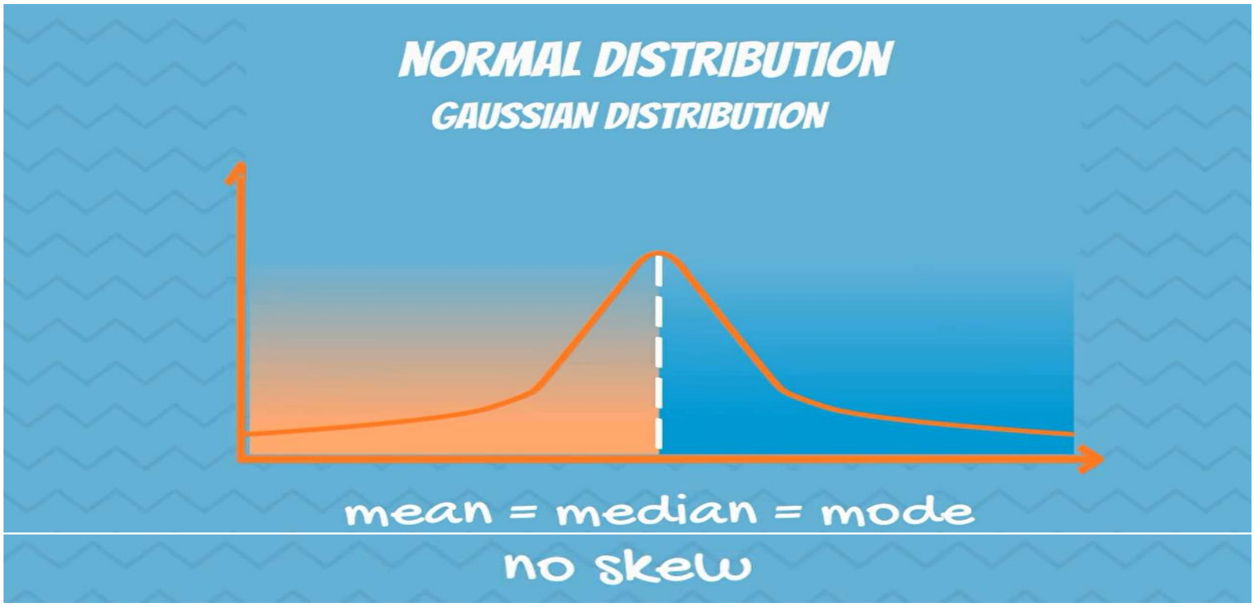
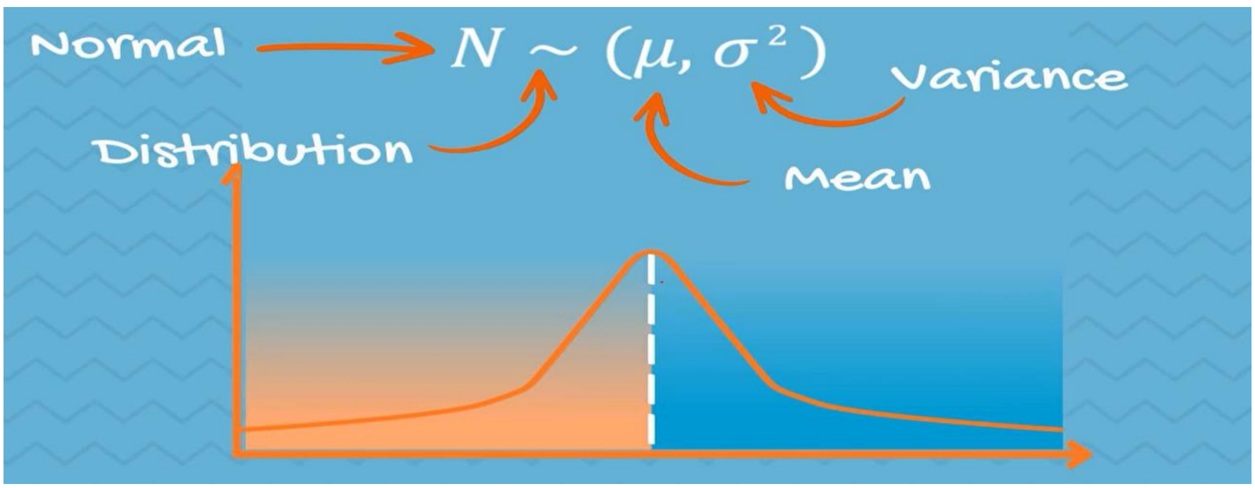
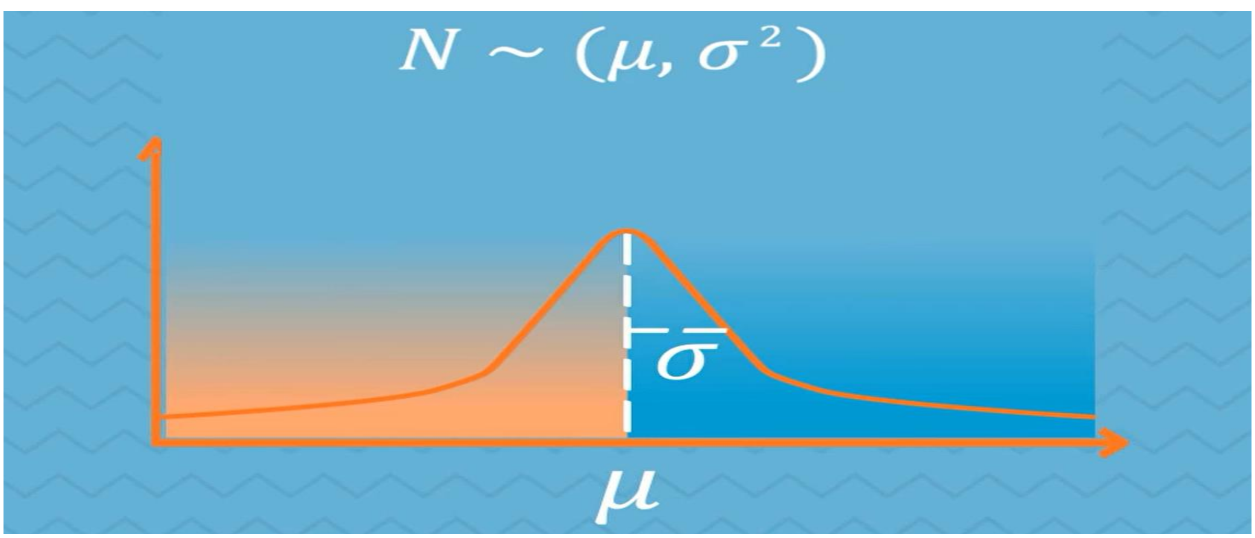
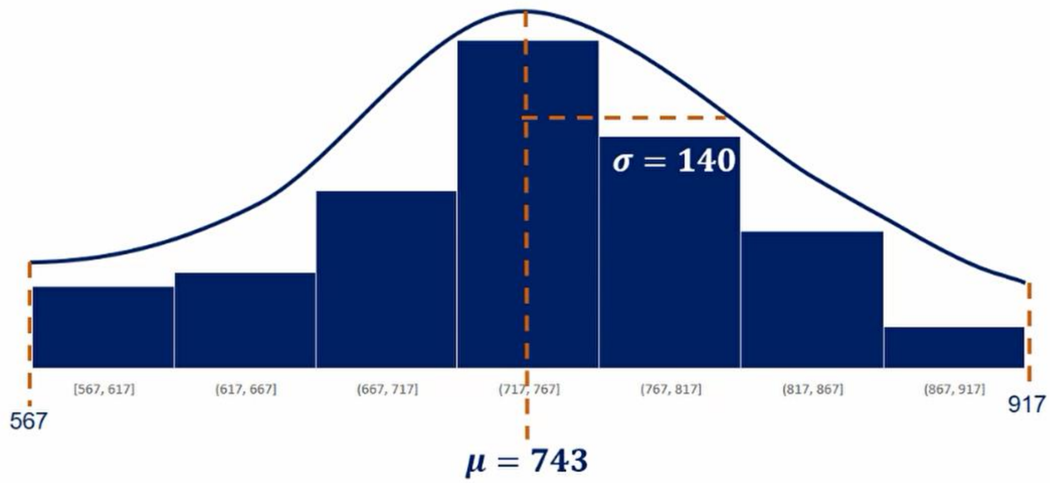
4. The Normal Distribution



