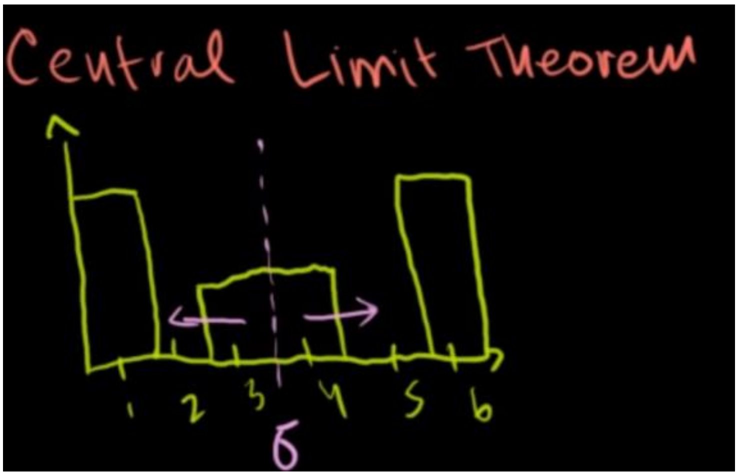
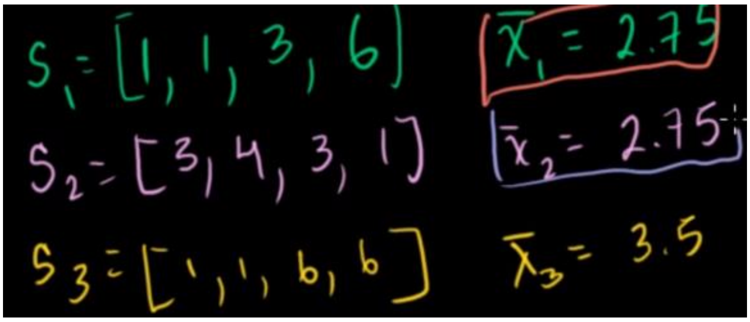
Friday, May 10, 2019 9:58 PM

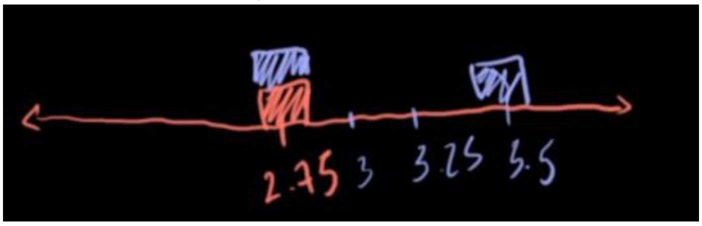
Central limit theorem:



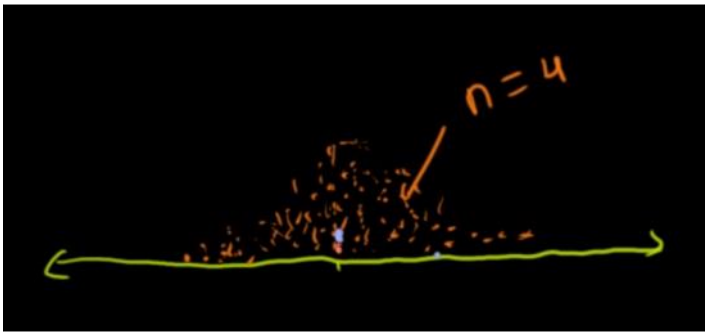
Supposing this is the population of your data points,
Every time you take a sample of size 4
Sample size refers to the number of individual in one sample



And you plot the mean of every sample on another plot.

