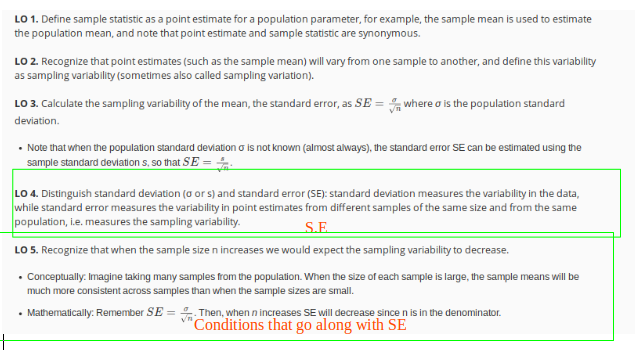
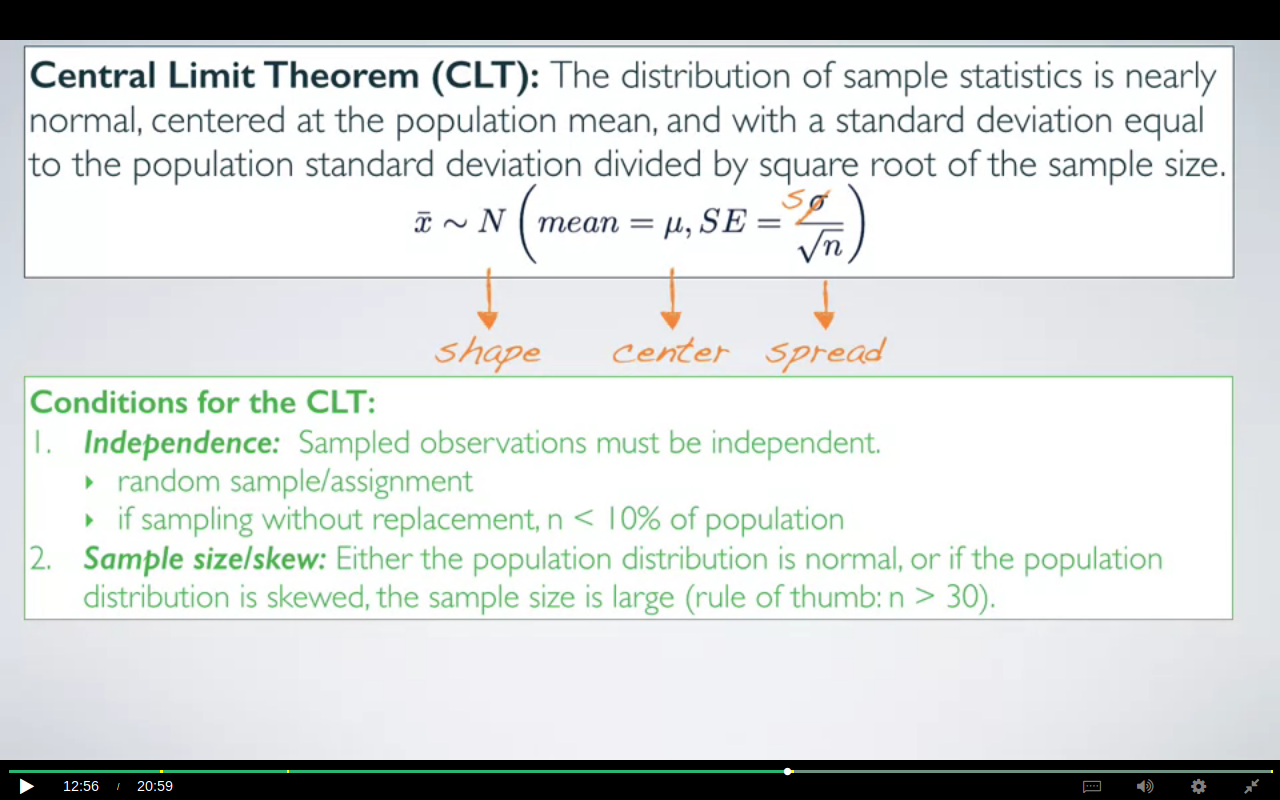
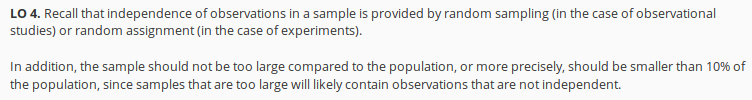
Week 1