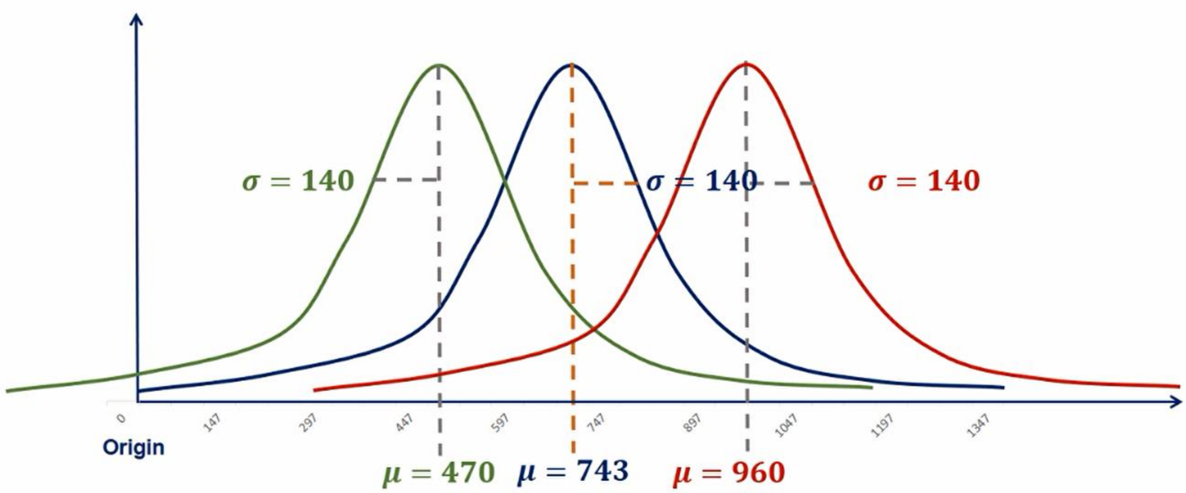




Normal distribution



Normal distribution. Controlling for standard deviation

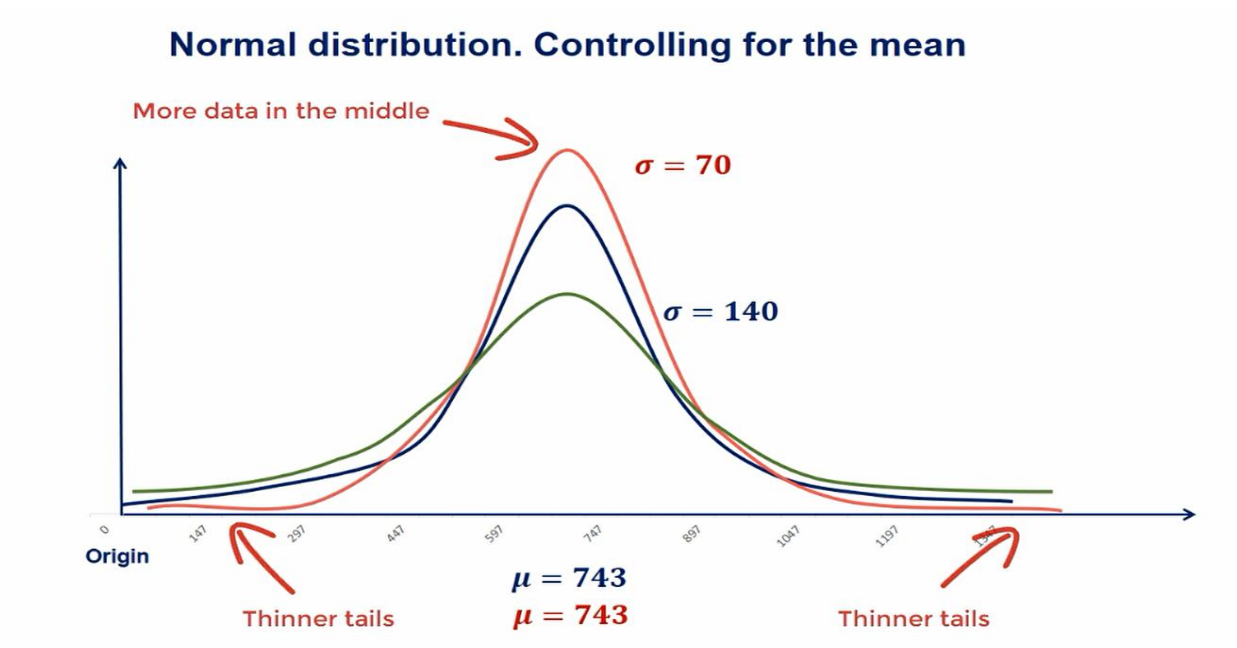


A lower mean would result in the same shape of the distribution, but on the left side of the plane

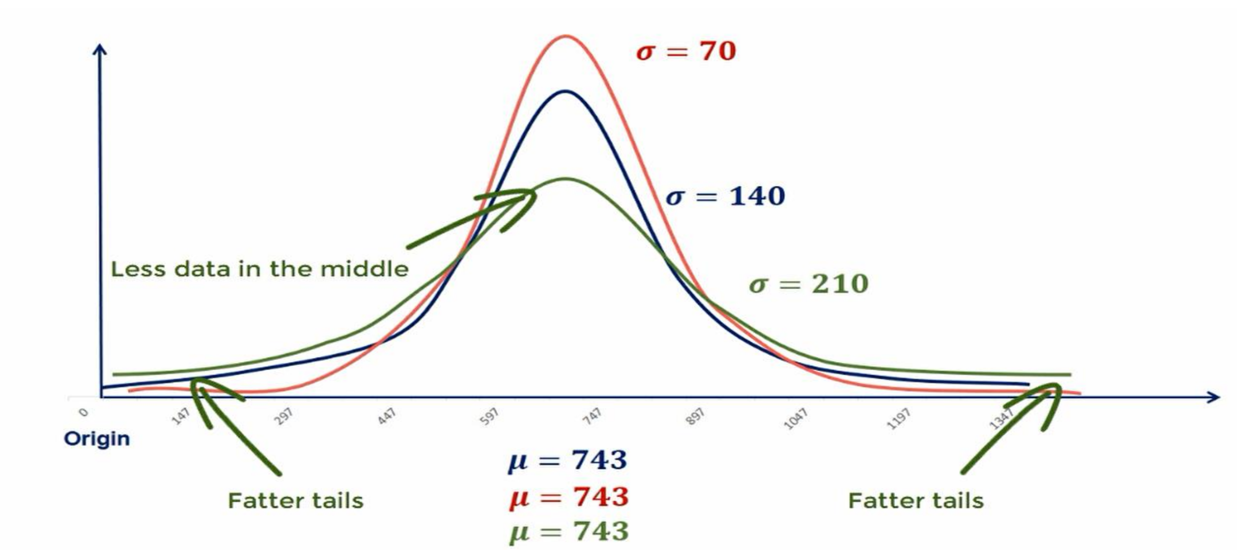
A bigger mean would move the graph to the right

Controlling for the mean and we can change the standard deviation.

The graph is not moving, but rather - reshaping!