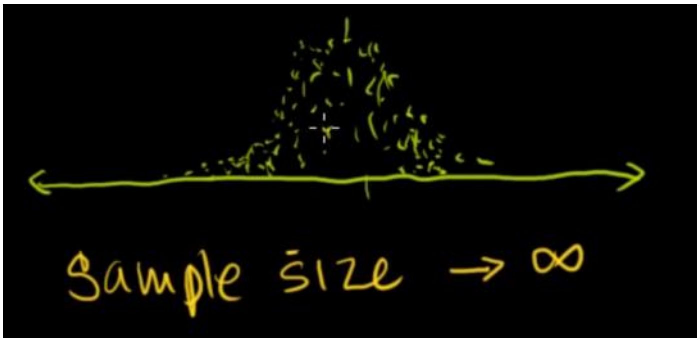


As you take greater and greater samples, you ended up with a plot like this



This is the plot of the mean of your many samples with the size of 4

If you take a larger sample size, you ended up with a more concentrating plot, with less standard deviation



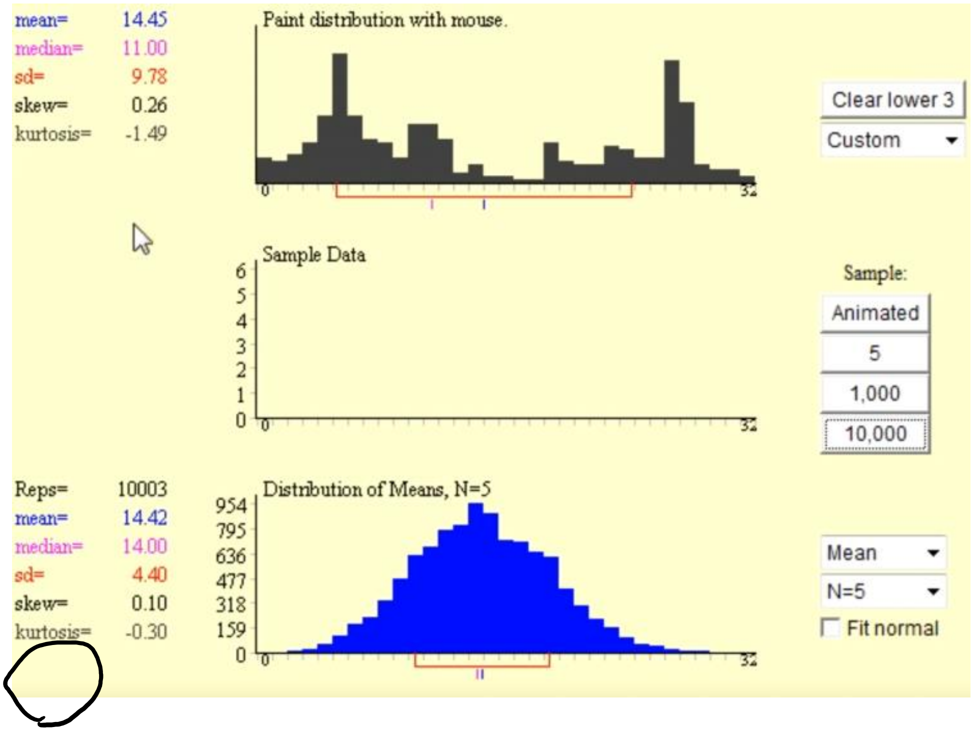
The distribution is approximately a normal distribution, even though the original distribution of population is not normal distribution

This is why normal distribution shows up so much in statistics -- you can always convert a certain distribution to a normal distribution
It is a good estimation of a lot of processes

