Inference and confidence interval for mean

Inference and confidence interval for mean

We can distinguish **two types of statistical inference methods**. We can **estimate population parameters** and we can **test hypotheses** about these parameters. In the present module we'll talk about the first type of inferential statistics.

In this section we argue that we can estimate the value of a population parameter in two ways: by means of a so-called **point estimate** (a single number that is our best guess for the population parameter) or by means of an **interval estimate** (a range of values within which we expect the parameter to fall).

An interval estimate is a range of numbers, which, most likely, contains the actual population value. The probability that the interval contains the population value is what we call the **confidence level.**

Inference and confidence interval for mean

I usually slept about eight hours every night. Now, I only sleep about five hours per night. That means a reduction of three hours per night. That's about 20 hours per week and about 1000 hours per year. That's about 40 days.

In other words, if my son continues with his present sleeping schedule, after a year I will have lost about **forty full days** of sleep!!

Just to let you know ☺

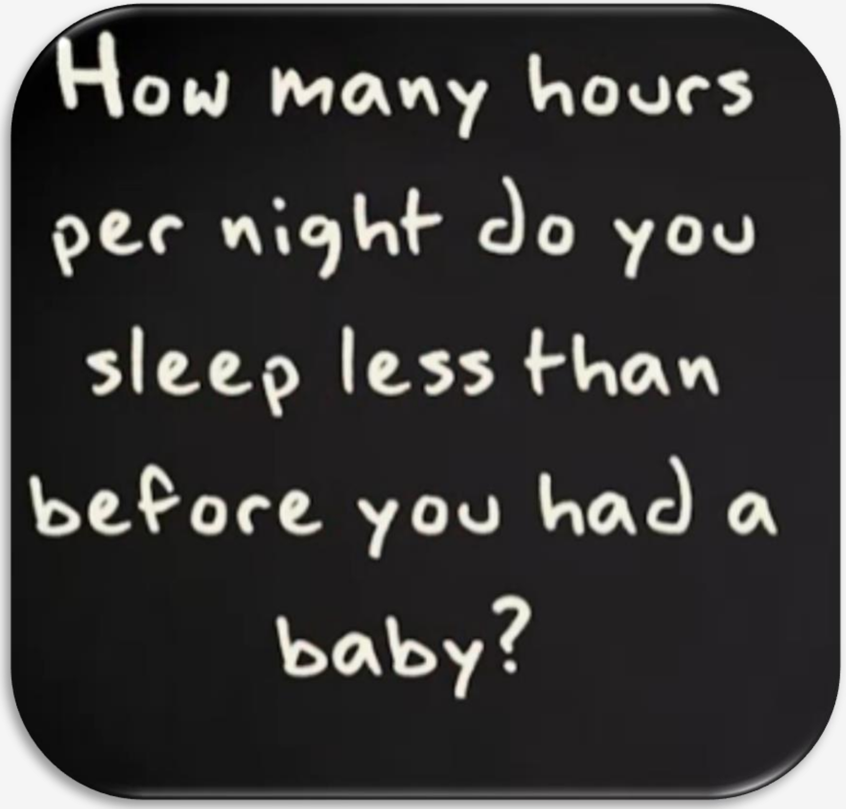
Inference and confidence interval for mean

And I was wondering how much sleep other new parents in my home city of KL lost after they had their first baby. Are me and my wife the only ones losing so much sleep?

Suppose I contacted the local authorities and asked them the contact details of all new parents in KL.

Let's say that new parents are those who got a baby within the last six months.

I drew a simple random sample of 60 new parents and asked them **how much hours per night they slept less than before they had a baby.**

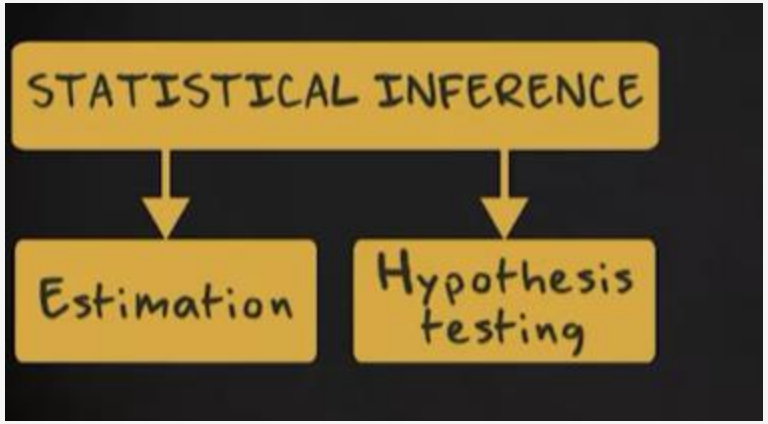


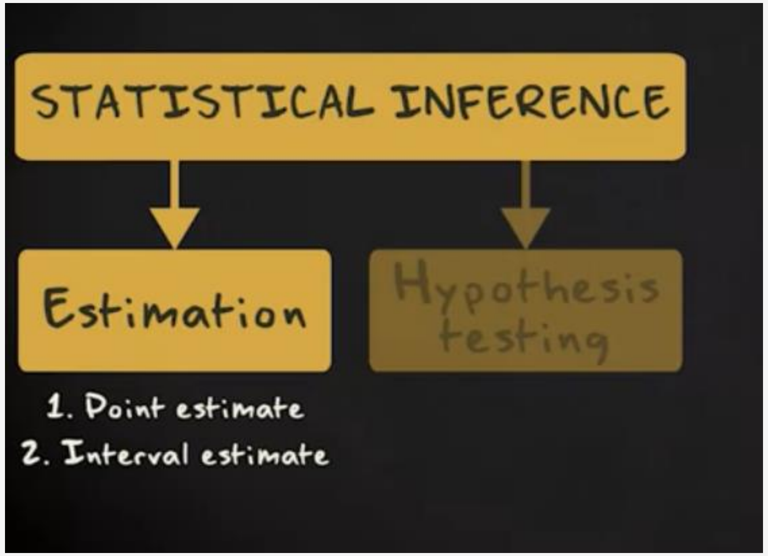
Inference and confidence interval for mean

In this module I will talk about statistical inference. I will, on the basis of sample information, draw conclusions about the entire population from which the sample was drawn.

In this module we will talk about the first type of inferential statistics.

There are two ways in which we can estimate the value of a population parameter. The first one is the so-called **point** estimate. It is a single number that is our best guess for the population parameter. The second one is the **interval** estimate. It is a range of values within which we expect the parameter to fall.





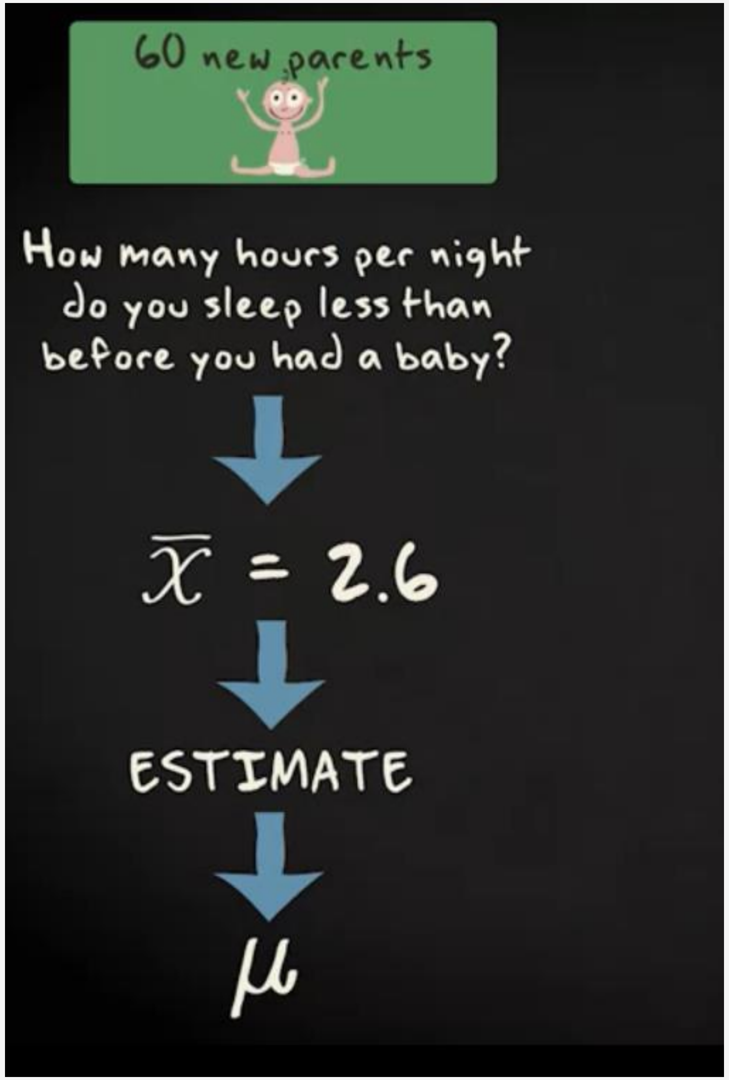
Inference and confidence interval for mean