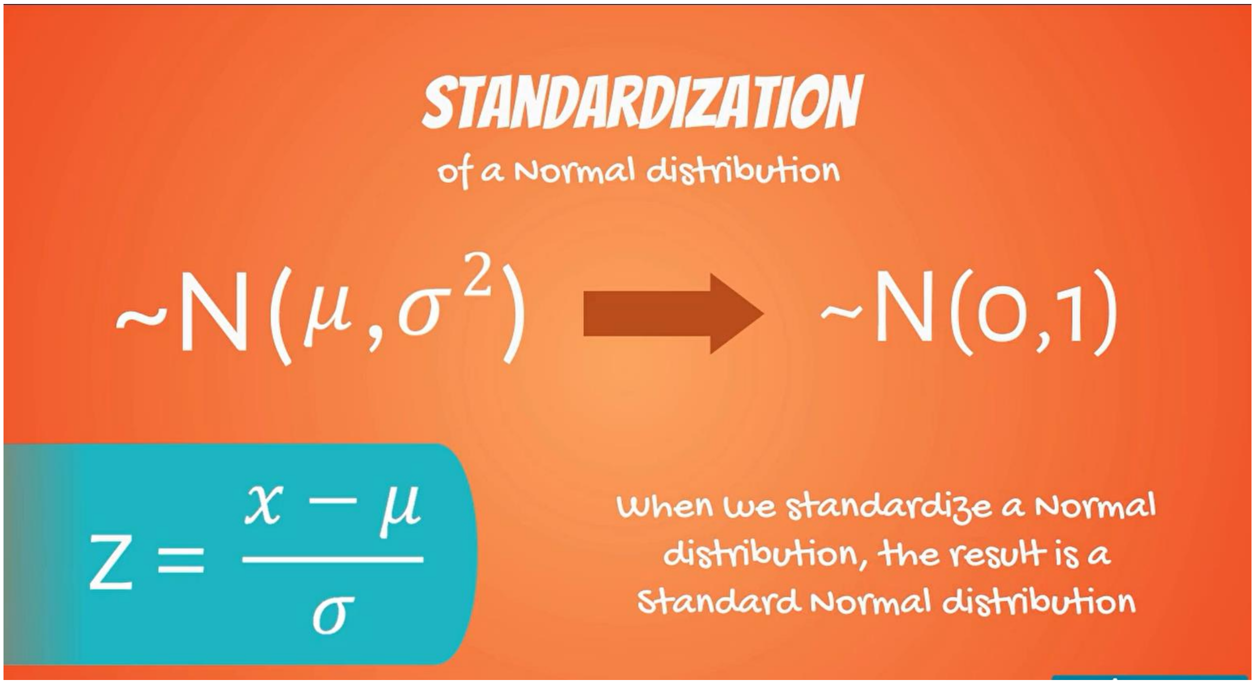


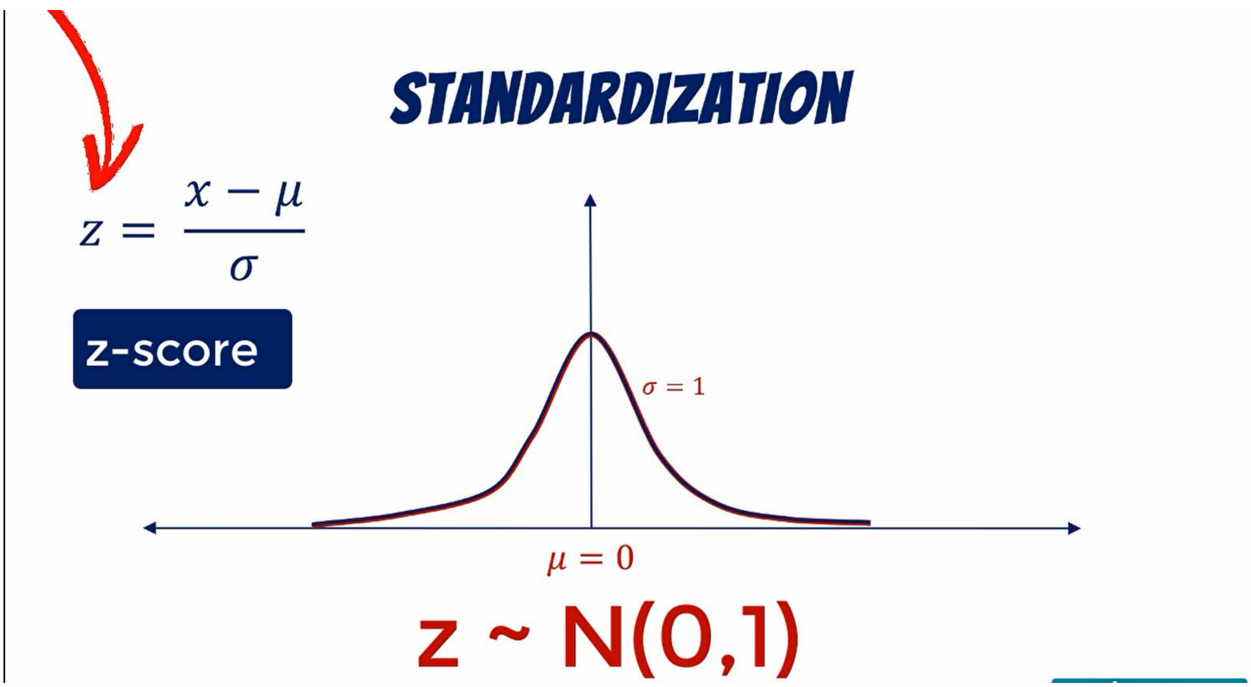
Lower standard deviation result in a lower dispersion, so more data in the middle and thinner tails.