S.E=std of error of mean is less than the pop std ,because we expect the means of sampling dist to be less variable than the pop



<https://gallery.shinyapps.io/CLT_mean/>

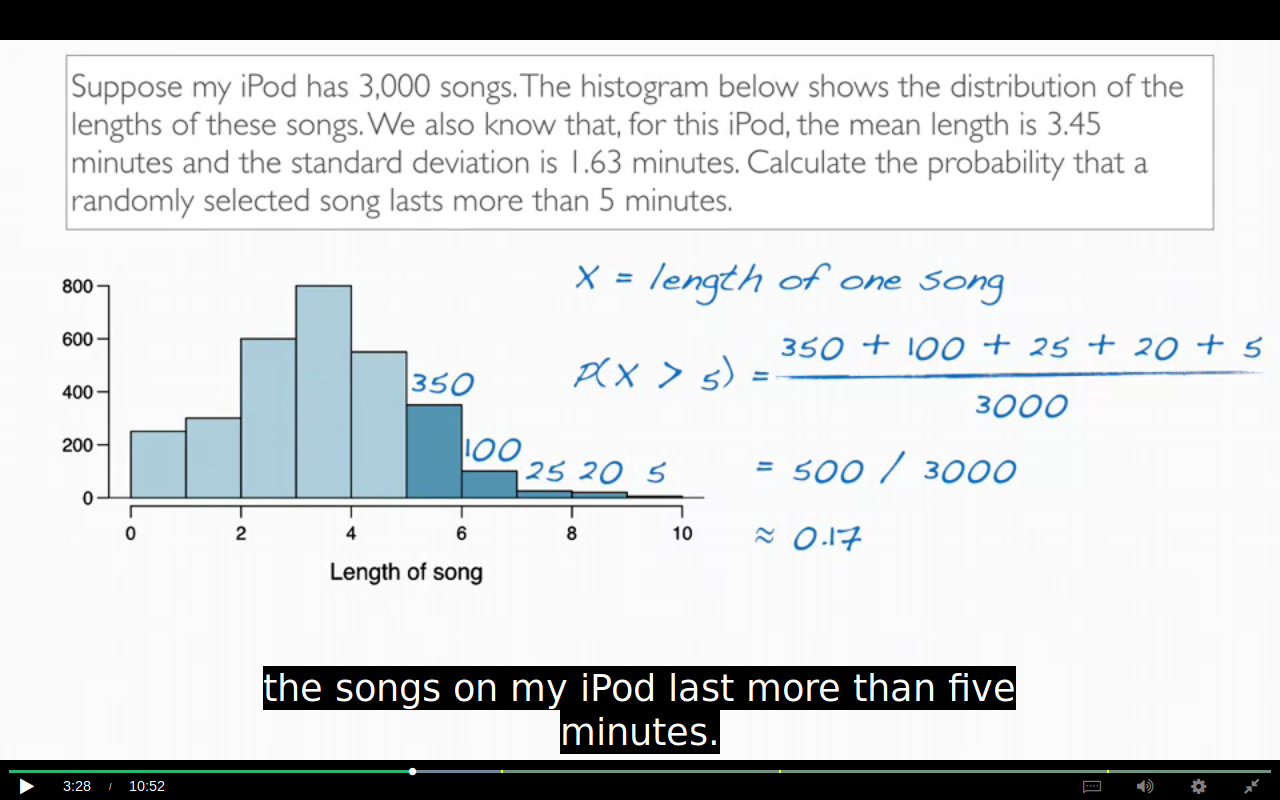
***THE SECOND CONDITION HERE IS AT MOST 10% AND NOT ATLEAST 10%***