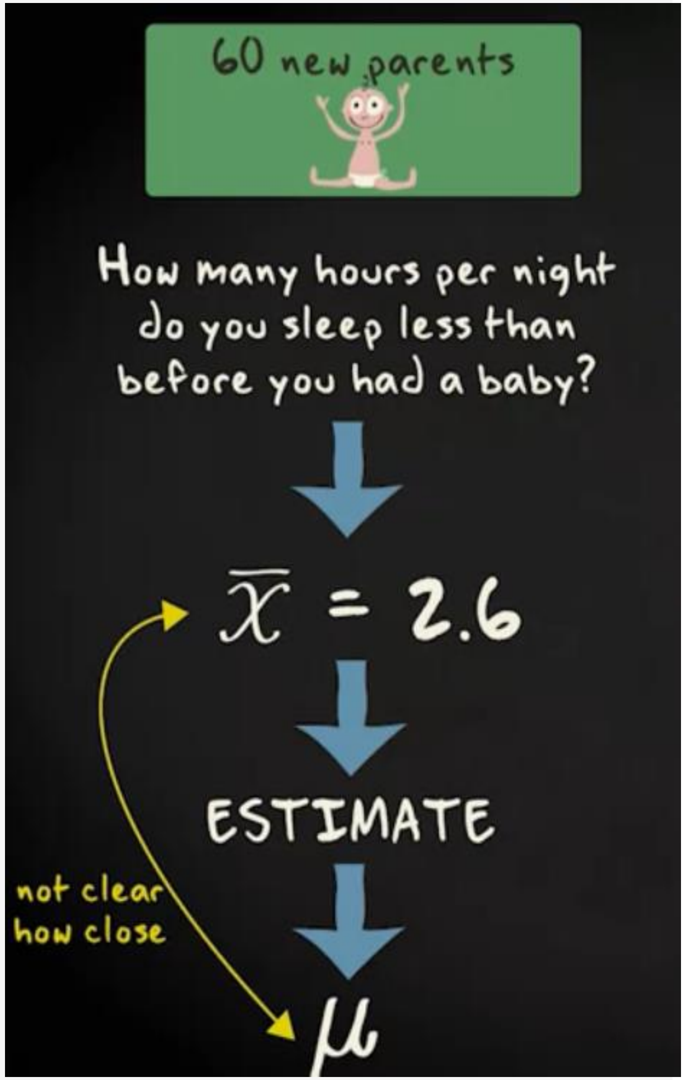
Let's assume that the mean number of hours that the 60 respondents in my sample slept less after they had their first baby is **2.6**.

That means that a good point estimate for the mean number of lost sleeping-hours in the population is, well, 2.6. In other words, the **statistic X-bar** (which in our case is 2.6 hours) is a good point estimate for the parameter **Mu**.

However, one individual point estimate **doesn't** tell us if this estimate is close to the population parameter we are interested in or not.

Therefore, next to a point estimate, researchers often also want to know the likely precision of this point estimate.



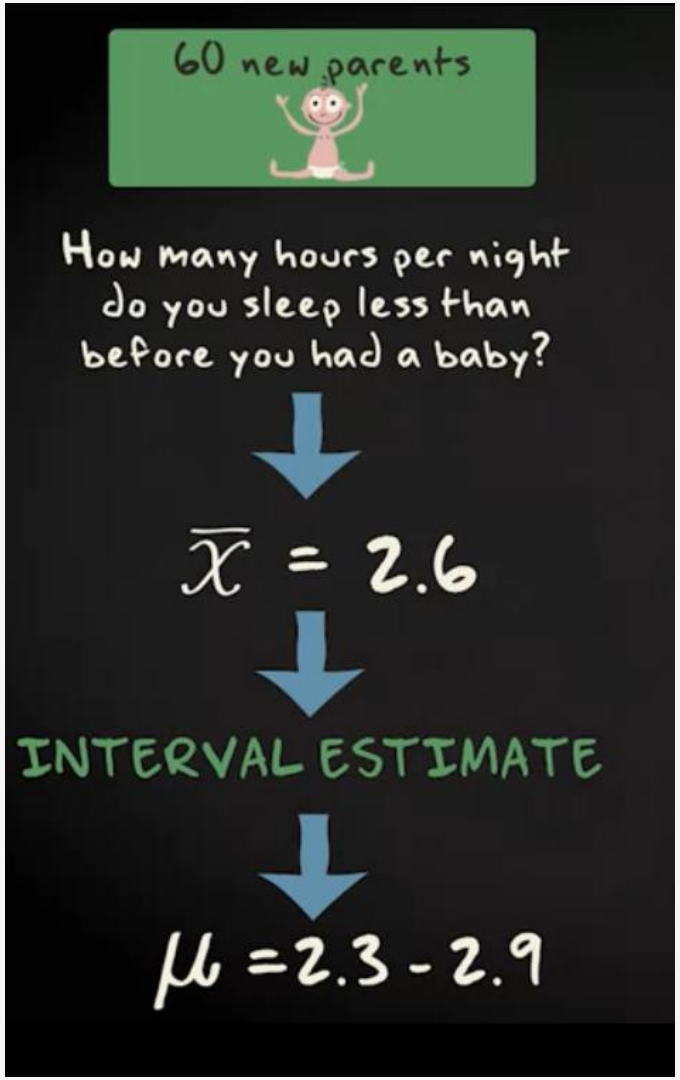


Inference and confidence interval for mean

They show this likely precision by also computing an interval estimate.

An interval estimate is a range of numbers, which, most likely, contains the actual population value.

On the basis of our sample mean of 2.6 hours we might predict, for instance, that the mean lost sleeping-hours of all new parents in Amsterdam lies somewhere between 2.3 and 2.9.



Inference and confidence interval for mean

The probability that the interval contains the population value is what we call the **confidence level**.

The confidence level always has a value close to one.

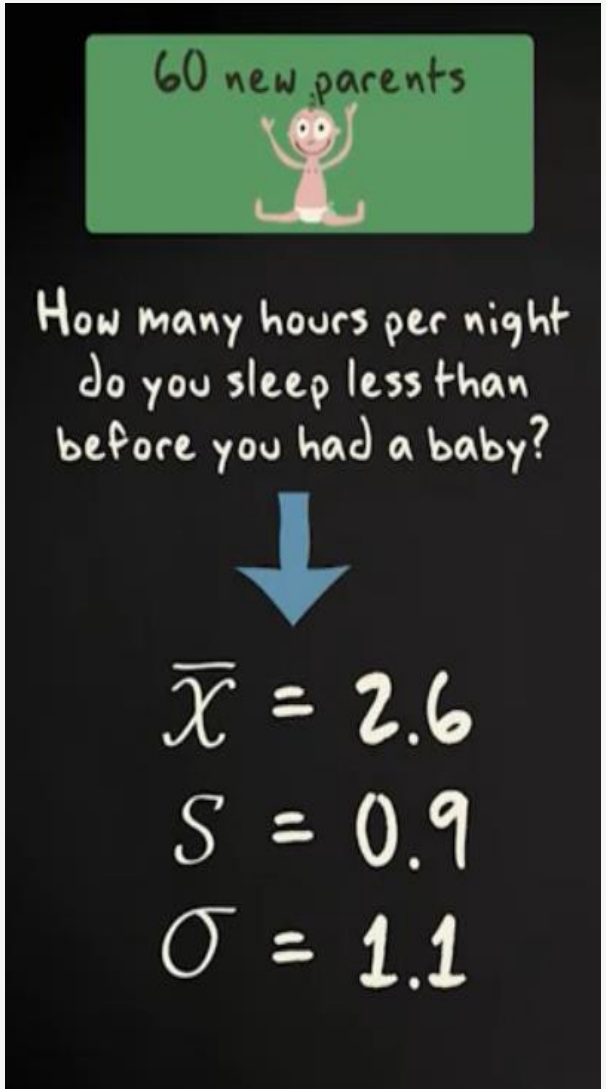
In most cases it's 0.95. In that case we talk about the 95 percent confidence interval.

In the next slides we'll discuss how we can construct such confidence intervals.

CI for mean with known population SD

In this section we will show you how you can **construct a confidence interval** based on the information from your **sample** and the **population standard deviation (SD)**.

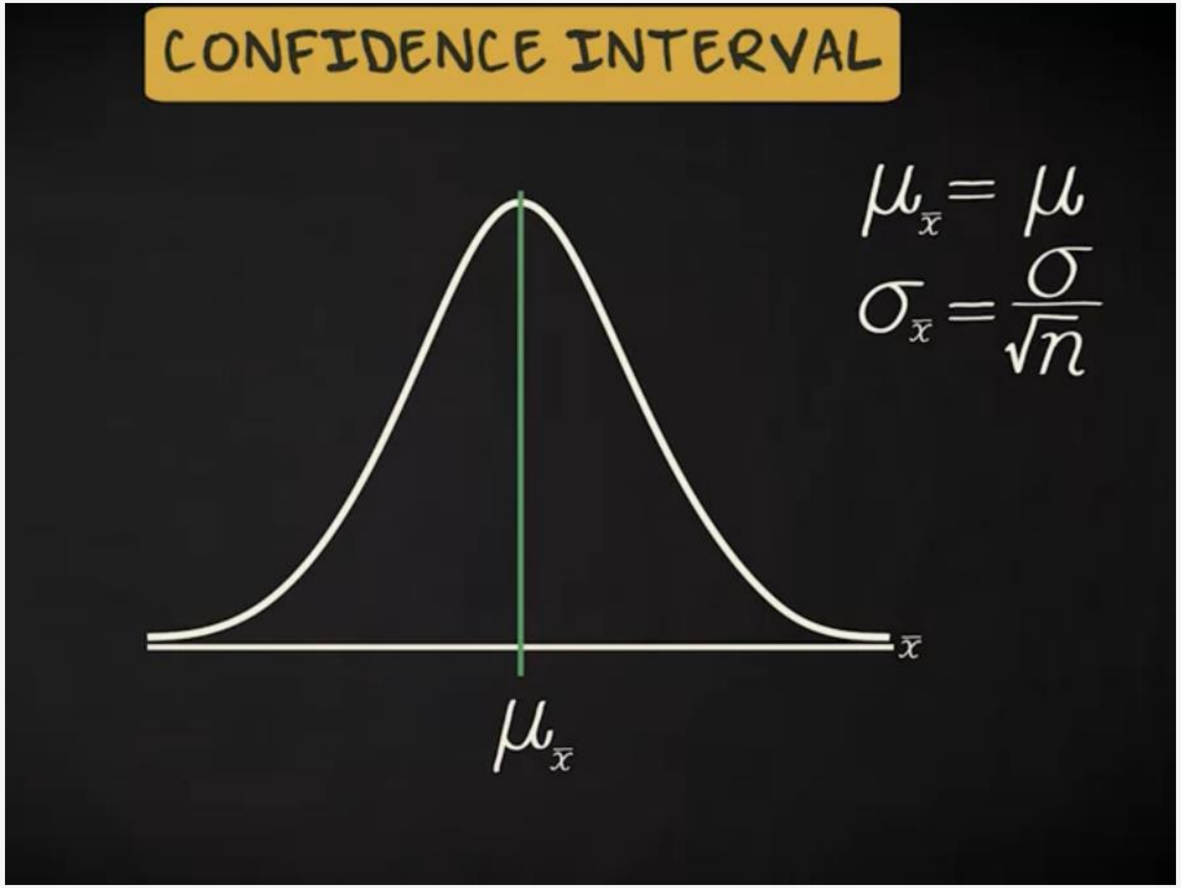
We'll explain how such a confidence interval should be interpreted



CI for mean with known population SD

In this section we will show you how you can **construct a confidence interval** based on the information from your **sample** and the **population standard deviation (SD)**.