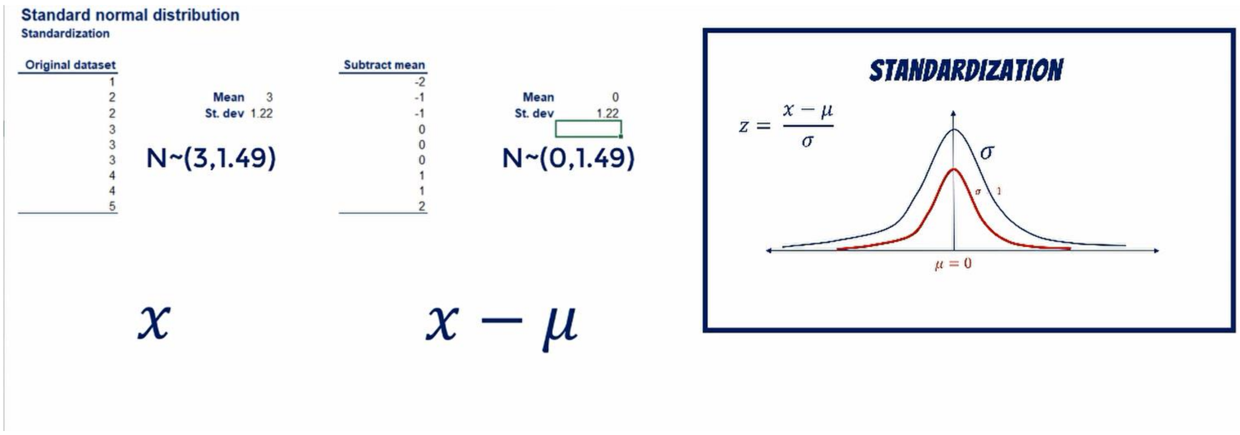


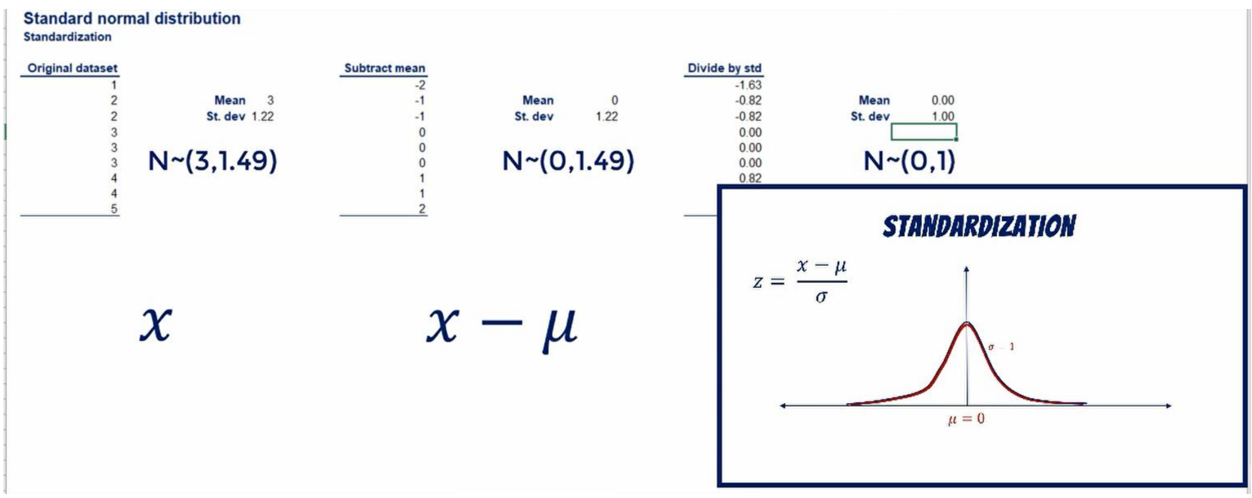
higher standard deviation result in a higher dispersion, so less data in the middle and fatter tails.