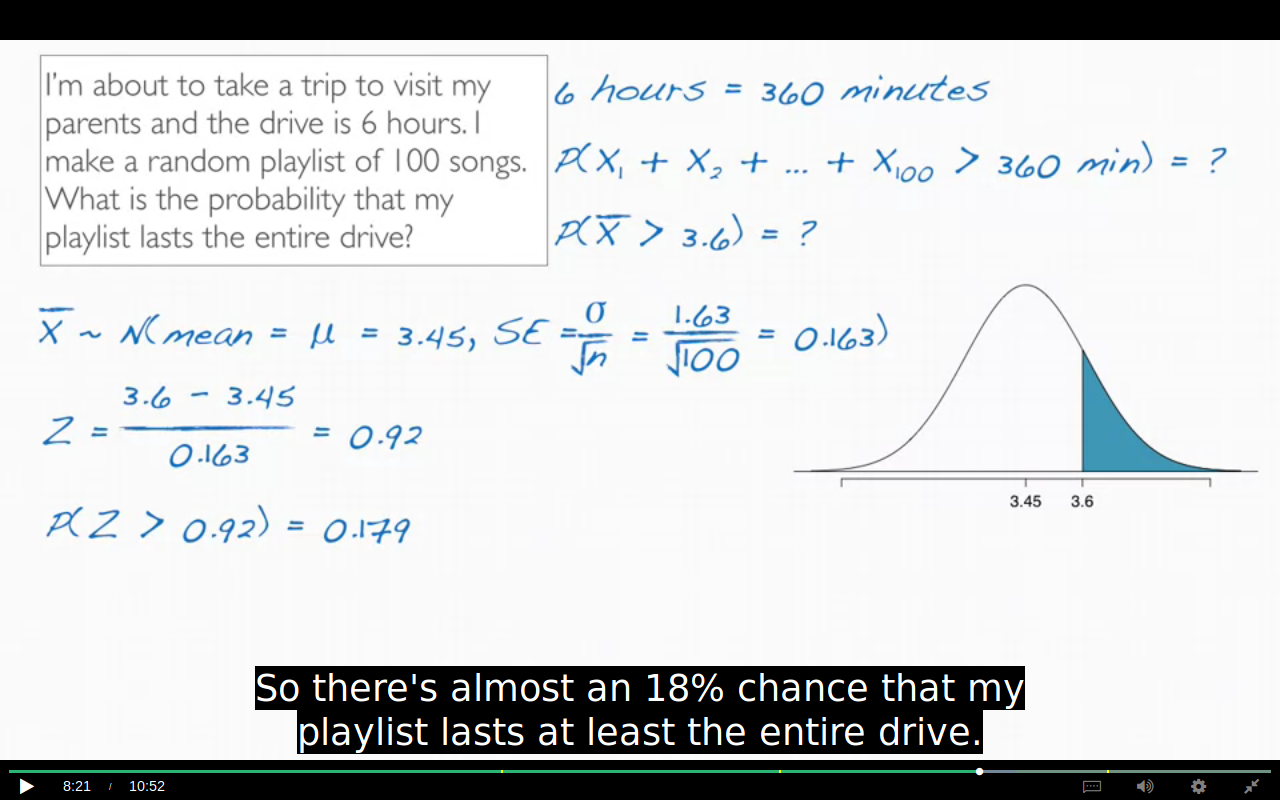
if the pop dist is normal then the sampling dist is also going to be normal regardless of the sample size however if it is not then we would have to take the sample size>= 30. the more skewed a dist is the larger the s.s will be needed in order to follow the CLT.

Often times We do not know what dist our population is following,so for those scenarios we plot our sample data and assume that both follow the same dist.

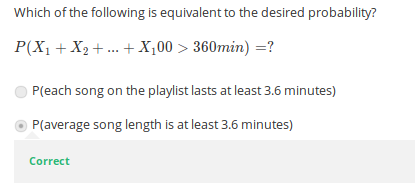
We need to concentrate on sample dist before doing any calculations ,for instance the dist of the incomes can’t be the normal as there’s no certain limit to what a person can earn.



The methods for calculating the prop using Z table can;t be applied here as the dist is right skewed.



Think why this is true ?



## Q. When do I use sigma /sqrt(n)?

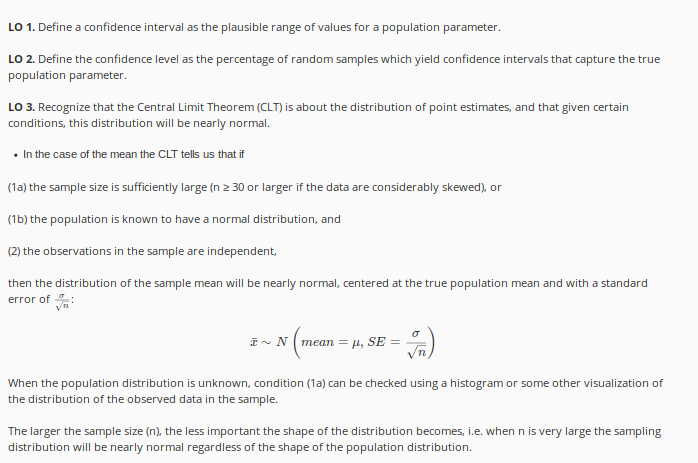
A. You always divide by sqrt(n). However, occasionally the square root of n sometimes equals 1 (making it just σ in the denominator. for example, if you are choosing one person and trying to figure out the probability their weight is under x pounds, then n=1. In other words, if you are calculating a [z-score](http://www.statisticshowto.com/probability-and-statistics/z-score/), you can always use √(n).