We'll explain how such a confidence interval should be interpreted

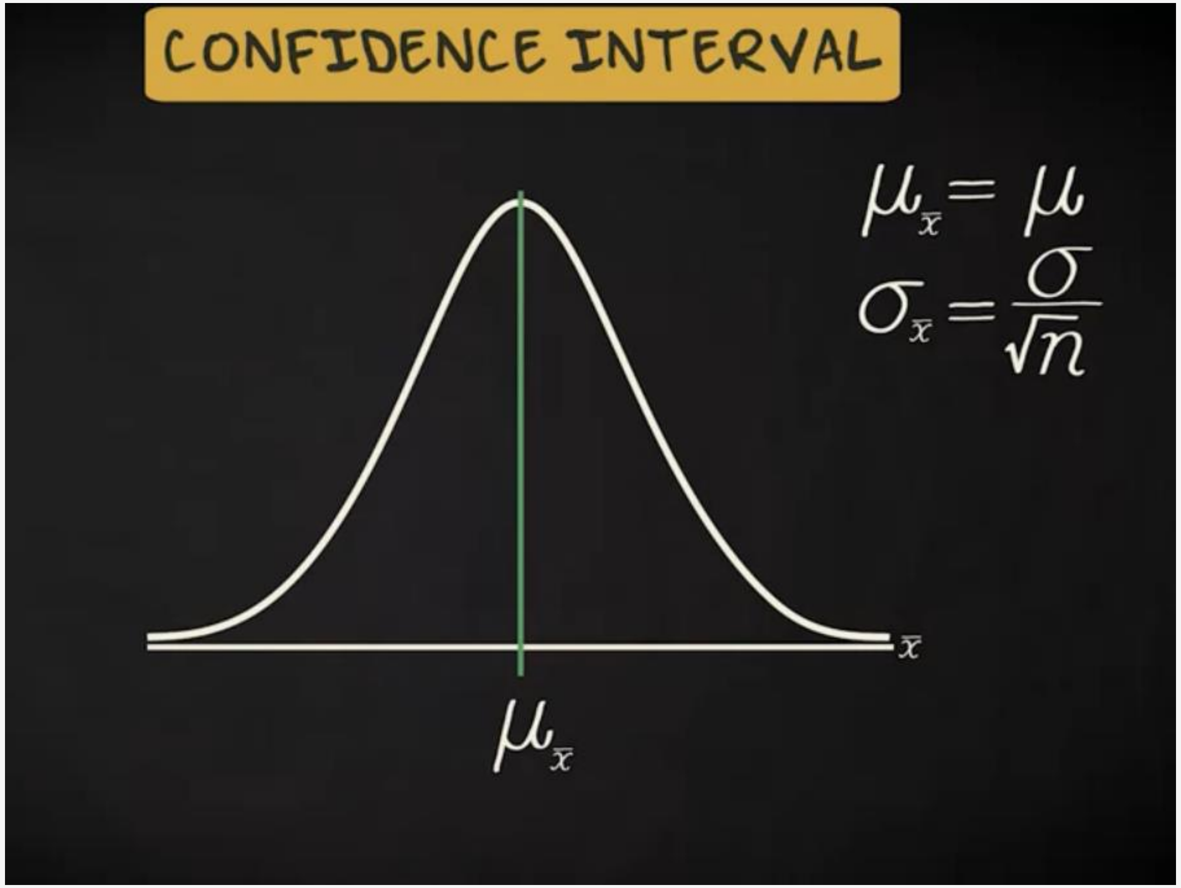


CI for mean with known population SD

We know that, as long as our sample is sufficiently large, the sampling distribution is normally distributed with a mean that is equal to the population mean Mu

and

a standard deviation that is equal to the population standard deviation divided by the square root of n.

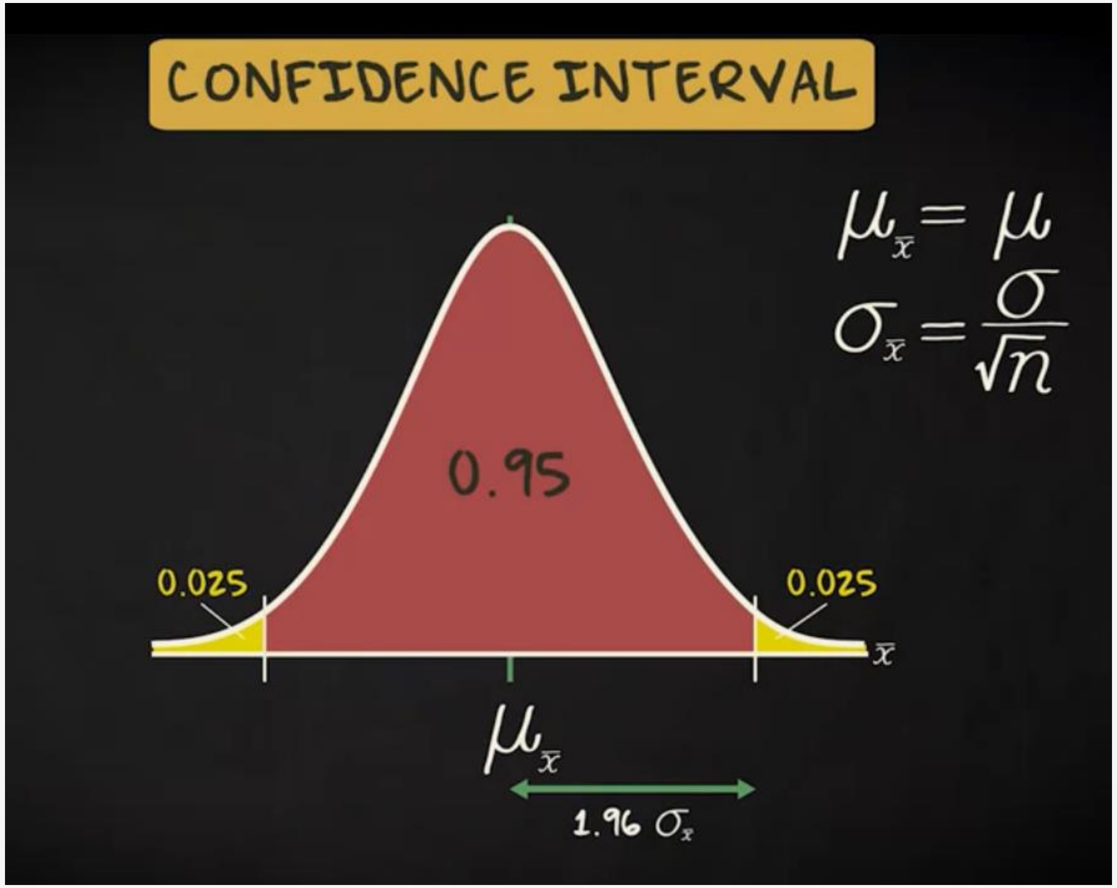


CI for mean with known population SD

We also know that the probability of finding a sample mean of less than about 2 standard deviations from the mean is 0.95.

More precisely, if we look up the z-scores which correspond to this probability, we'll find values of minus 1.96 and 1.96.

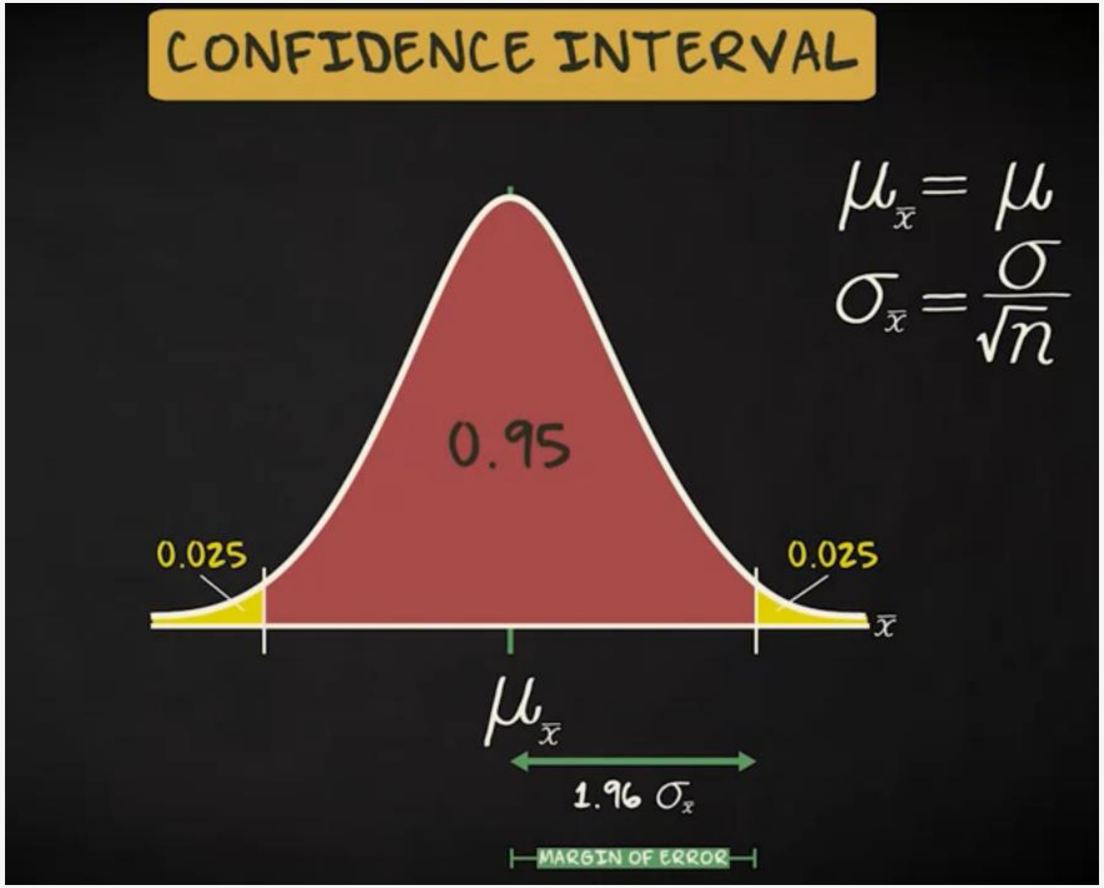
This means that we have a 95% chance that our sample mean will fall within 1.96 standard deviations of population mean Mu