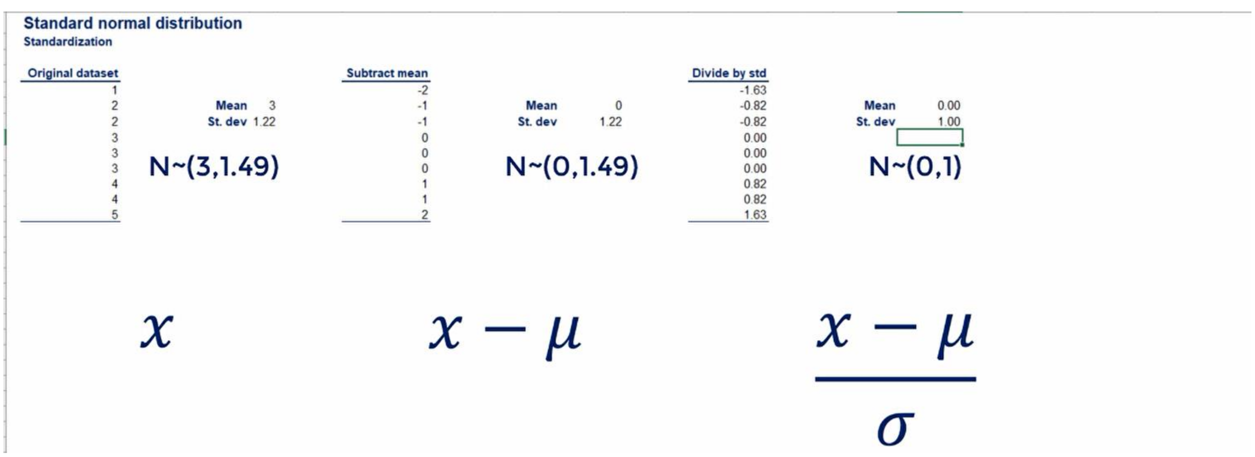
6. The Standard Normal Distribution



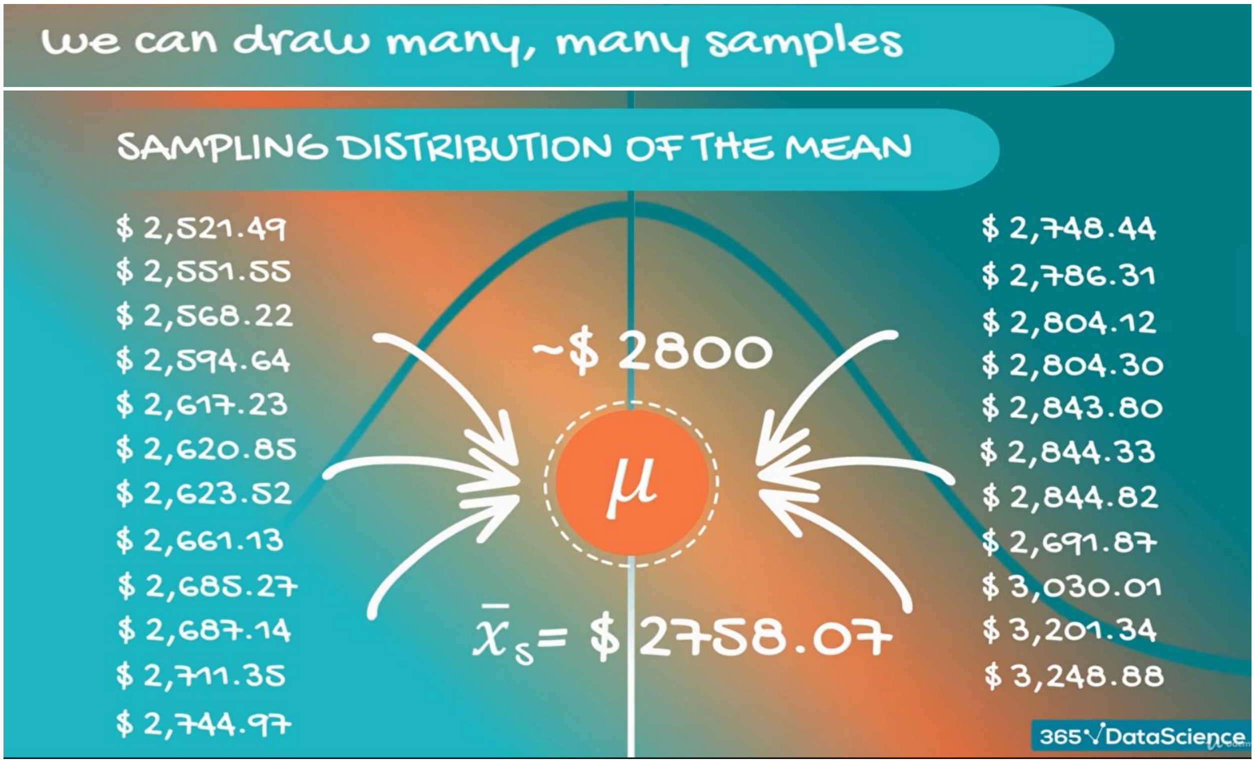


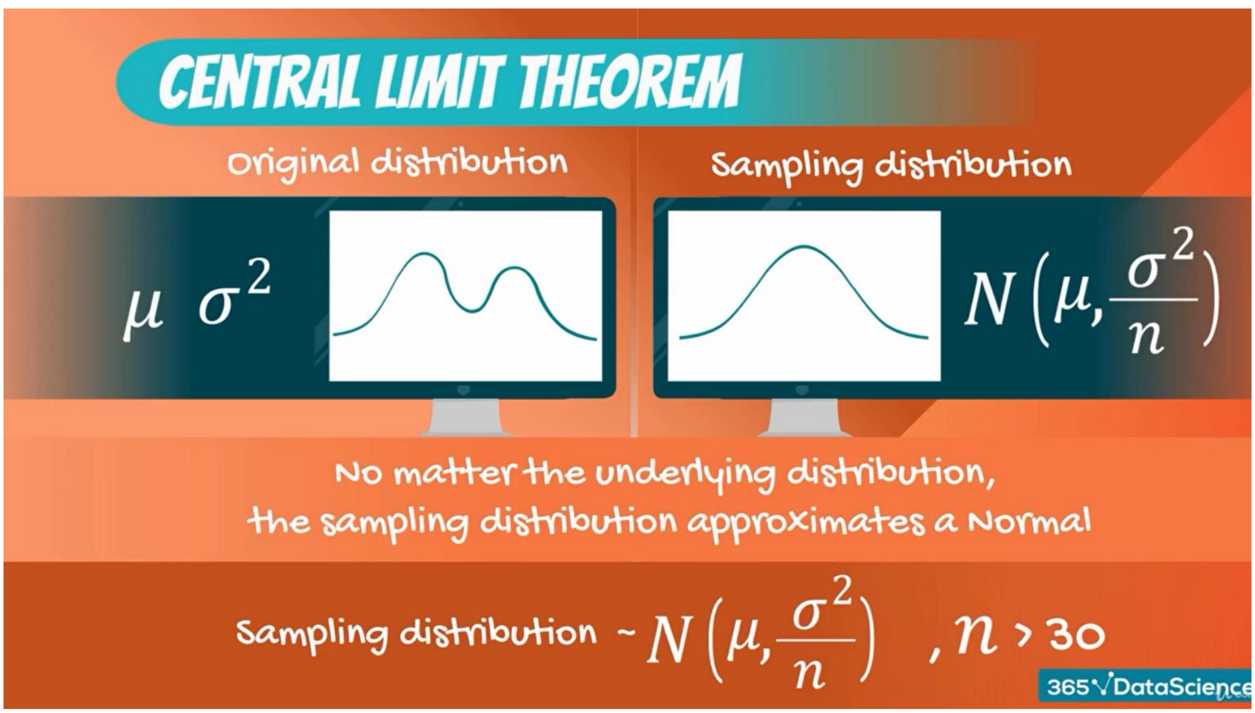

Standard normal distribution
Standardization
Original dataset
1 2 2 3 3 3 4 4 5
Mean 3
St. dev 1.22
N~(3,1.49)
Subtract mean
-2 -1 -1 0 0 0 1 1 2
Mean 0
St. dev 1.22
N~(0,1.49)
STANDARDIZATION
$z = \frac{x-\mu}{\sigma}$
σ
$\mu = 0$
x
$x - \mu$


Standard normal distribution
Standardization
Original dataset
1 2 2 3 3 3 4 4 5
Mean 3
St. dev 1.22
N~(3,1.49)
Subtract mean
-2 -1 -1 0 0 0 1 1 2
Mean 0
St. dev 1.22
N~(0,1.49)
Divide by std
-1.63 -0.82 -0.82 0.00 0.00 0.00 0.82
Mean 0.00
St. dev 1.00
N~(0,1)
STANDARDIZATION
$z = \frac{x-\mu}{\sigma}$
$\mu = 0$
x
$x - \mu$


Standard normal distribution
Standardization
Original dataset
1 2 2 3 3 3 4 4 5
Mean 3
St. dev 1.22
N~(3,1.49)
Subtract mean
-2 -1 -1 0 0 0 1 1 2
Mean 0
St. dev 1.22
N~(0,1.49)
Divide by std
-1.63 -0.82 -0.82 0.00 0.00 0.00 0.82 0.82 1.63
Mean 0.00
St. dev 1.00
N~(0,1)
x
$x - \mu$
$\frac{x-\mu}{\sigma}$

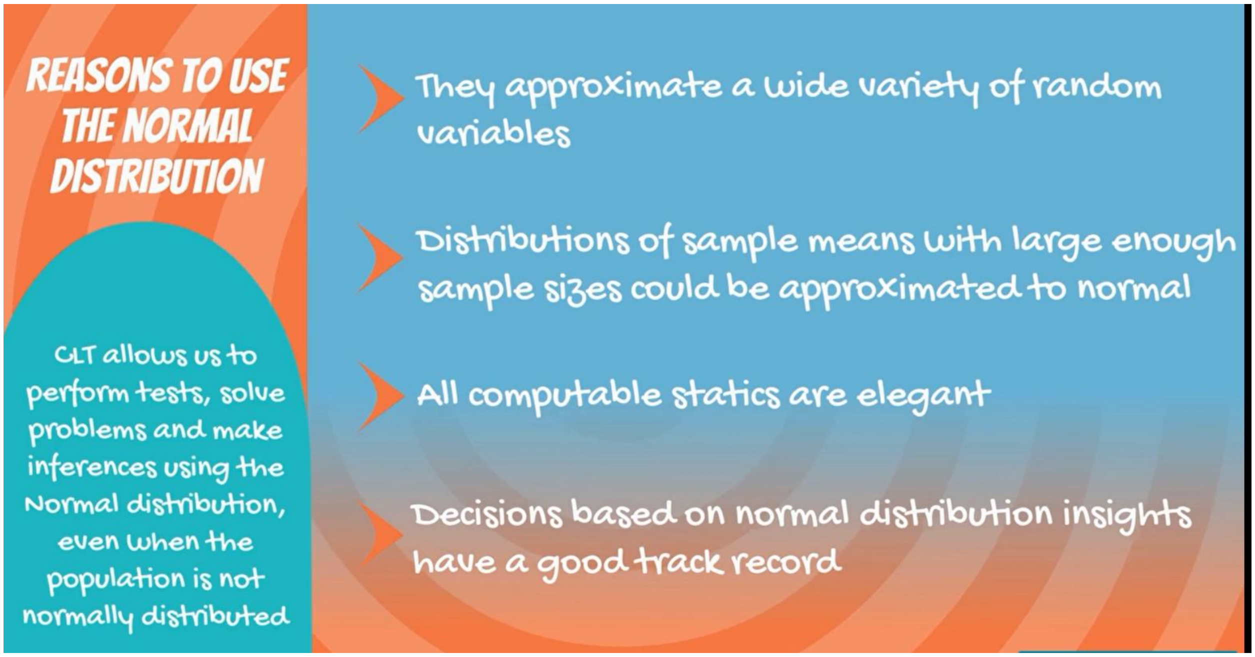
9. Central Limit Theorem





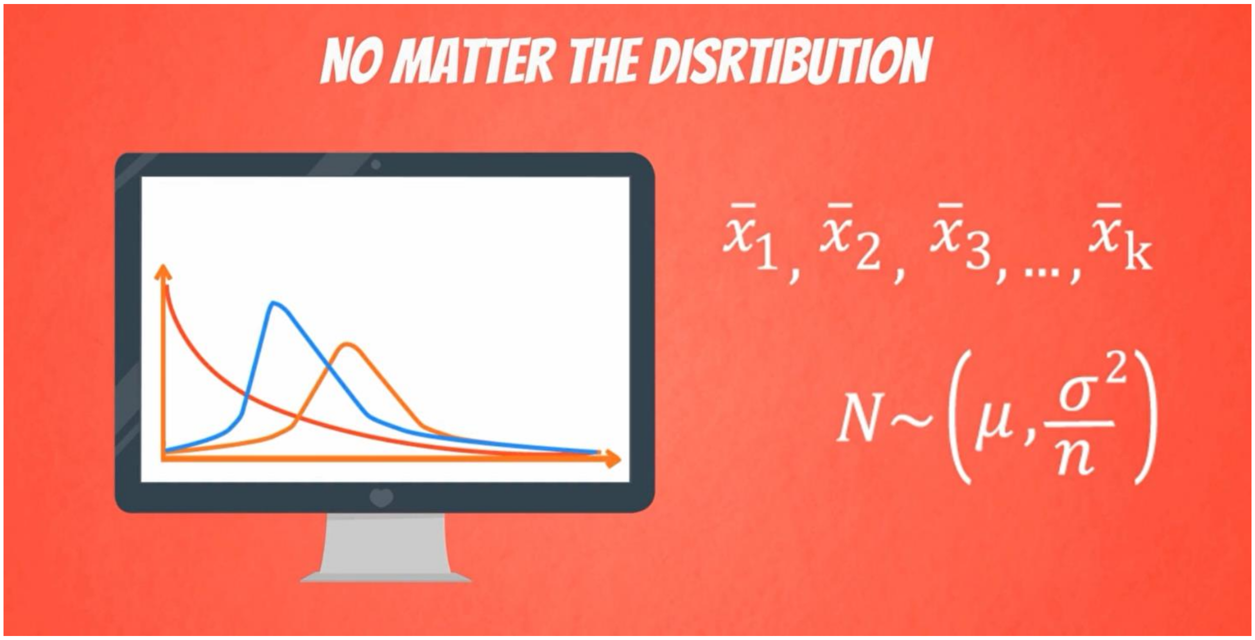
If we are able to draw bigger samples our statistical results will be more accurate usually for Central Limit Theorem apply.