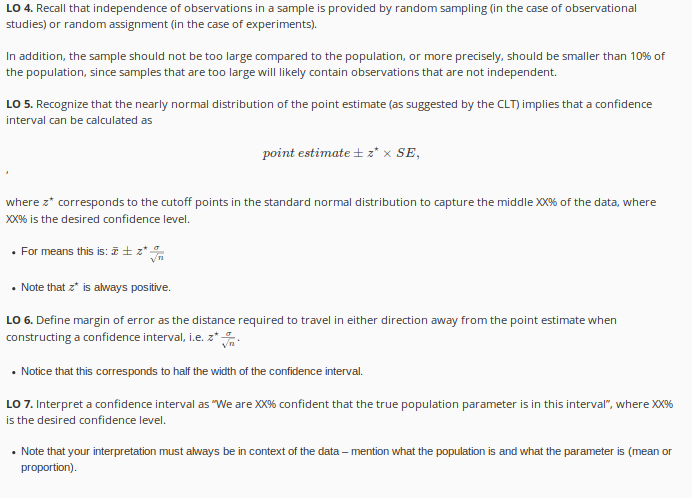
We measure the variability of individual observations with standard deviations. We measure the variability of sample means with standard errors.So whatever the observation is that you plug in in

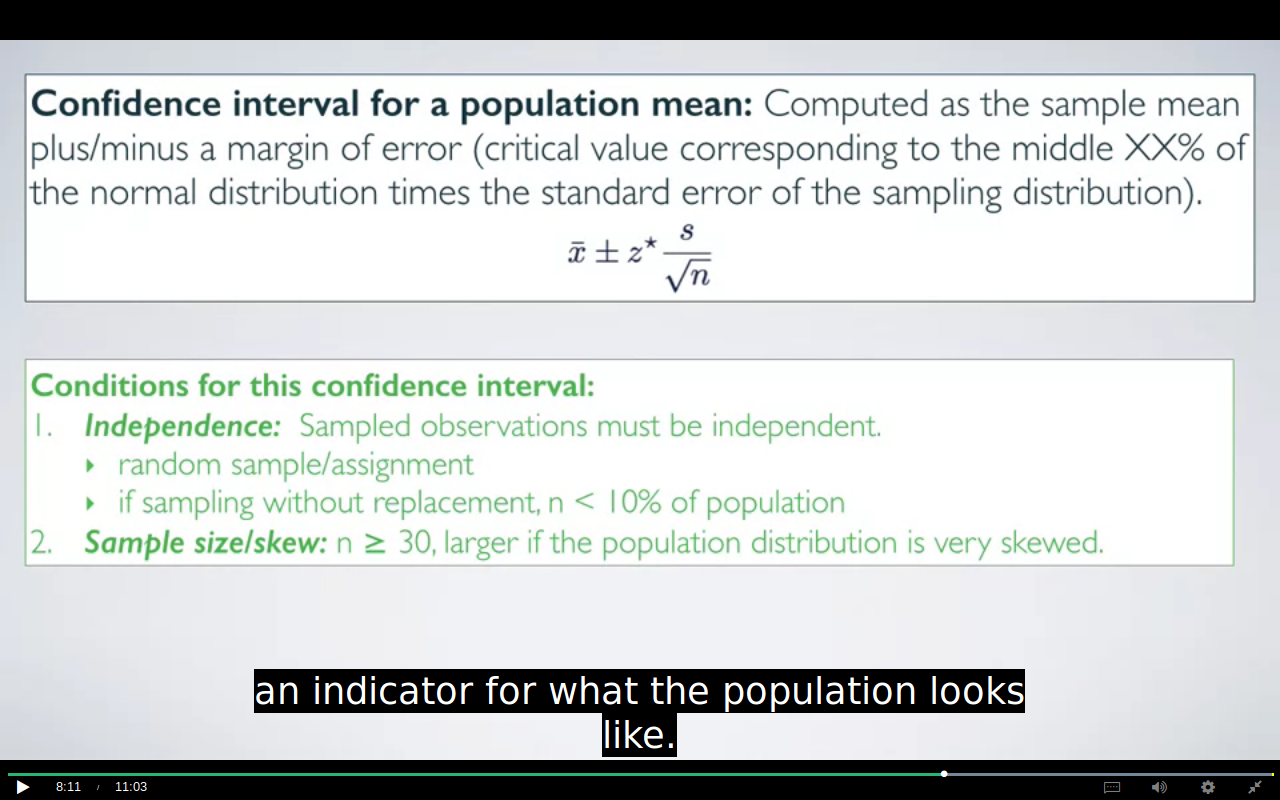
the numerator in your Z-score, its variability belongs in the denominator.