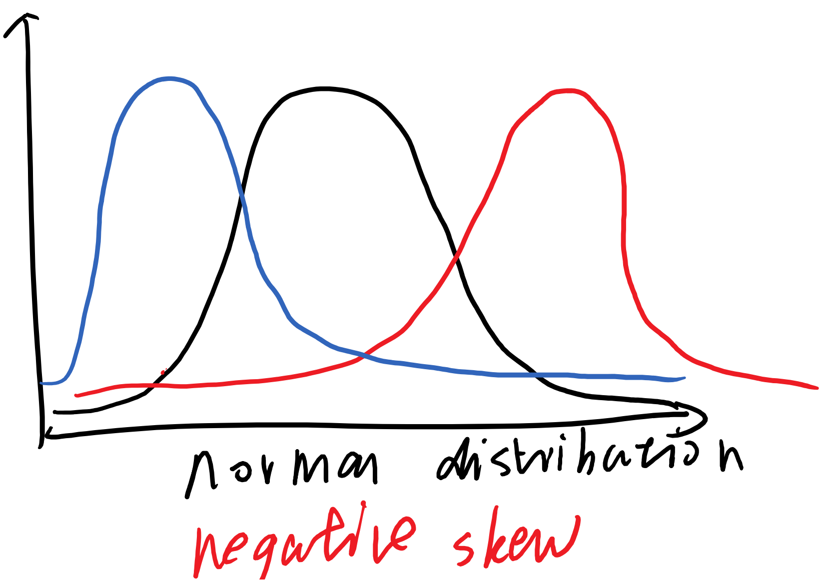
In your original data set, the individual is your observed individual value

But in the normal distributed data set, the individual is the mean of your samples, every sample is composed of n of your observed individuals -- that is the **"sample distribution of the sample mean"**

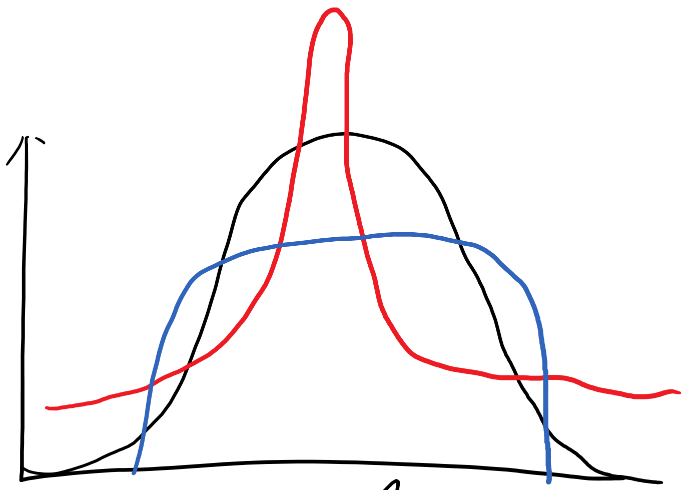
N = 5



Kurtosis and Skew tell how much this distribution is like a normal distribution



negative skew
positive skew

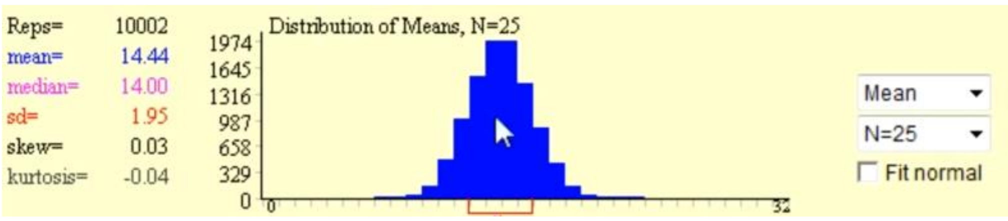


normal

positive kurtosis
negative kurtosis

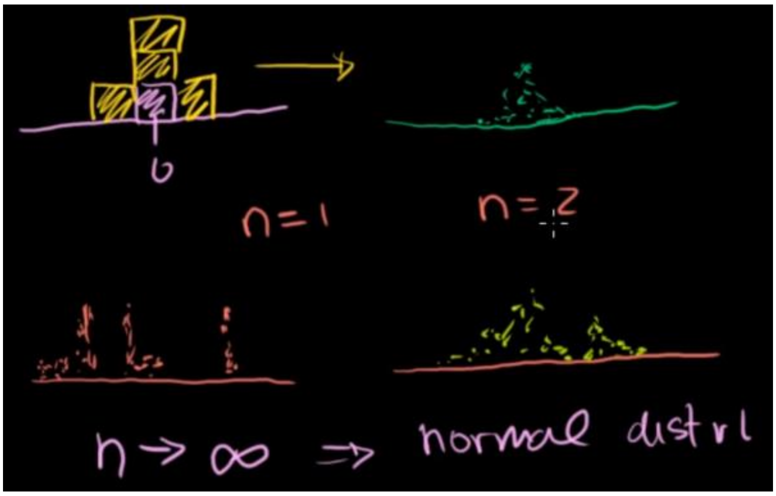
Positive kurtosis: more pointed peak but fatter tail