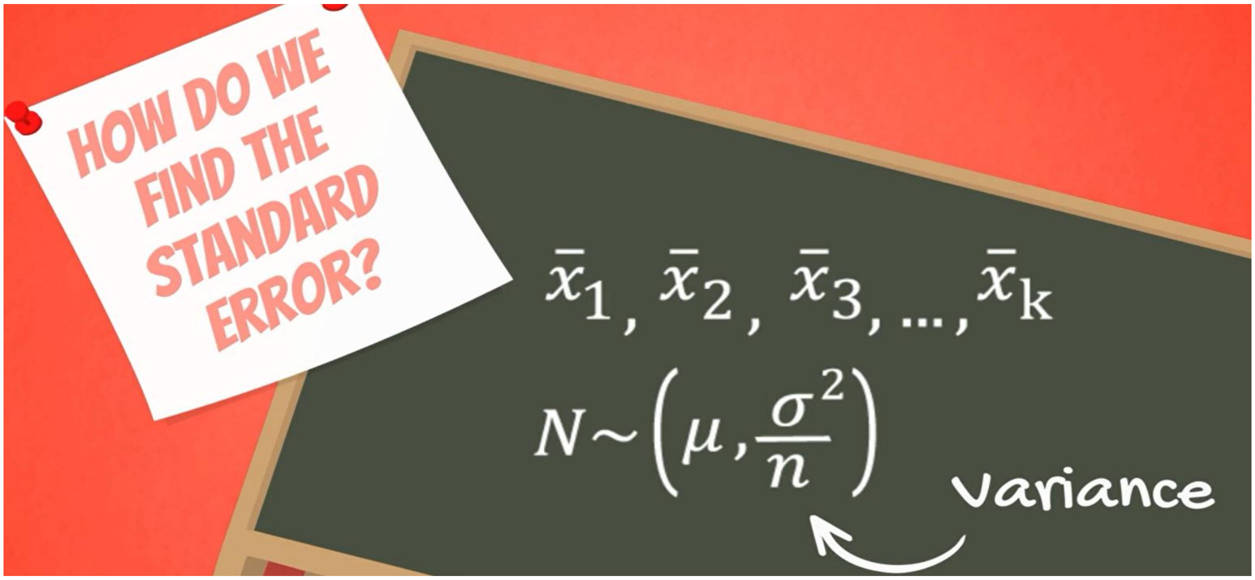
We need a sample size of at least 30 observations.

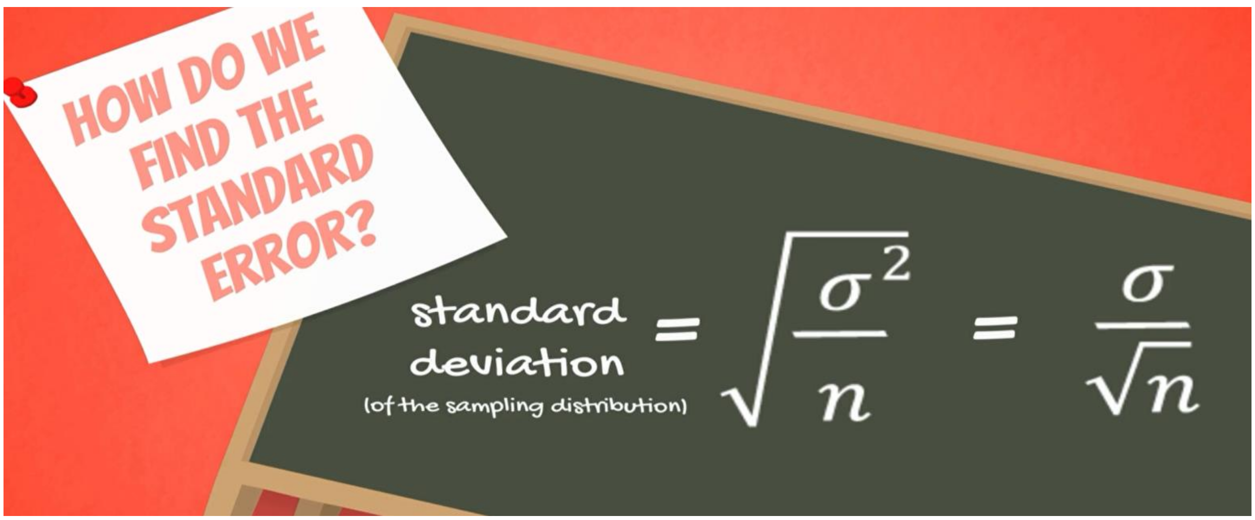


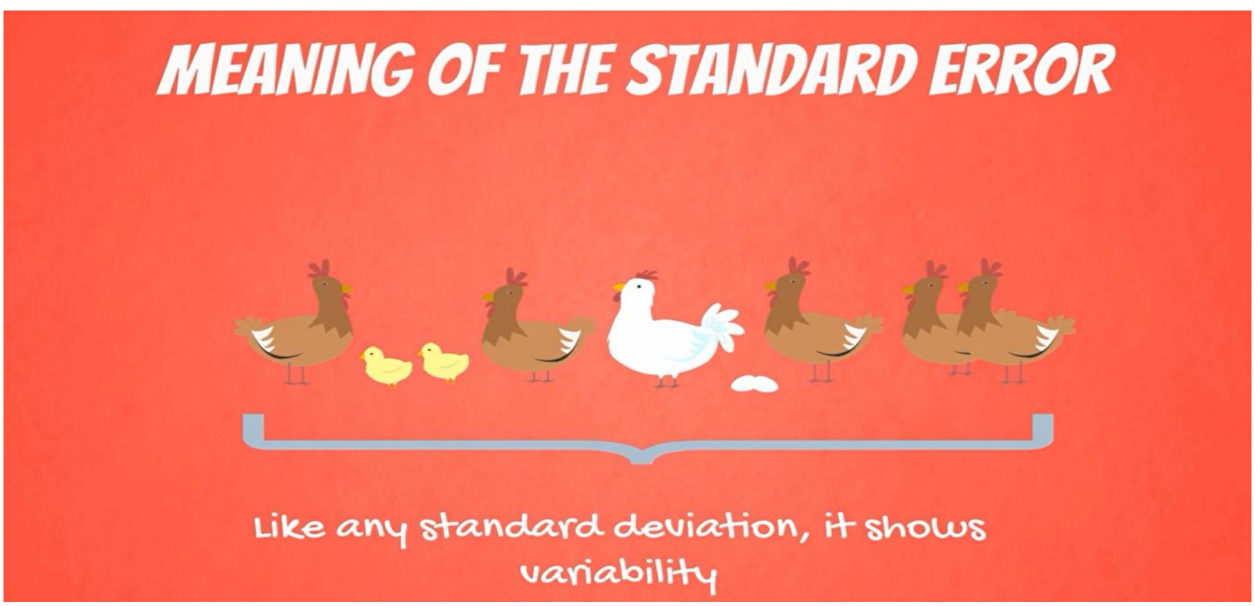
11. Standard error

Standard error is the standard deviation of the distribution formed by the sample means.




