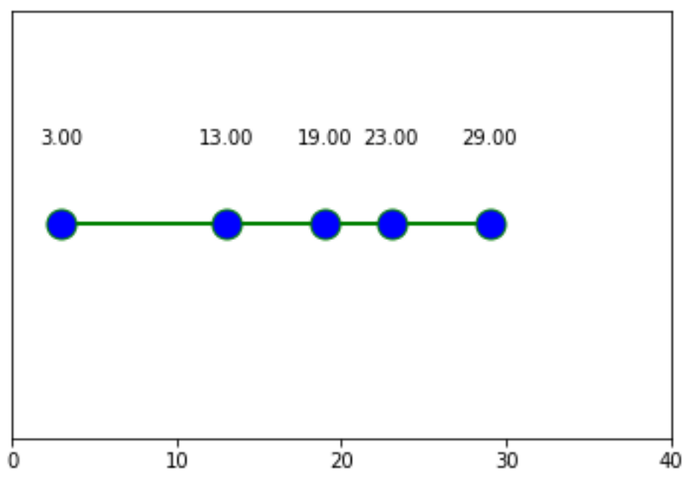
MEAN, VARIANCE, STANDARD DEVIATION

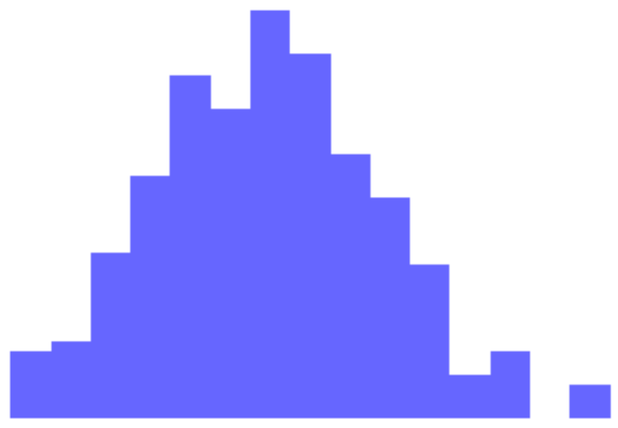
Assume we counted the quantity on T-shirts from 5 different store

1st dot represents only 3 T-shirs were there in store 1 > 2nd dot represents only 13 T-shirs were there in store 2

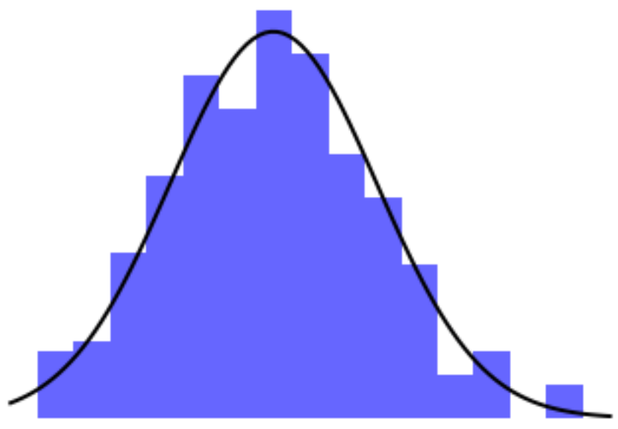
> 3rd dot represents only 19 T-shirs were there in store 3
>
> 4th dot represents only 23 T-shirs were there in store 4
>
> 5th dot represents only 29 T-shirs were there in store 5



same way if we did count for 2 billion stores we can plot a histogram with measurements



And if we want to fit a normal curve to this Histogram like the below picture we need calculate the **Population mean** ,**Population variance** ,**Population Standard Deviation**



Population Mean : To calculate population mean we take the average of all the 2 billion stores

$$Populationmean(mu) = \frac{3 + 13 + 15 + 19 + 23 + \ldots.}{2000000000} = 20$$

we get 20 as population mean and we center the cure on population mean

NOTE : This calculation is performed on whole 2 billion so this will be considered as population mean not estimation mean

We do population mean if we have enough patience else generally we do an estimation Mean with small sample

$\bar{x}$: Estimation mean

μ : $Populationmean$

$\bar{x}$ is different from μ but with more and more data $\bar{x}$ should be closure to μ

Now we need to determine how wide to make the curve by calculating variance and Standard deviation other words we are calculating how data spread around mean

Variance = $\frac{\sum(x-\mu)^2}{n}$

$x - \mu$: we are subtracting Population mean with each measurement

$(x - \mu)^2$: Squaring ti eliminate negative values

$\sum$: here we will be Adding the values after $(x - \mu)^2$

n : to get the average of 2 billion measurements

NOTE : To eleminate negative values we are squaring the values but why not MOD ?

reason for that is when we use MOD we get a SHARP V SHAPED CURVE and since we have sharp edge at minimum value we can't apply derivative then finding the min value is hard with absloute than with square

Assume we get Variance = $\frac{\sum(x-\mu)^2}{n}$ = 100

But what we did is square of each term soo the output 100 are the units T-shirt square, however the units on x-axis is not squared

so to solve this we have to include Square root to variance which is Population Standard Deviation

$$PopulationStandardDeviation(\sigma) = \sqrt{\frac{\sum(x-\mu)^2}{n}} = \sqrt{100} = 10$$

mean(μ) = 20 SD (σ) = 10 and we get the normal curve

NOTE : we never have the population data so we only calculate the Estimation Mean , Estimation Variance , Estimation Standard deviation because we everytime work with small samples

soo rewiting the formulae we get

Variance = $\frac{\sum(x-\bar{x})^2}{n-1}$

n-1 : we are dividing it with n-1 instead of n becuase to compenstate the fact that we are calculating sample mean not population mean