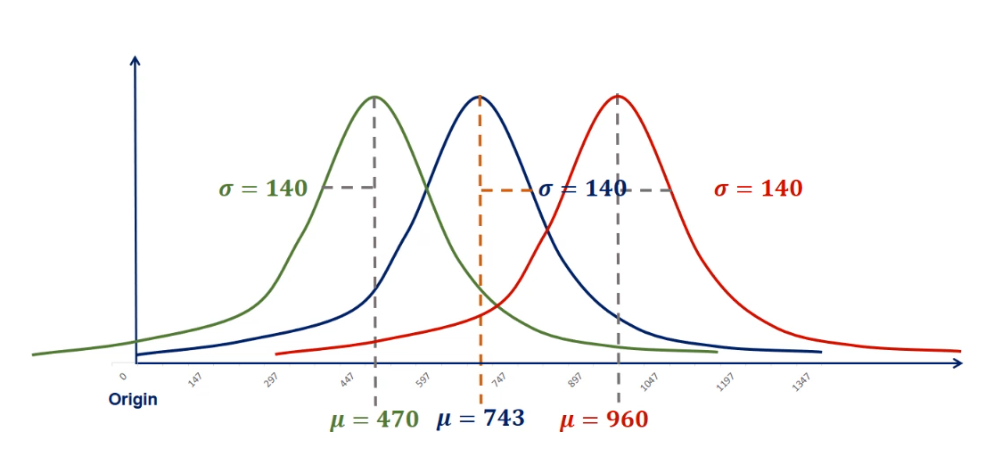
Binomial Distribution

So Distribution shows frequency at which all the possible values occur.

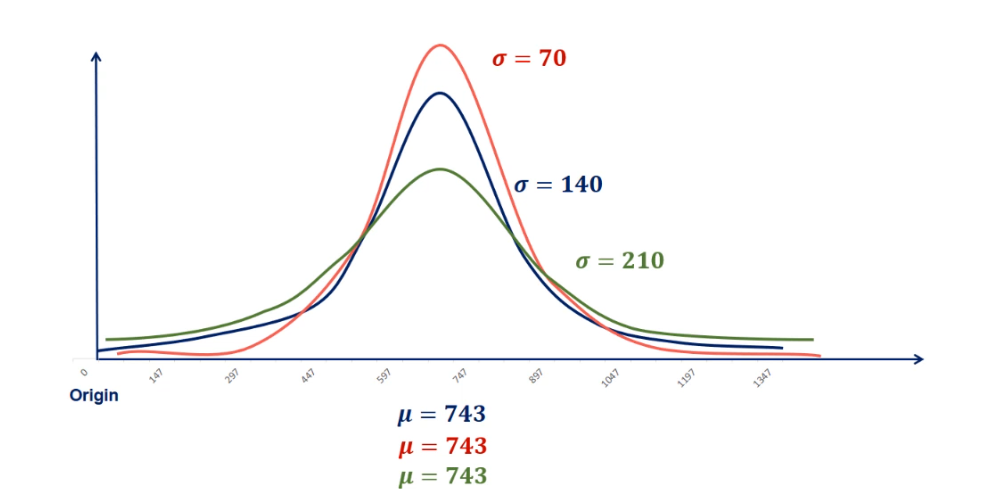
We will mostly use Normal Distribution and Student’s T Distribution.



Normal Distribution (aka Gaussian Distribution or Bell Curve)

Controlling SD



Controlling Mean

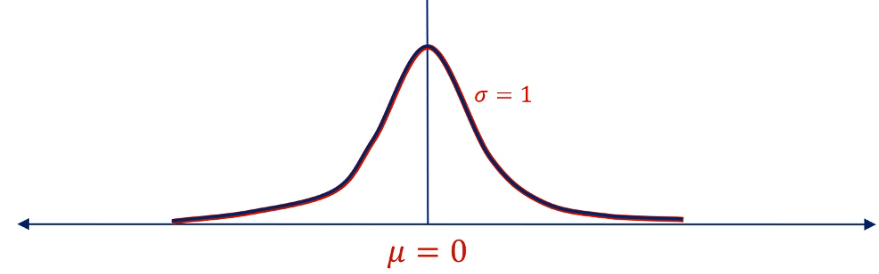
**Standardization:**

The Standard Normal distribution is a particular case of the Normal distribution. It has a mean of 0 and a standard deviation of 1.

Every Normal distribution can be ‘standardized’ using the standardization formula:

It will result in:

𝑁~(0,1)



Standardization

**Central Limit Theorem:**

For understanding CLT, let’s understand Sample Distribution.

Example: Take different samples from an old car market. Every sample has different mean. Take all the means and make a **Sample Distribution** from it. The mean of this sample distribution will be very much closer to the population mean.

*So, no matter what the underlying distribution of the dataset is, the sampling distribution of the means would approximate a normal distribution.*

**Confidence Intervals:**

CI is an interval within which we are confident (with a certain percentage of confidence) the population parameter will fall.

Example: Average meal price in Peshawar is Rs: 300. It’s much safer if we say that the average price is around 250 to 350.

Let’s say for above example we are 95% sure that population parameter falls in this range. As you can’t be 100% confident unless you have gone through whole population. So we introduce 1- α.

**Confidence Level: 1-α**

Where,

Confidence Level = 95% means α= 5% or 0.05. Common values are 90%, 95%, 99%.

Confidence Intervals:

-Population Variance Known

-Population Variance Unknown

**For Population Variance Known:**

Example: Salary of a data scientist.

Population SD = PKR 15000

N=30 salaries.

Formula:

Where, is the standard Error.

Another example of CI -> Age of the class students.