**Central Limit Theorem**

The central limit theorem says that equally-weighted averages of samples from any distribution themselves are normally distributed. Consider the sample mean of Independent and identically distributed (IID) random variables X1, X2 . . . Such that E[Xi] = µ and Var(Xi)= σ^2

* Mathematically, the central limit theorem states:

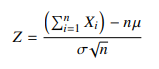


n - sample size

σ^2 - variance

µ - mean

* It is often expressed as a way of obtaining the standard normal, Z:



**The distribution of sample mean:**

* If sample mean is random variable, it follows normal distribution then population follows the normal distribution.
* Sample mean will be approximately normal even if the distribution of data in the population is not normal if the sample size is fairly large.
* Standard deviation:

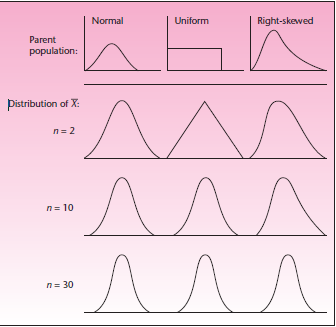


* This is referred as to as standard error of the mean

The standard error of the mean estimates the variation between samples whereas the standard deviation

* Some general rule of thumb telling us when a sample is large enough that we may apply the central limit theorem would be useful.
* Sample of 30 or more elements is considered as **large enough** for the central limit theorem to take effect.
* If sample size is small then the graph may seems to be more spread out and if the sample size is large it may be more steam

**The Effects of the Central Limit Theorem**: The distribution of **X̄** for different  populations and different Sample Sizes

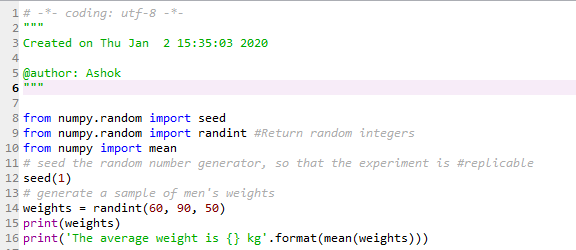


***The History of the Central Limit Theorem:***

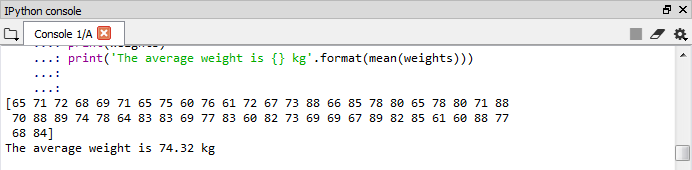
The first such theorem was discussed at the beginning of Normal distribution as the discovery of the normal curve by Abraham De Moivre in 1733. Recall that De Moivre discovered the normal distribution as the *limit* of the binomial distribution. The fact that the normal distribution appears as a limit of the binomial distribution as *n* increases is a form of the central limit theorem. Around the turn of the twentieth century, Liapunov gave a more general form of the central limit theorem, and in 1922 the final form we use in applied statistics was given by Lindeberg. The proof of the necessary condition of the theorem was given in 1935 by W. Feller.

**Python code :**

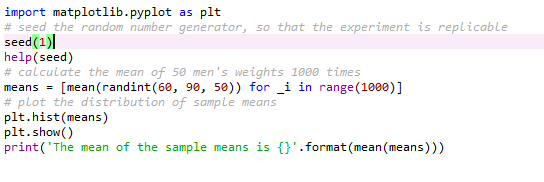
Creating random samples of men’s weight (range between 60 and 90kgs), size=50