CI for mean with known population SD

This distance of 1.96 standard deviations is what we call the **margin of error**.

The margin of error tells us how accurately our sample mean X-bar is likely to estimate our population mean Mu.

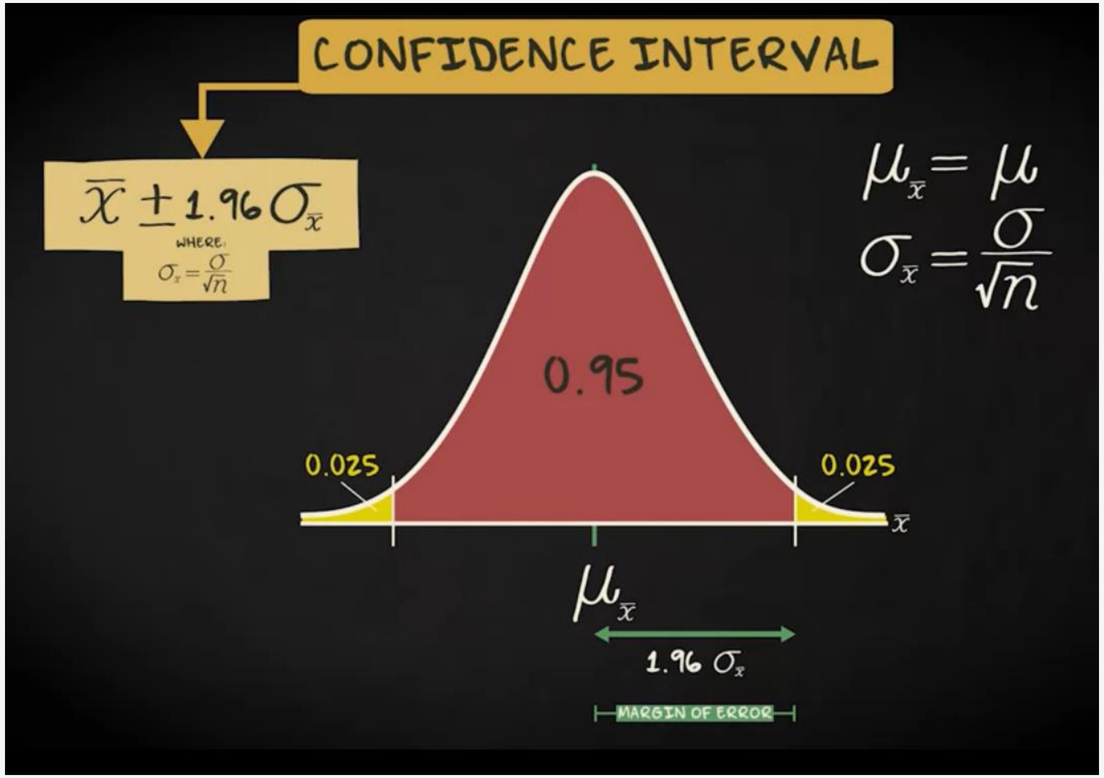


CI for mean with known population SD

Now,

the formula of the 95% confidence interval is the following. It is the point estimate (or the sample mean) plus and minus the margin of error (which equals 1.96 standard deviations).

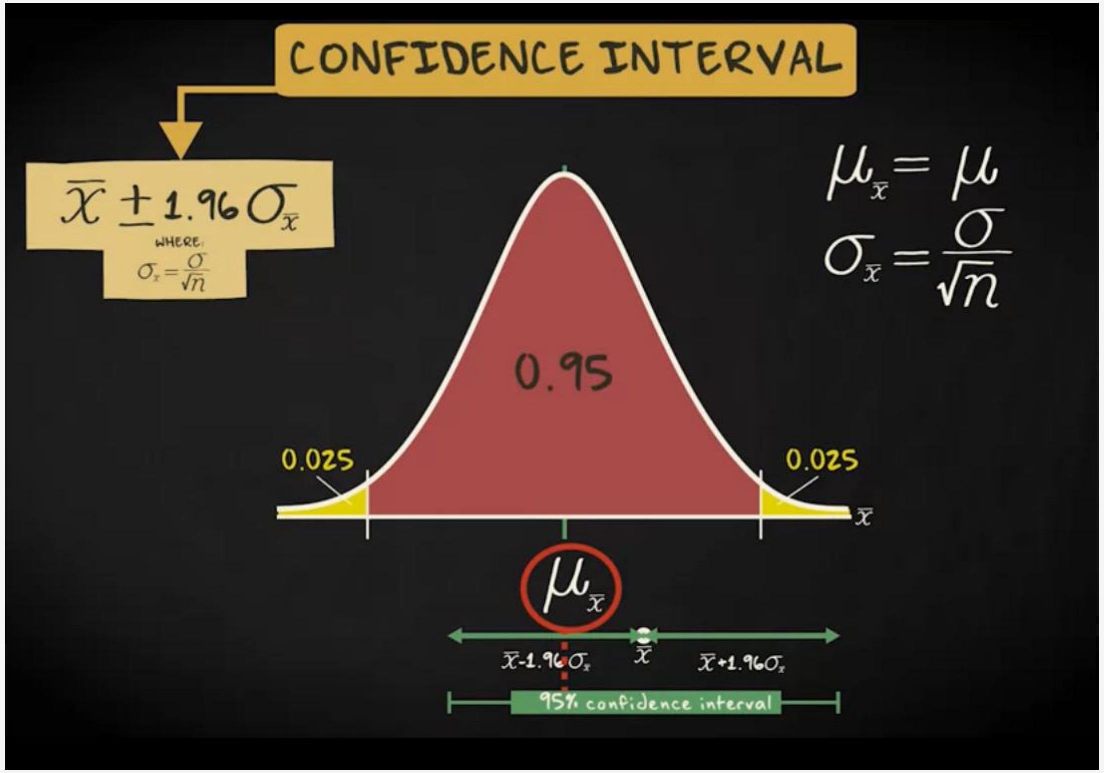
Note that we're dealing with the sampling distribution of the sample mean here, so the standard deviation equals Sigma divided by the square root of n.



CI for mean with known population SD

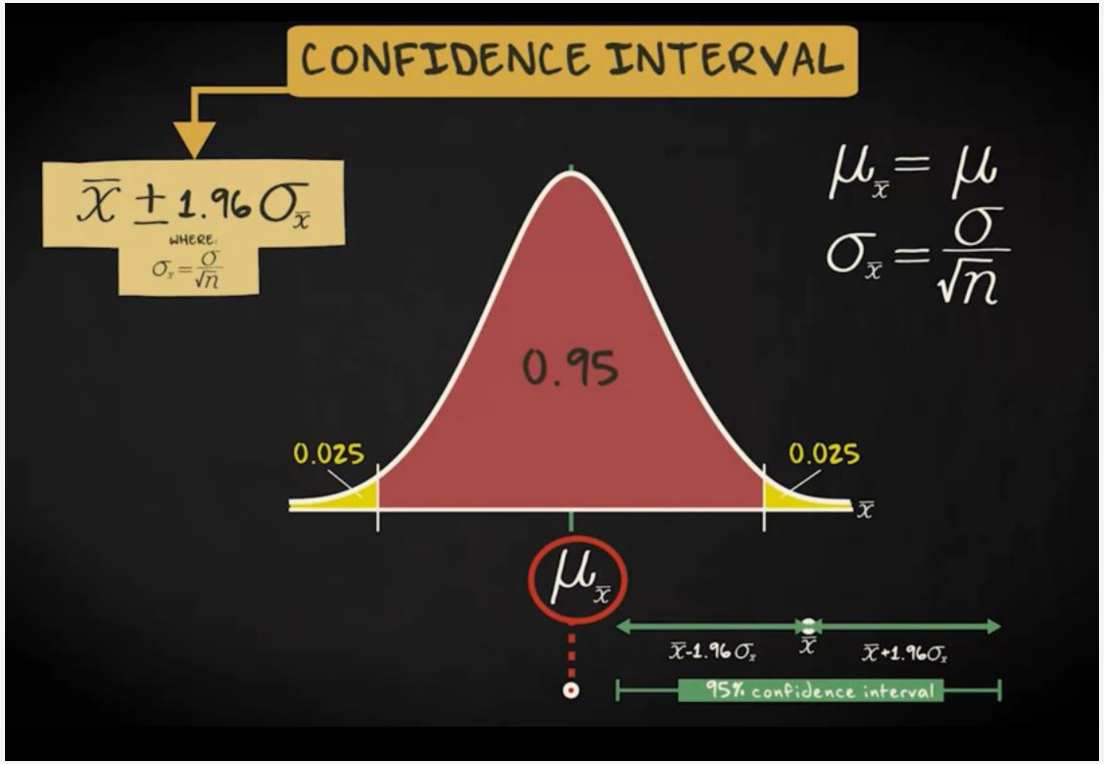
The mean of this sample is represented by this dot. The lines here represent the margins of error on both sides of the mean. Together they form the 95% confidence interval.