N = 25

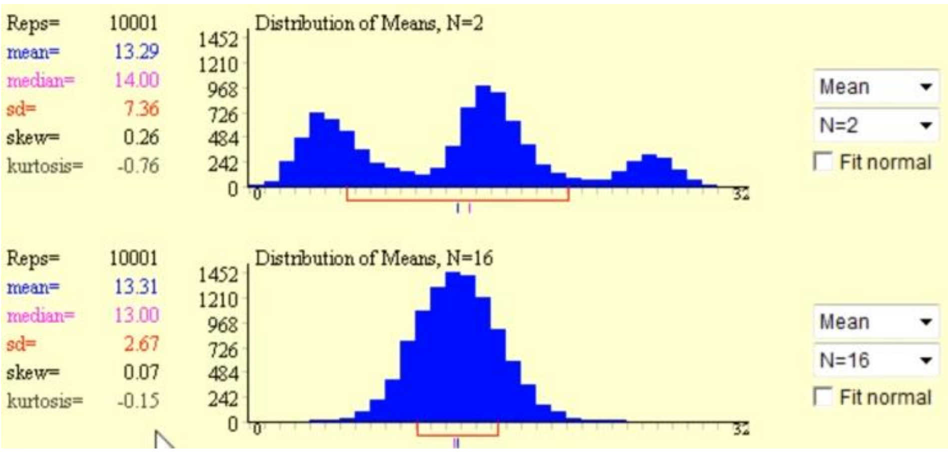


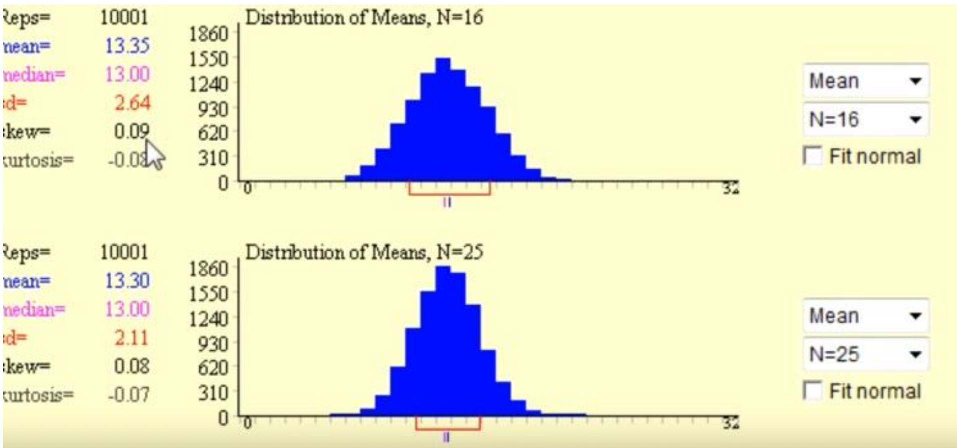
The graph is more normal than N=5

As your sample size N increases:



The mean of your sample distribution of the sample mean is equal to μ





$\sigma_{\bar{x}}^2$ is

The variance of your

standard deviati on

standard deviation on

of your sample mean)

$$= \frac{\sigma^2}{n}$$

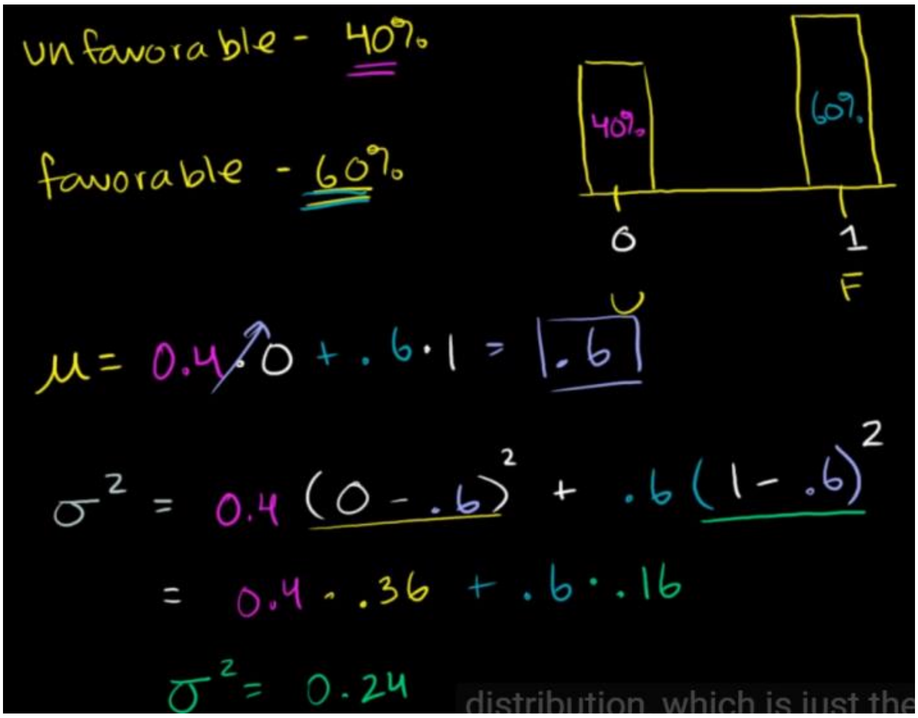
(σ^2 is the standard deviation of the original data set)

$\boxed{\sigma^2}$ n

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$$

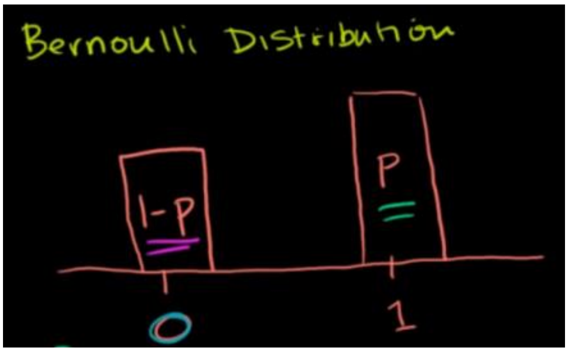
$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

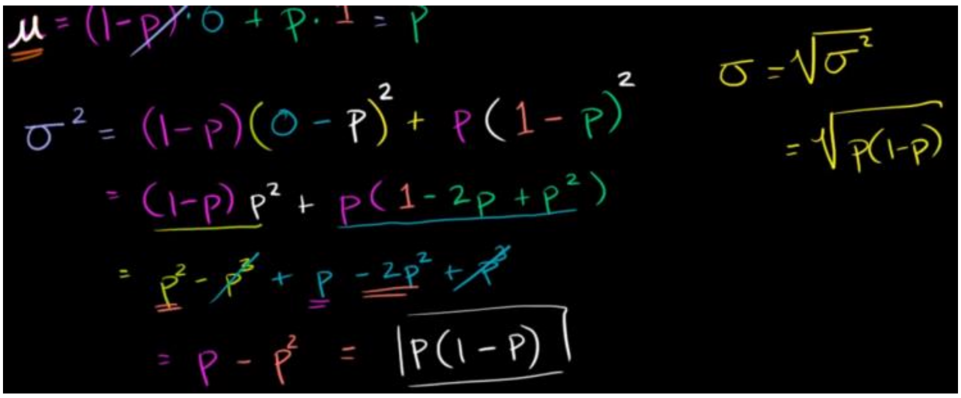
Bernoulli distribution:



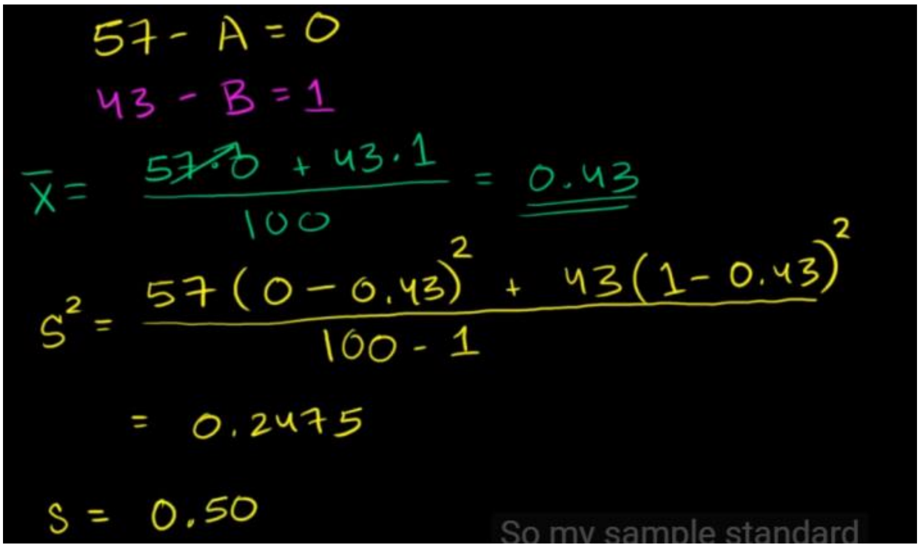
the problem of favor rate of the president
To calculate the possibility of getting elected

The possibility of favorable = p
The possibility of unfavorable = 1-p

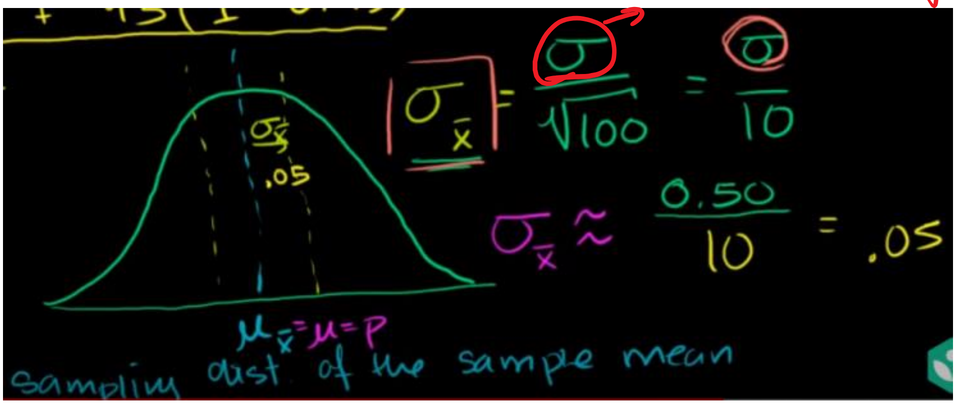




Let's randomly survey 100 people about who do they want to vote for



While the 100 people is just a sample from the population,
In the sample distribution of the sample mean, 0.43 is just a data point. But we are roughly sure the mean of the sdsm is 0.43, now we need to calculate the standard deviation of the sdsm. this is the SD of population



The real value of population SD is impossible to get, but the SD of the randomly surveyed 100 people is a good estimation of the population SD

Find an interval such that "reasonably confident" that there is a 95% chance that the true population mean is

Confidence interval:

2 SD =95.4%