In other words, our observation is an X bar, and not an X.

**CONFIDENCE INTERVAL:**

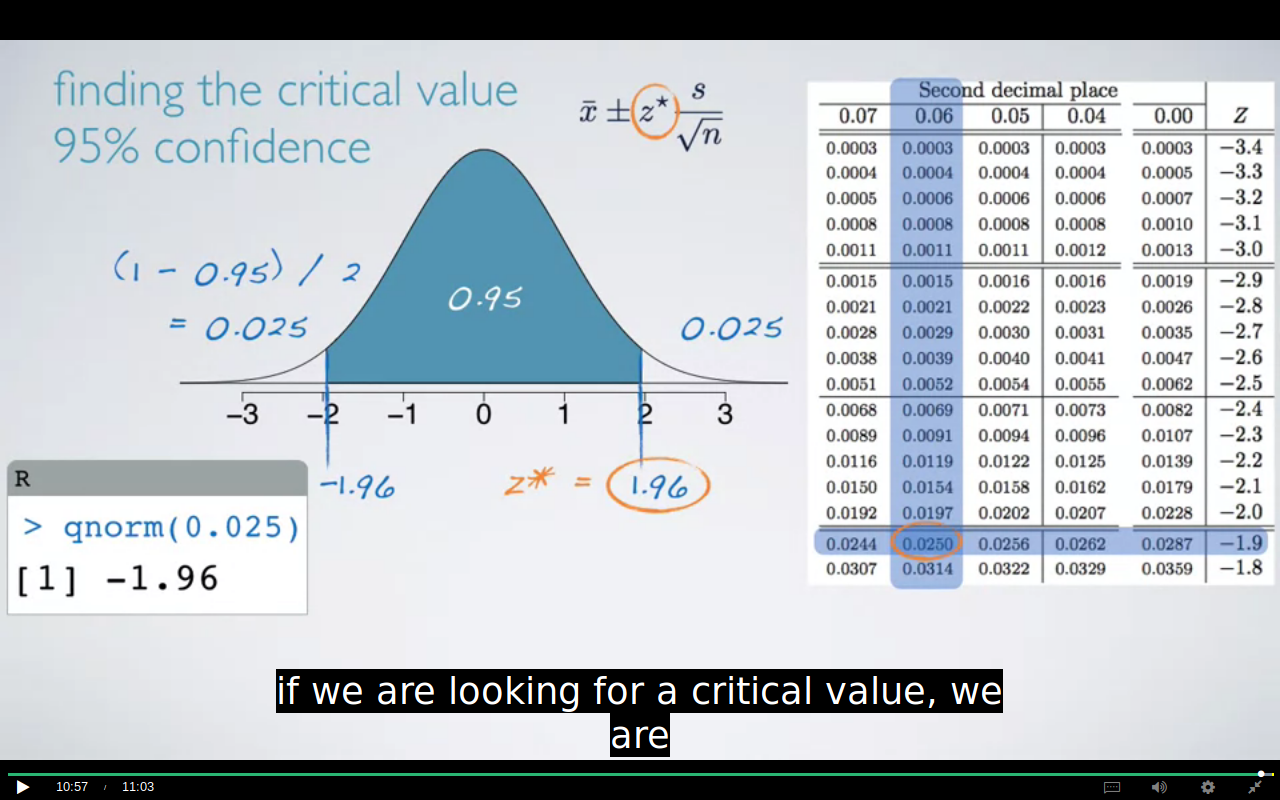


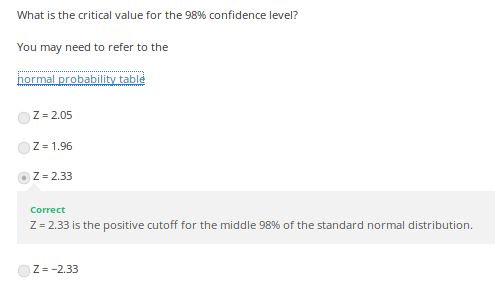




THE CONDITIONS ARE SAME AS OF CLT BUT THE SECOND IS MORE STRICT HERE

***THE SECOND CONDITION HERE IS AT MOST 10% AND NOT ATLEAST 10%***





Given that the critical