HOW DO WE FIND THE STANDARD ERROR?
$\bar{x}_1, \bar{x}_2, \bar{x}_3, \ldots, \bar{x}_k$
$N \sim \left(\mu, \frac{\sigma^2}{n}\right)$
Variance


HOW DO WE FIND THE STANDARD ERROR?
standard deviation (of the sampling distribution) $= \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$

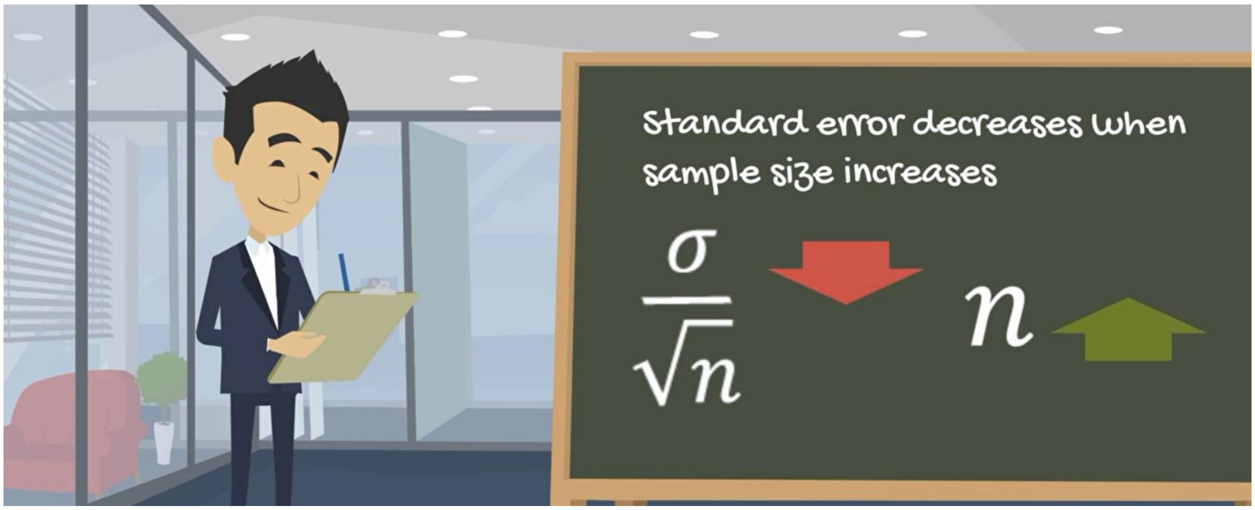

MEANING OF THE STANDARD ERROR
Like any standard deviation, it shows variability

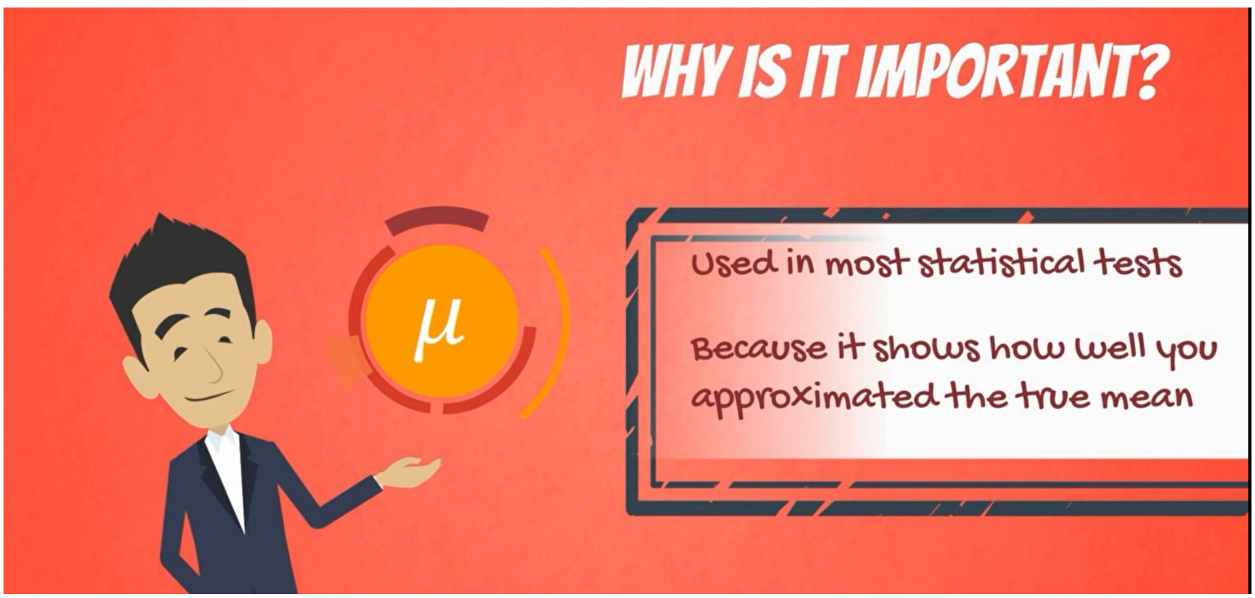
Standard error shows variability

It is the variability of the means of the different samples we extracted

Note: Standard error decreases as the sample size increases.

Bigger sample give a better approximation of the population





13. Estimators and Estimates

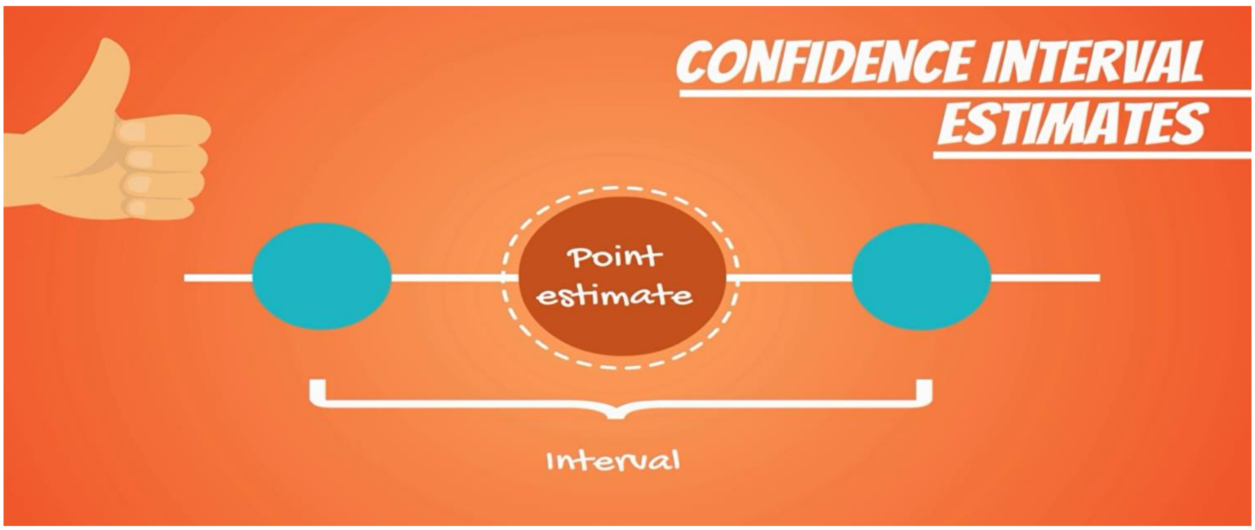
Let's continue by introducing the concept of an estimators of a population parameter.