If the sample mean falls within the red area, than the confidence interval contains the population mean Mu.



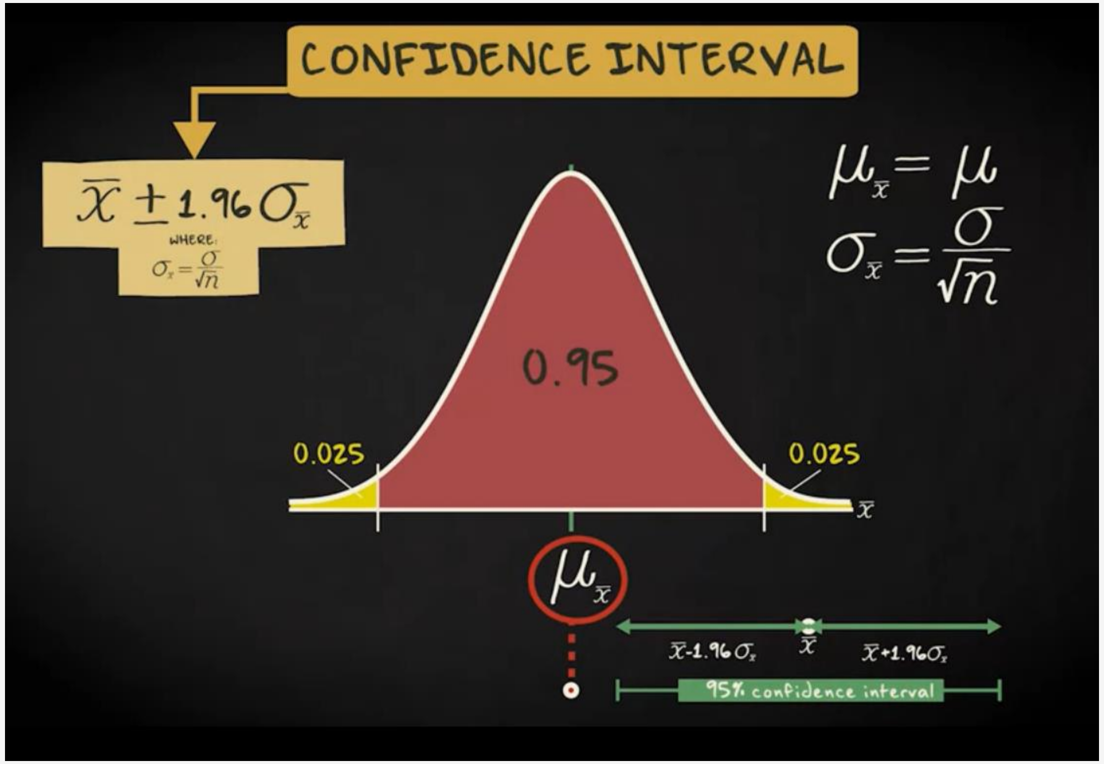
CI for mean with known population SD

If the sample mean does not fall within the red area, the confidence interval does not contain population parameter Mu.



CI for mean with known population SD

If the sample mean does not fall within the red area, the confidence interval does not contain population parameter Mu.

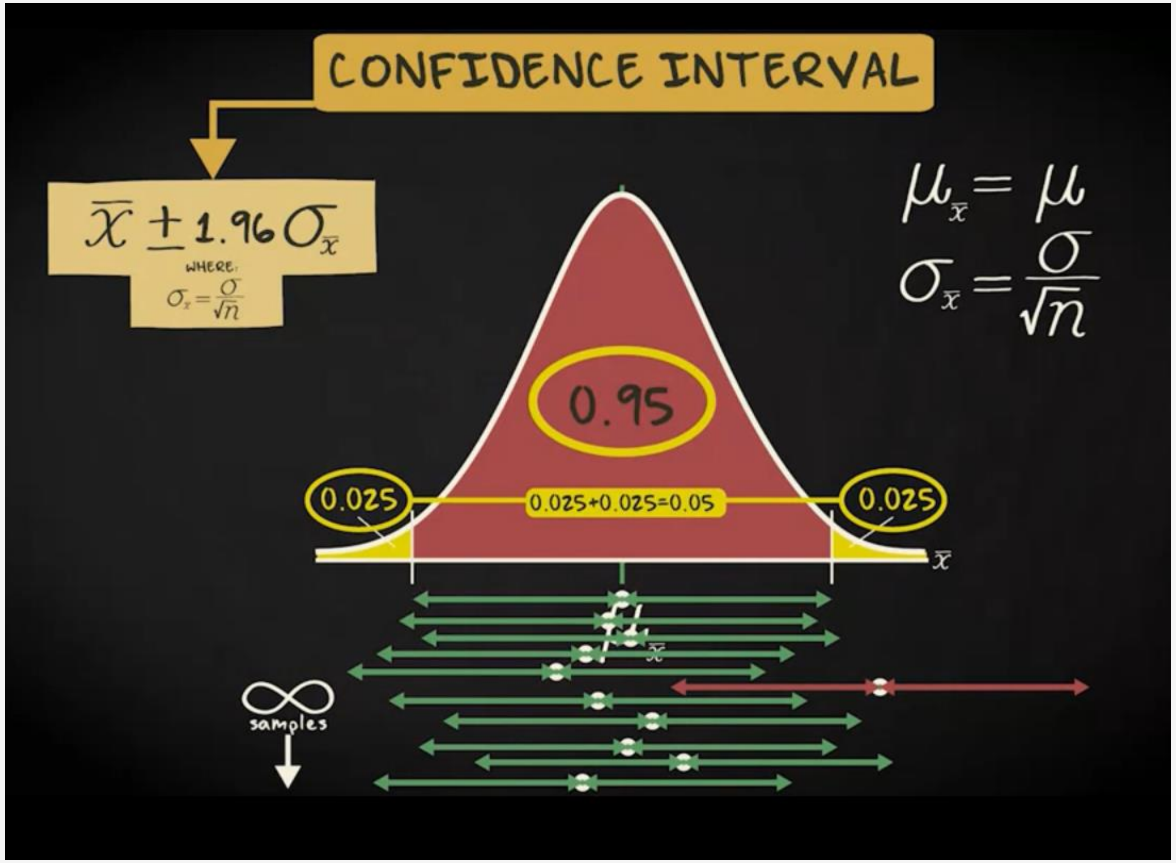


CI for mean with known population SD

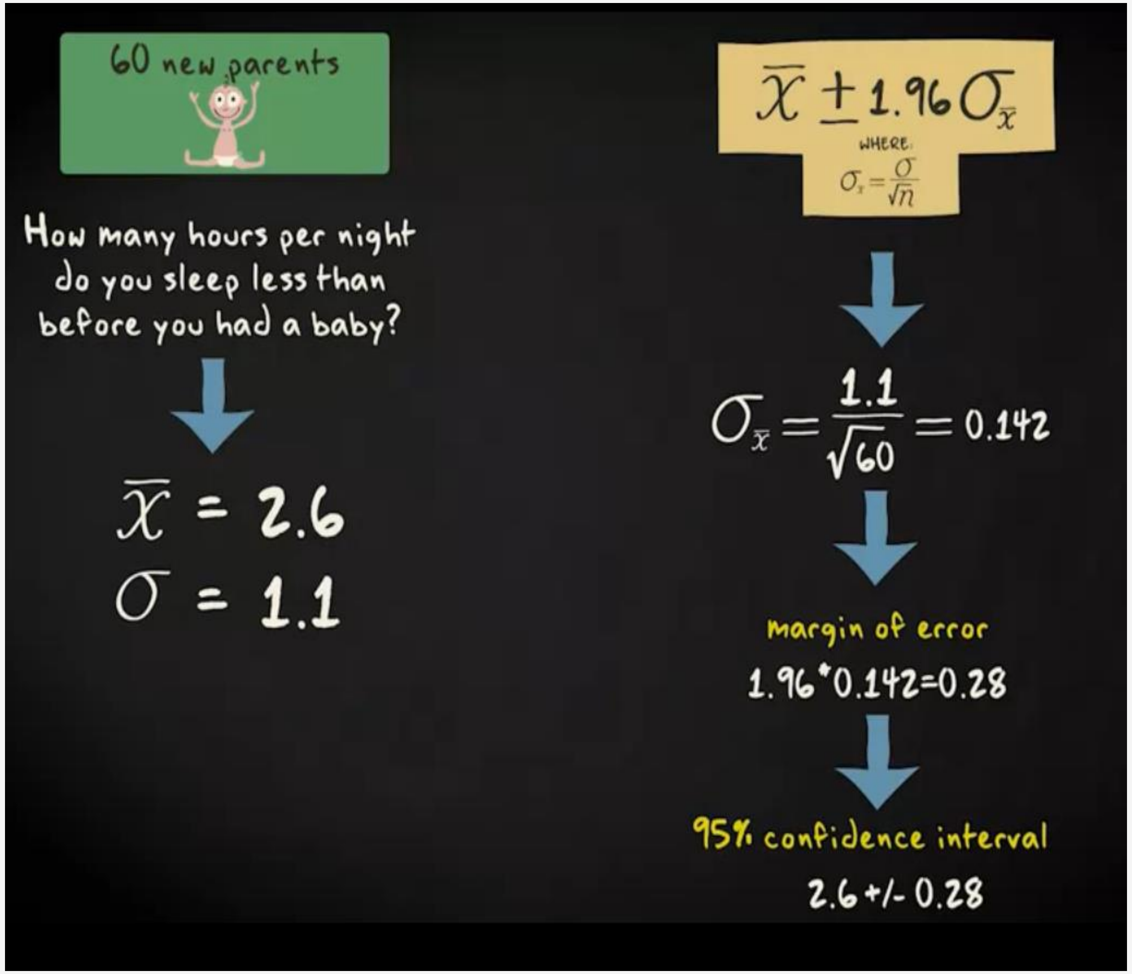
We're talking about the 95% confidence interval.

