

Why does sampling work?

Remember these terms:

- sample
- population
- statistic
- parameter

Remember we use **statistics** to estimate **parameters**.

BUT ...

a sample is just a sample.

How can we be so sure our estimate will be reasonable?

Sampling Distribution

- Consider all the possible samples of a certain size
- Think of the mean of every one of those possible samples.
- The **sampling distribution** is the distribution of all those possible means.

(There is a different sampling distribution for every size of sample.)

(You could also do a sampling distribution for other statistics.)

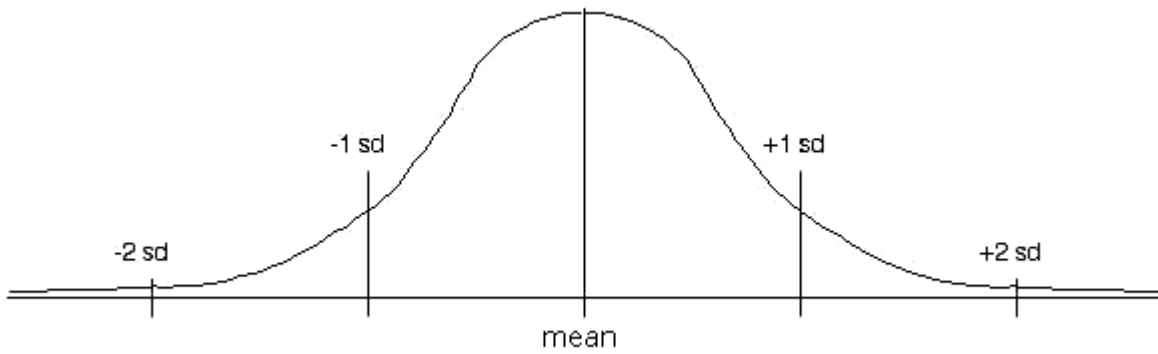
Central Limit Theorem

A sampling distribution has:

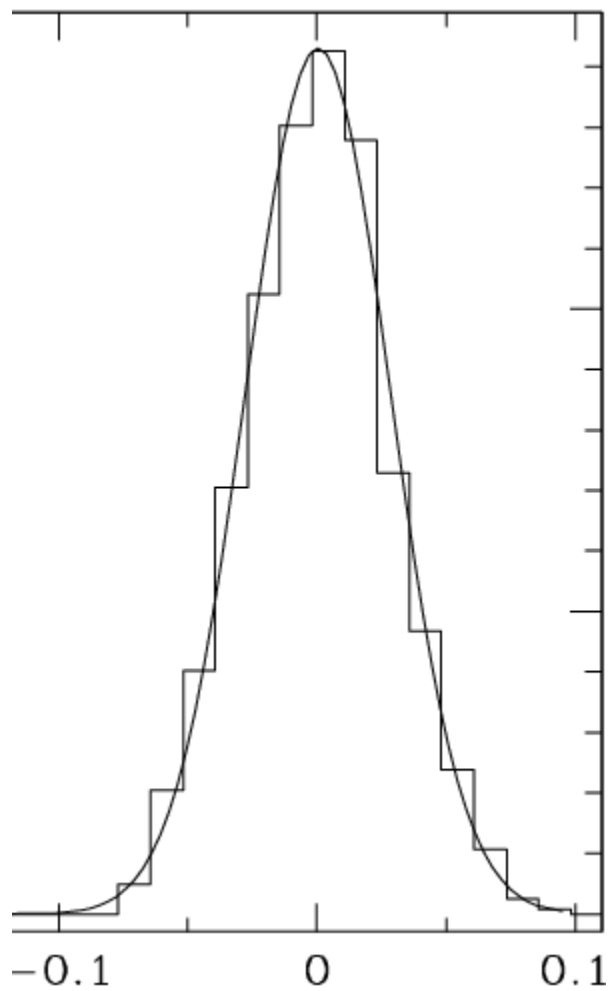
- the **same** mean as the population
- a much **smaller** standard deviation than the population
- So the sampling distribution is much more tightly packed than the population.



Population:

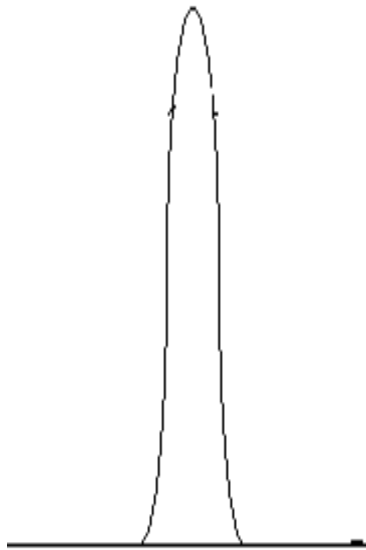


Sampling Distribution:



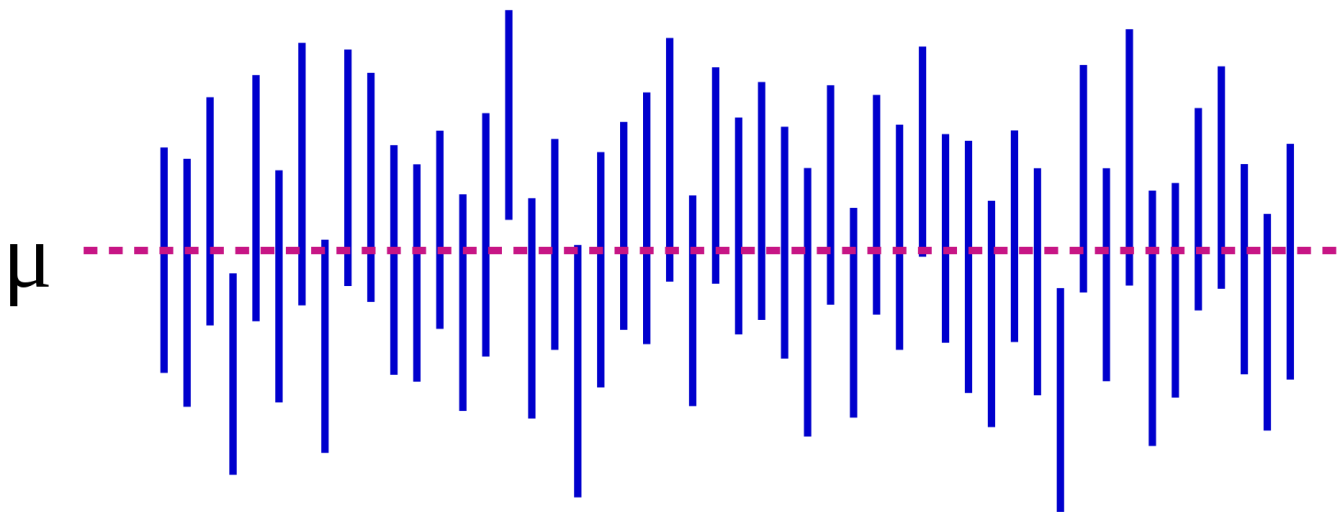
- This means there is a very low probability of picking a sample whose mean is far away from the mean of the population.

- As n increases, the distribution is even more tightly packed.



- So when you have a **big sample** it is almost impossible to have a mean that is far from the actual mean of the population.

FOR INSTANCE,
Consider all the 95% confidence intervals for a certain statistic:



The term “95% confidence” means 95% of the samples will produce an estimate that includes μ , the actual mean of the population.

- A few will be too high or too low
- The vast majority will include μ .