value for a 95% confidence interval is 1.96, for a 98% confidence interval is 2.33, and for a 99% confidence interval is 2.58, what happens to the width of the confidence interval as the confidence level increases (all else held constant)?

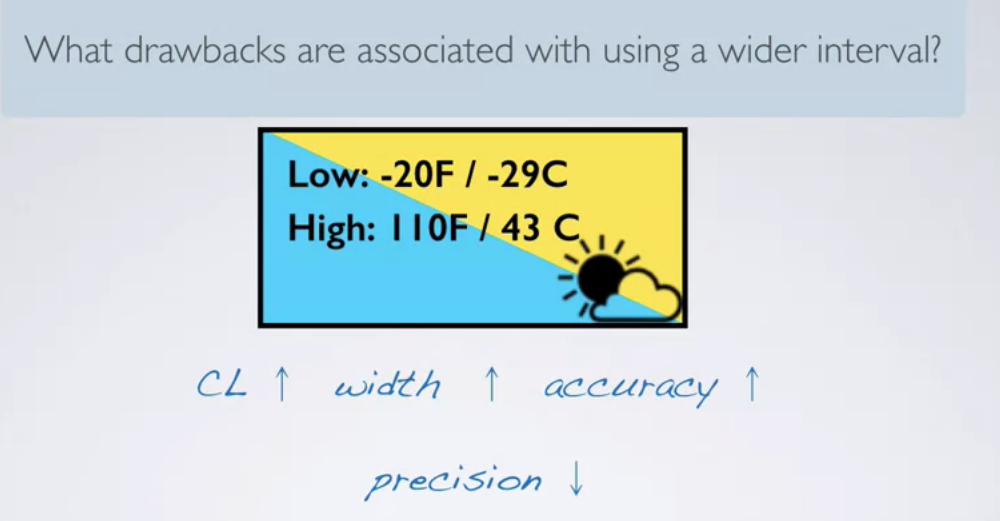
As the confidence level increases the interval gets wider.