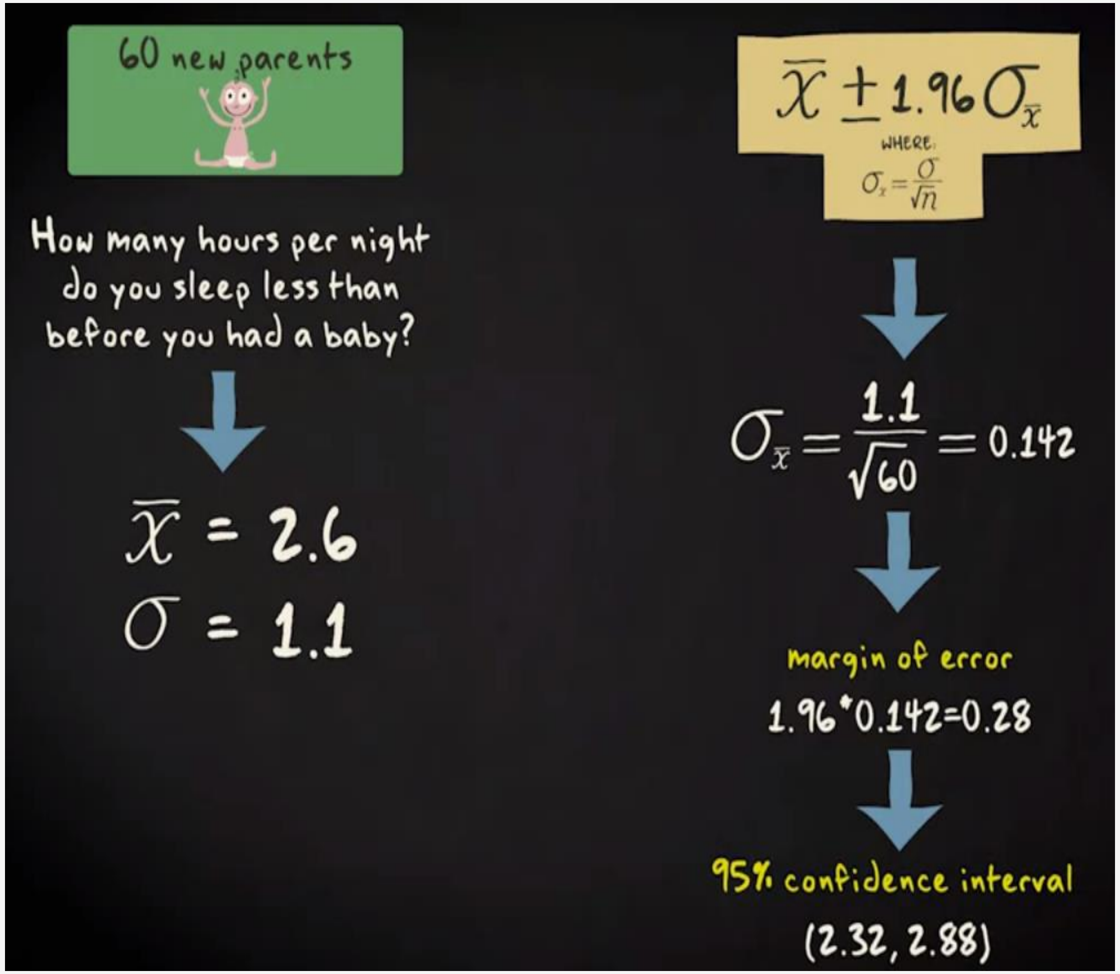
That means that the probability that the confidence interval of a randomly selected sample contains the population parameter is 0.95. The probability that it does not contain the population mean is 0.05.

In other words, if we would draw an infinite number of samples from our population, in 95% of the cases our confidence interval would contain population mean Mu.