There are two types of estimates:

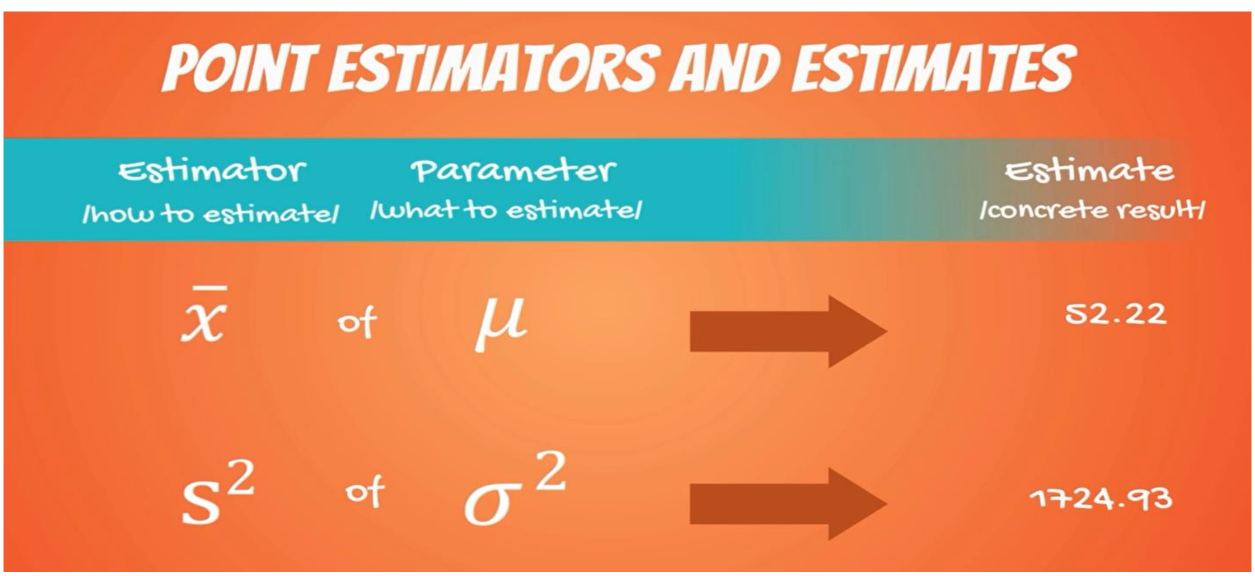
1. Point estimates
2. Confidence interval estimates.

A point estimate is a single number

A confidence interval naturally is an interval the two are closely related.



The point estimate is located exactly in the middle of the confidence interval.



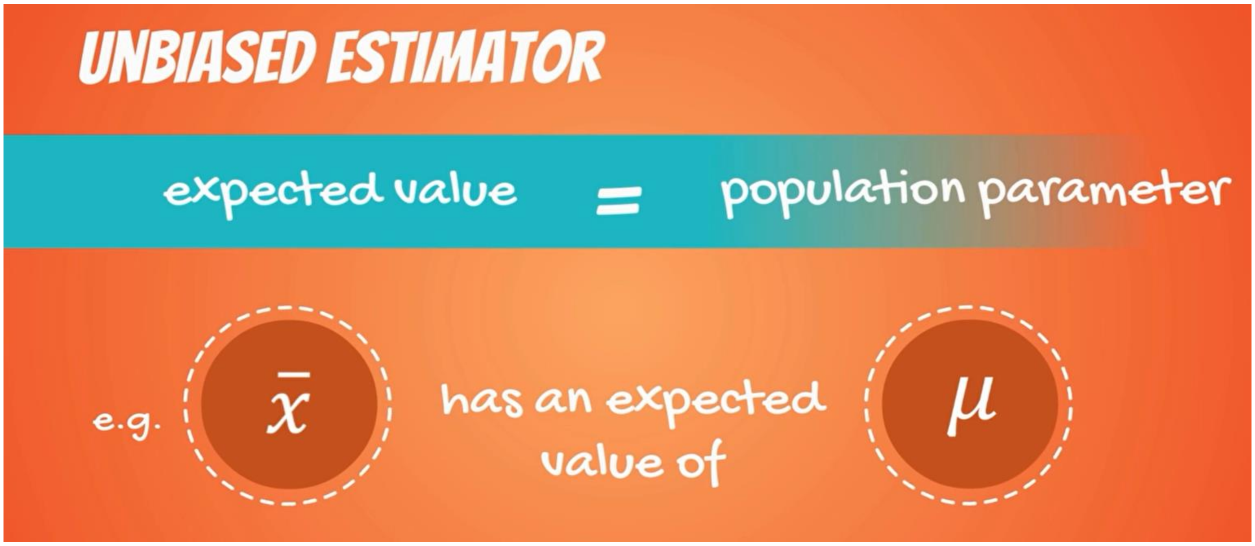
The Sample mean X bar is a point estimate of the population mean mew.

The sample variance S^2 was an estimate of the population variance Sigma squared.

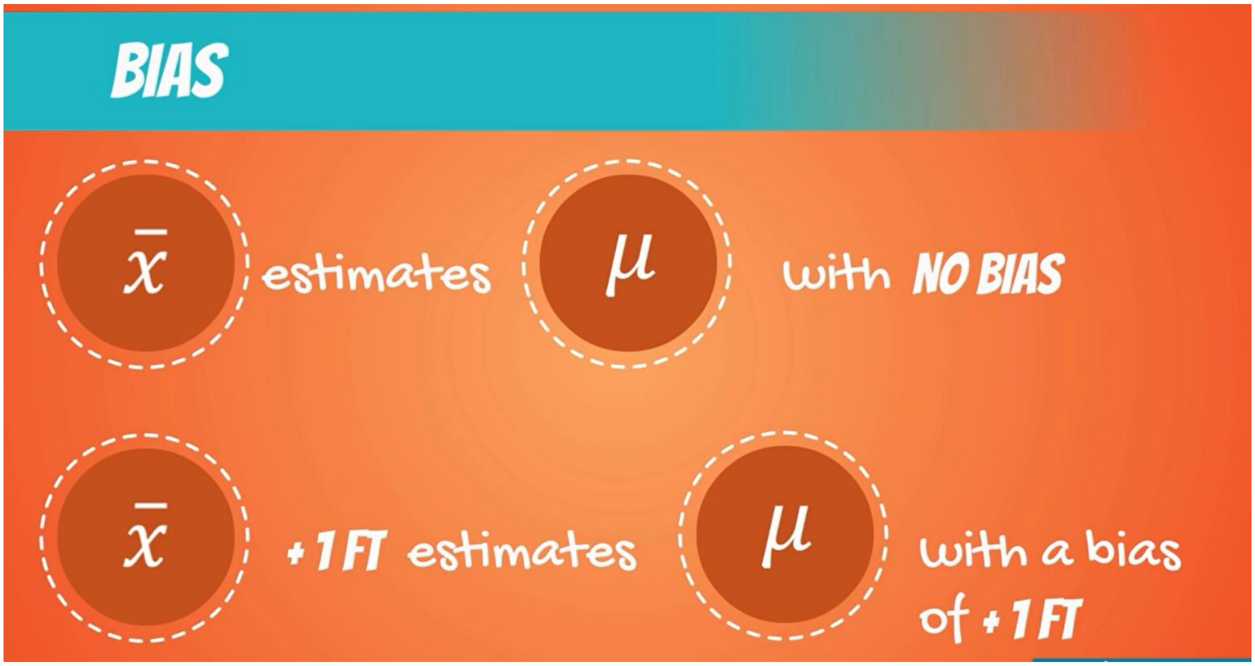
They all have two properties:

1. Bias
2. Efficiency

We are looking for the most efficient unbiased estimators and unbiased estimator has an expected value equal to the population parameter.



Bias Estimators




EFFICIENCY
The most efficient estimator is the unbiased estimator with smallest variance


STATISTICS
ESTIMATORS
broader term
a type of statistic