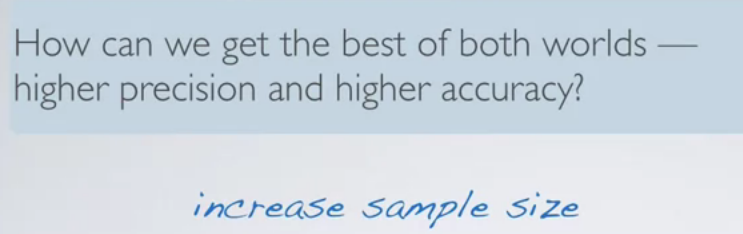
Correct

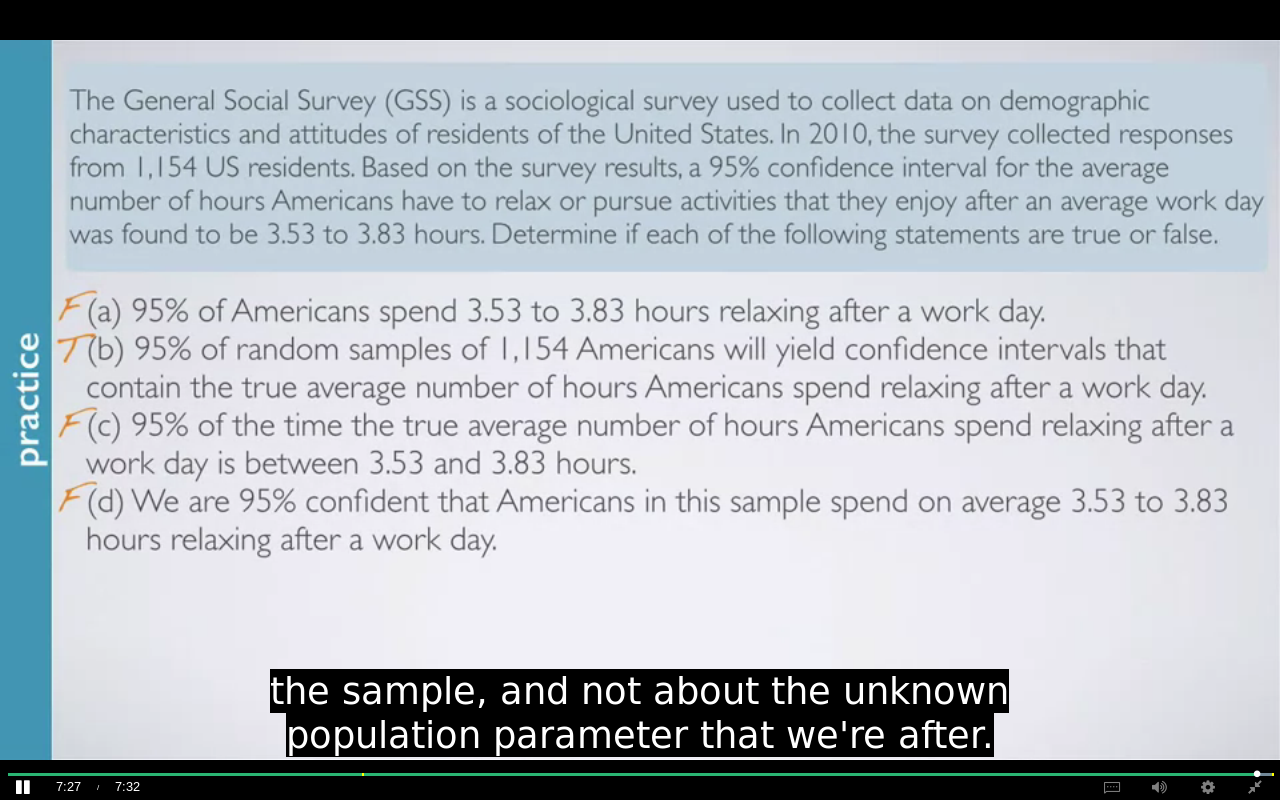
The higher the confidence level, the larger the critical value, hence the larger the margin of error, and hence the width of the confidence interval.

As the confidence level increases the interval gets narrower.

As the accuracy increases the precision decreases



as we increase our sample size the SE decreases and hence our precision will not decrease drastically.



MOST IMPORTANT PART

A says that 95% of Americans spend between

3.53 to 3.83 hours relaxing after a work day.

This is not true because remember that the confidence interval is not about

individuals in the population but instead about the true population parameter.

B says that 95% of ***random samples of 1,154 Americans(and not one individual)***

will yield confidence intervals that contain the true average

number of hours Americans spend relaxing after a work day.

This is indeed the definition of the confidence level.

The percentage of random samples that will yield confidence intervals

that contain the true population parameter.

So this is true.

C says 95% of the time the true average

number of hours Americans spend relaxing after

a work day is between 3.53 and 3.83 hours.

This is not true because the population parameter is not this moving target

that is sometimes within an interval and sometimes outside of it.

And lastly, D says we are 95% confident that Americans ***in this sample***

spend on average 3.53 to 3.83 hours relaxing after a work day.

This is not true because remember that the confidence interval

is not about the sample mean, but is instead about the population mean.

We know exactly what the sample mean is.

It has to be between these values because we construct the confidence interval

around the sample mean.

Therefore, we could actually say that we are 100% confident that Americans

in this sample spend on average between 3.53 and

3.83 hours relaxing after an average work day.

EXAMPLE:

The General Social Survey (GSS) is a sociological survey used to collect data on demographic characteristics and attitudes of residents of the United States. In 2010, the survey collected responses from 1,154 US residents. Based on the survey results, a 95% confidence interval for the average number of hours Americans have to relax or pursue activities that you enjoy after an average work day is 3.53 to 3.83 hours. Which of the following is false?

Increasing the confidence level would result in a more accurate but less precise confidence interval.