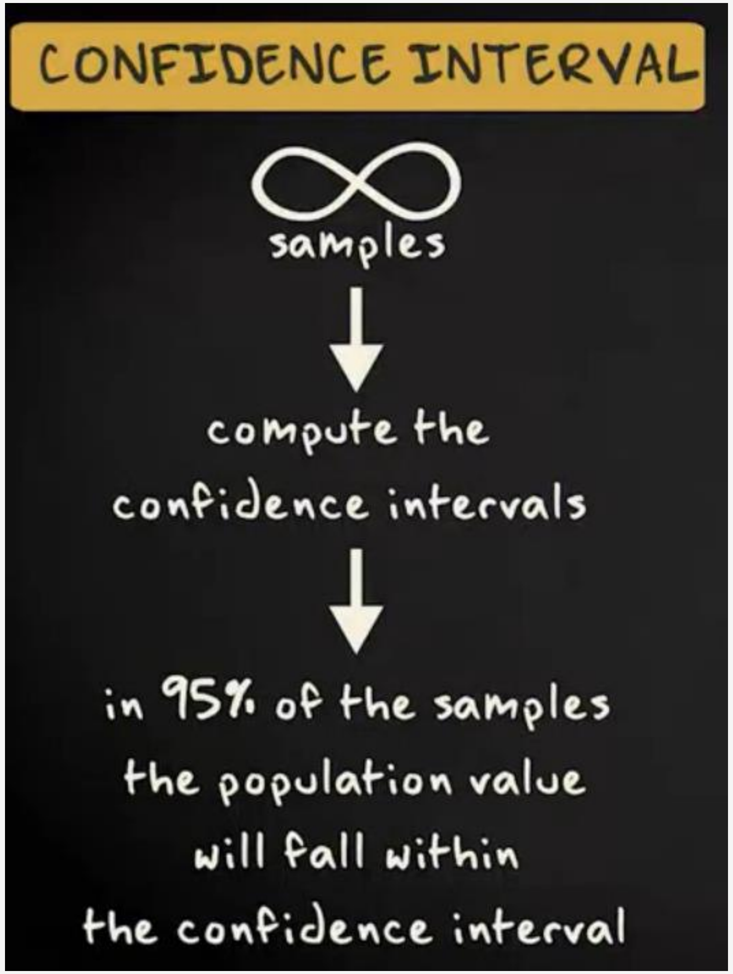


CI for mean with known population SD

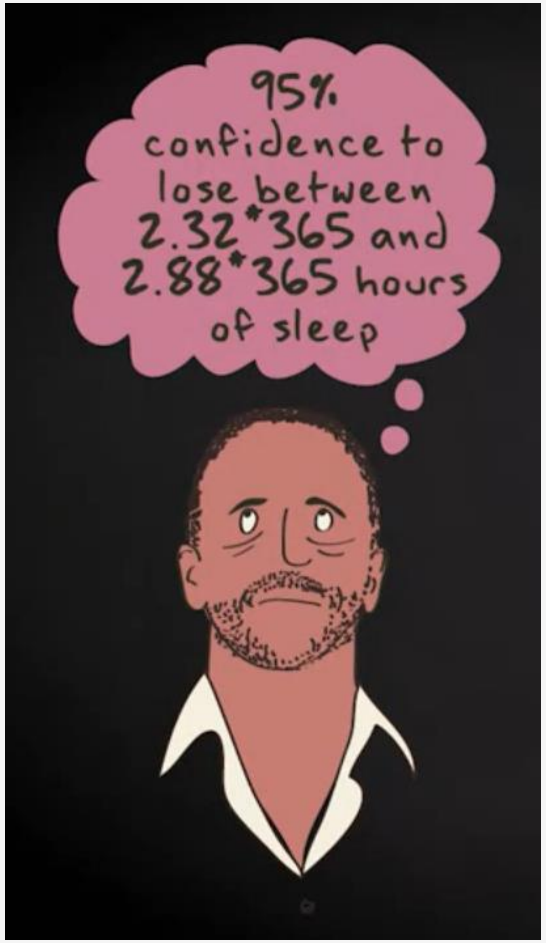


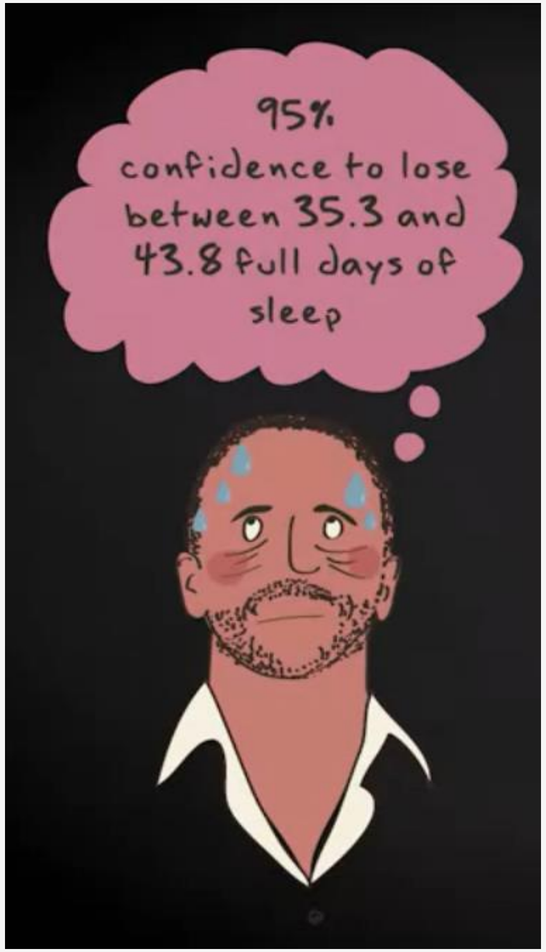


CI for mean with known population SD



CI for mean with known population SD



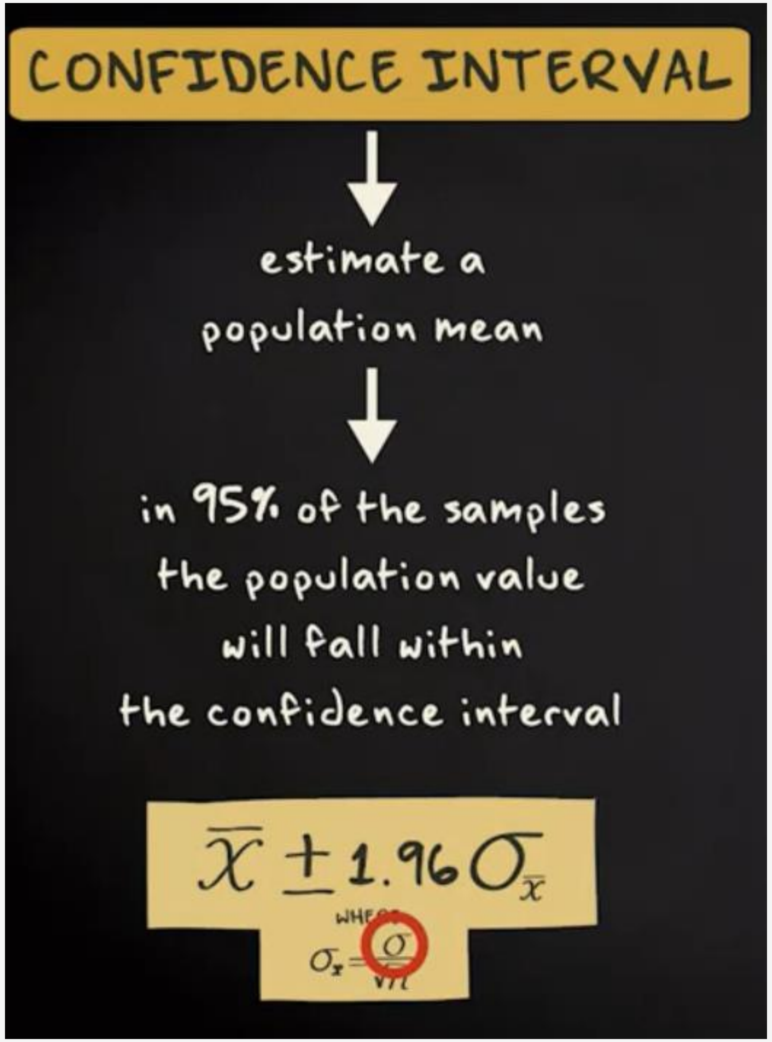


CI for mean with unknown population SD

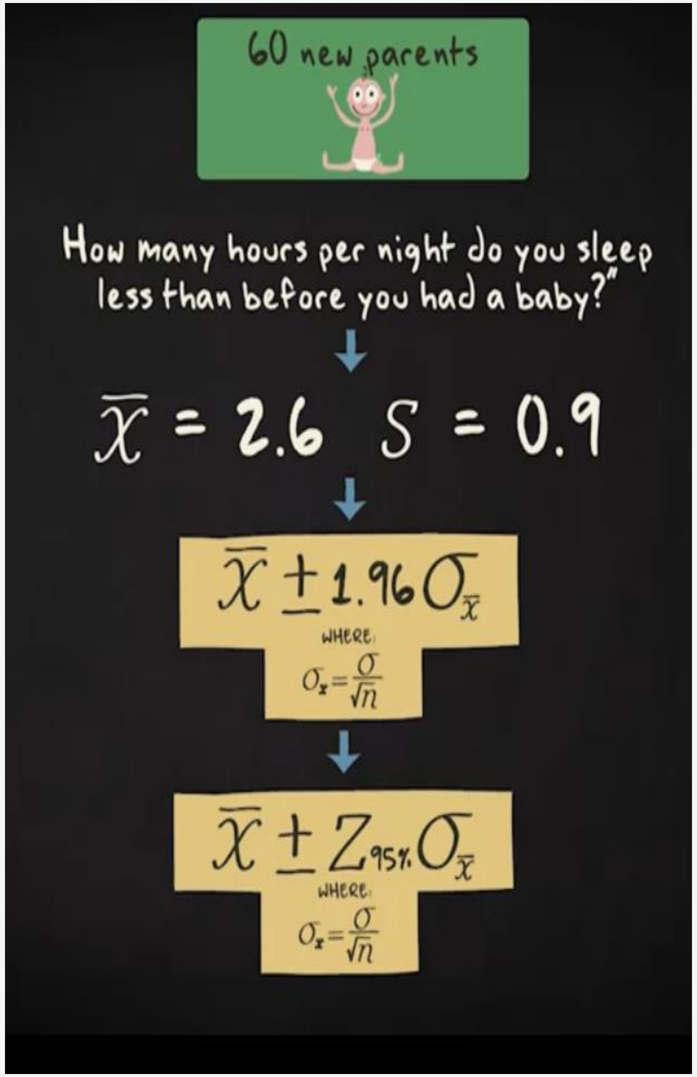
There is one big problem with this formula, however. To compute the confidence interval, you need to know the **population standard deviation**, and usually, we don't know this value. After all, we use the sample to draw inferences about our population parameters.

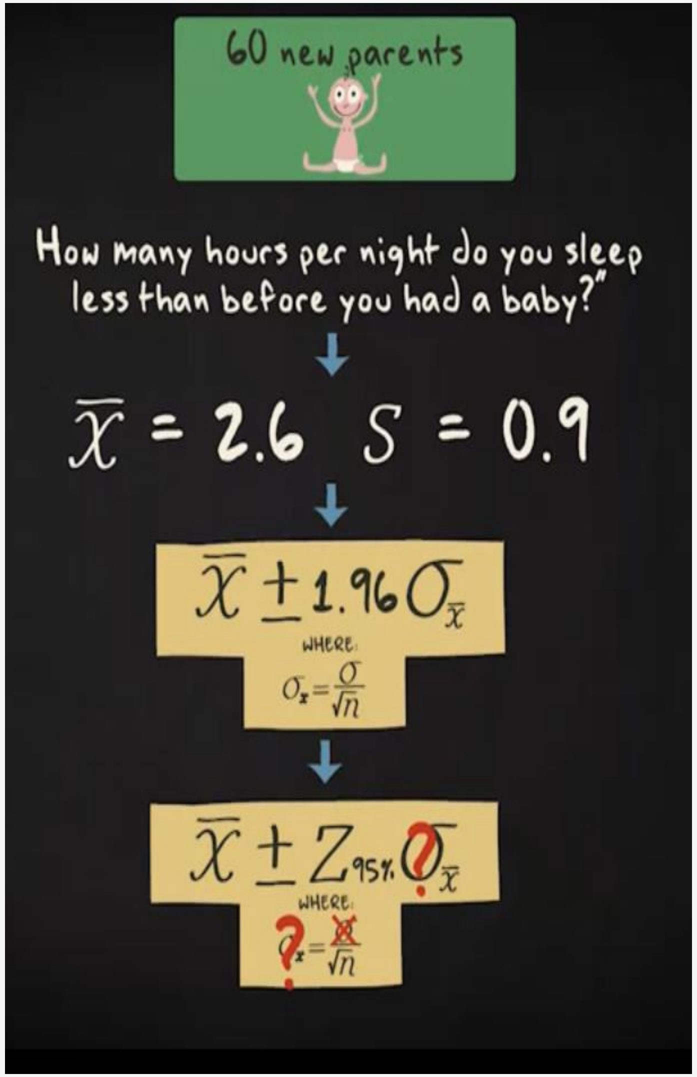
We'll show you how you can solve this problem. The solution is that we estimate the population standard deviation, and therefore have to employ another distribution than the standard normal distribution: the **t-distribution**.