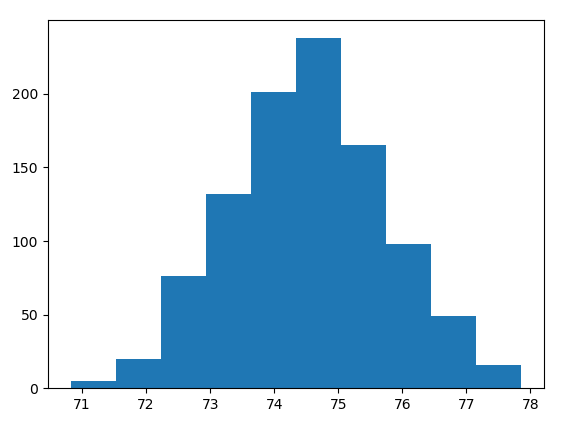
**Output of average sample weight:**



**Repeat the sampling simulation for 1000 times:**



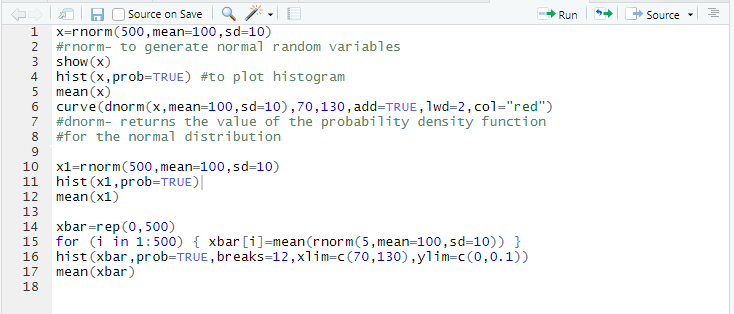
**Output :**

 **the mean of sample mean is 74.54**

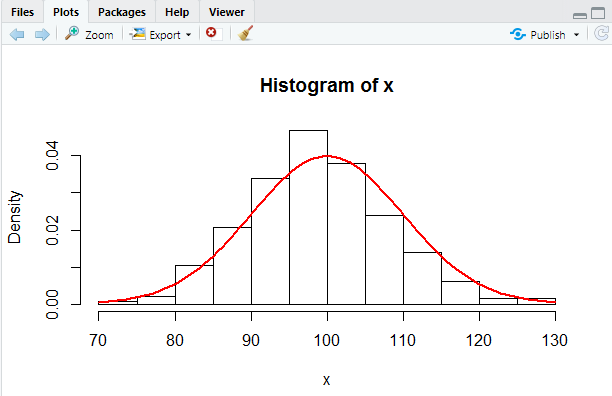
The mean of the sample means (74.54) should be a good estimate of the real parameter (According to the Central limit theorem)

**R-code:**

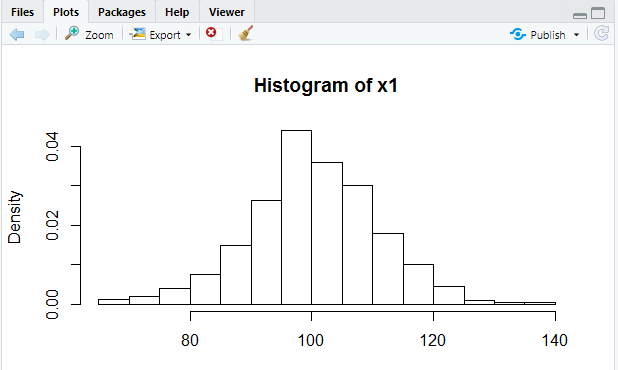
Use rnorm function to draw 500 numbers at random from a normal distribution

dnorm – density function

**Output of sample x:**



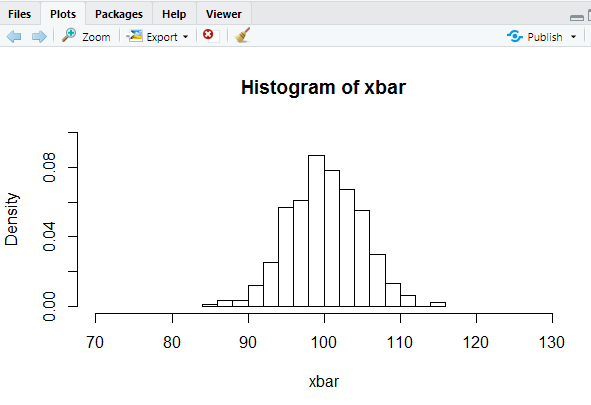
**Output of sample x1:**



**Output of xbar:**

For rep function repeat the entry zero 500 times

As a result, the vector **xbar** now contains 500 entries, each of which is zero



* If you draw samples from a normal distribution, then the distribution of sample means is also normal.
* The mean of the distribution of sample means is identical to the mean of the parent population, the population from which the samples are drawn.
* The higher the sample size that is drawn, the "narrower" will be the spread of the distribution of sample means.