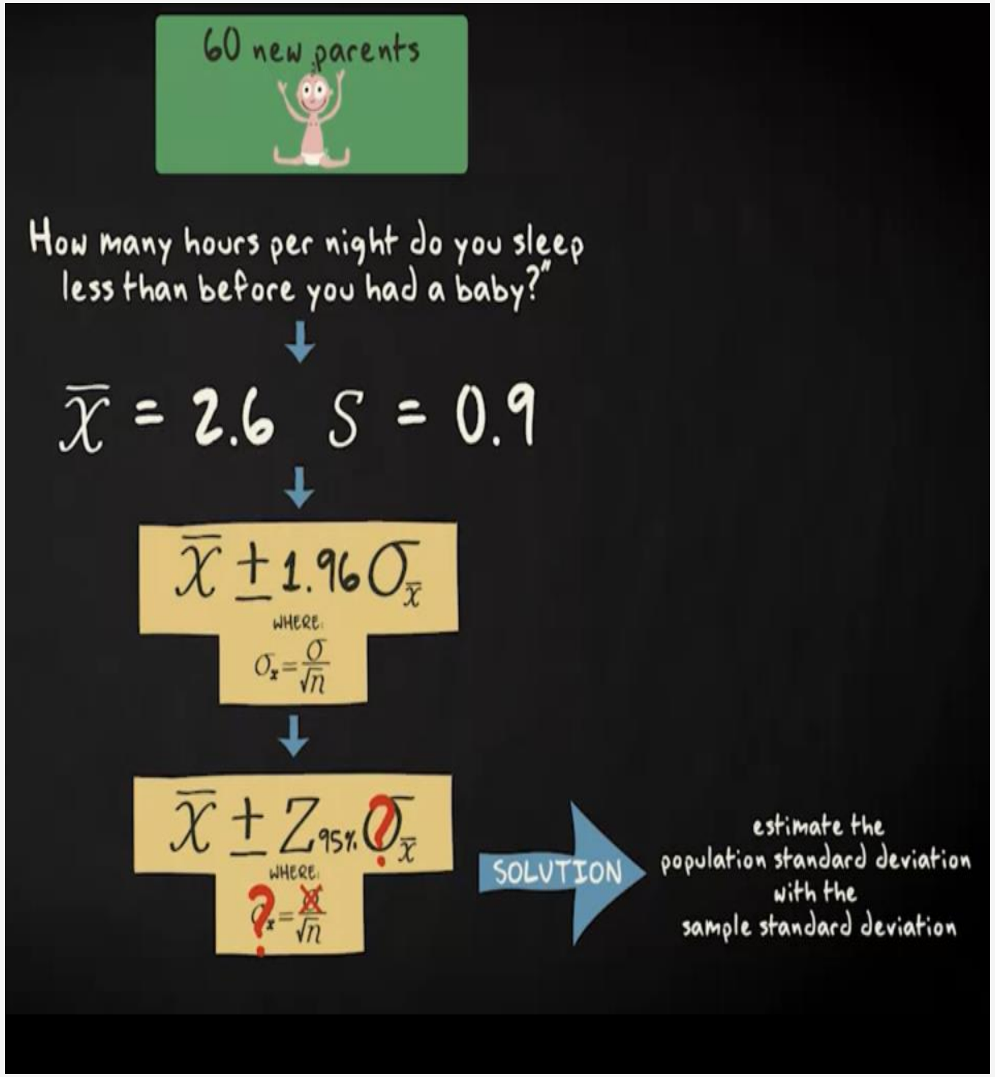
Let us tell you how that works.



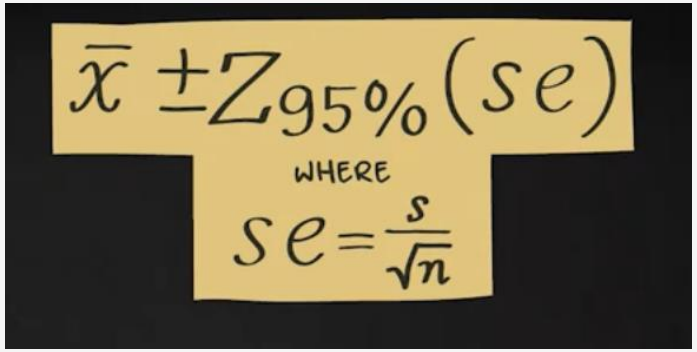
CI for mean with unknown population SD

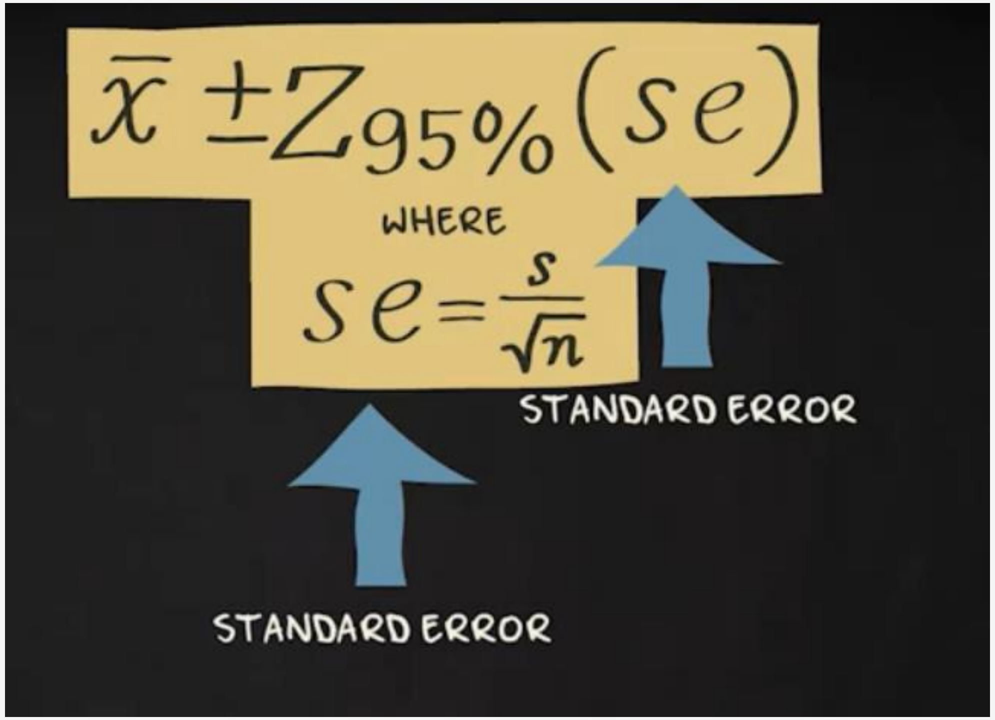




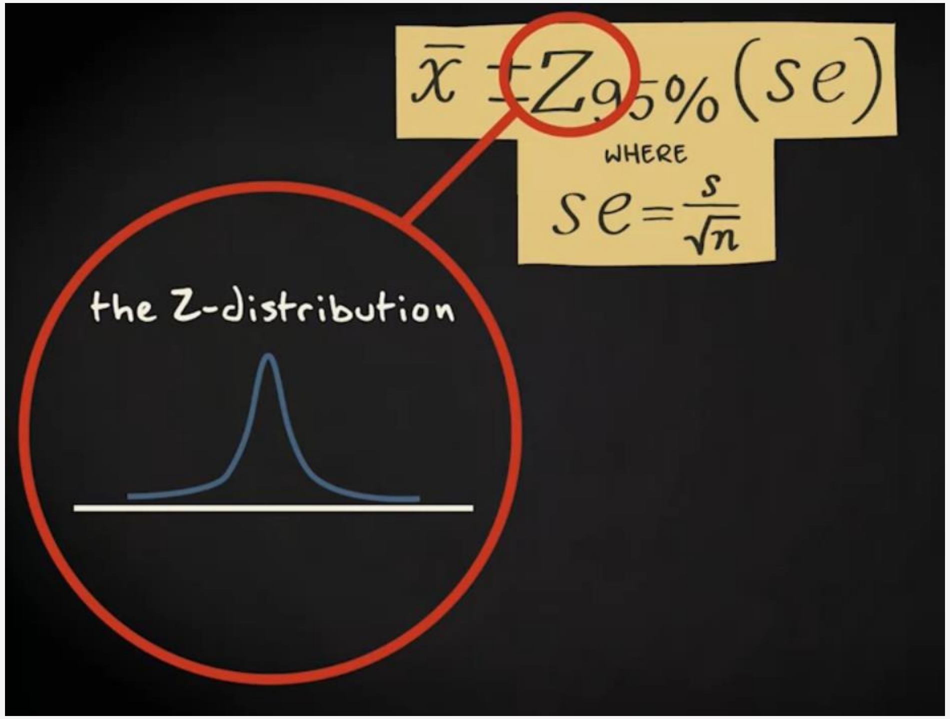


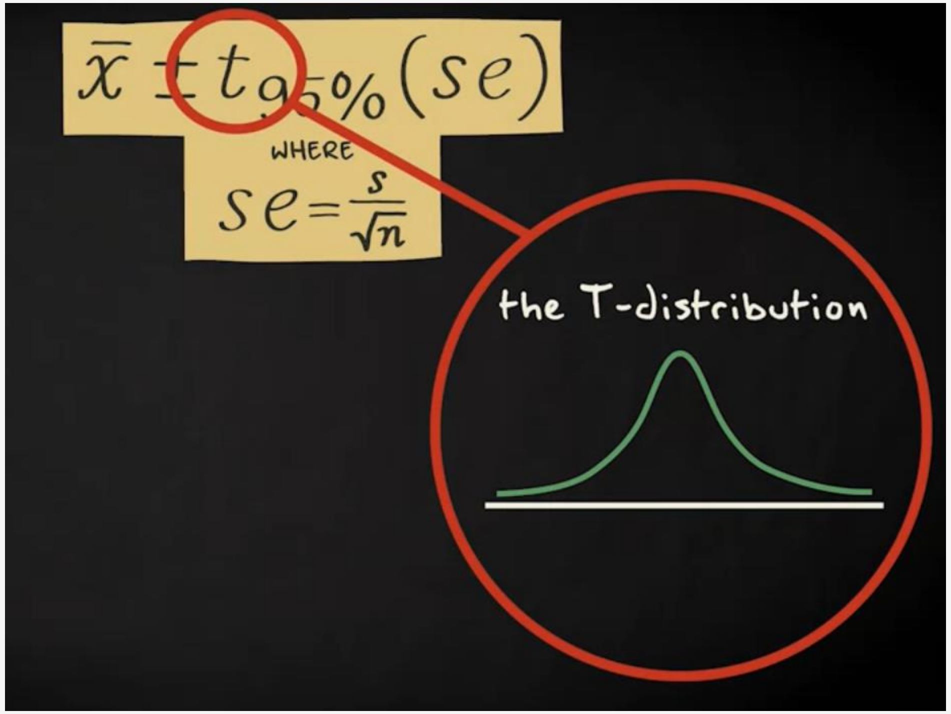
CI for mean with unknown population SD





CI for mean with unknown population SD





CI for mean with unknown population SD

Let me now tell you a little more about t-distributions and t-scores. The t-distribution strongly resembles the standard normal distribution. It is bell-shaped, symmetric and has a mean of zero.

Still, however, it is slightly different. Because we now estimate the standard deviation of the sampling distribution we introduce extra error.

That error can be substantial when we have a small sample. The t-distribution takes into account that extra error for small samples.