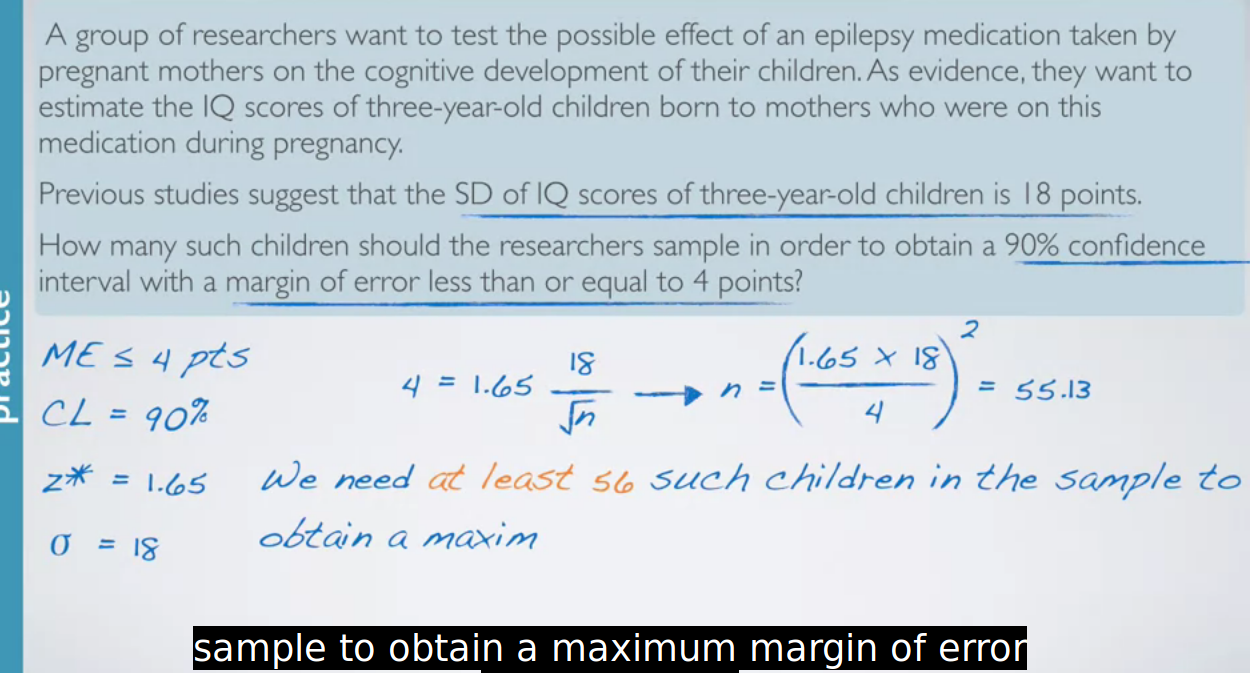
The standard error is approximately 0.075 hours.

The sample mean is 3.68 hours.

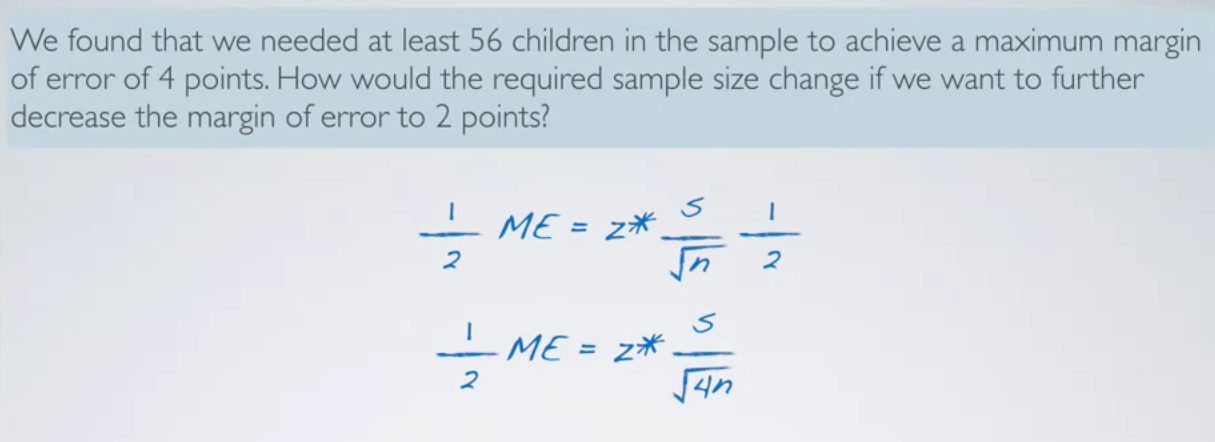
The margin of error is 0.3 hours.

Correct

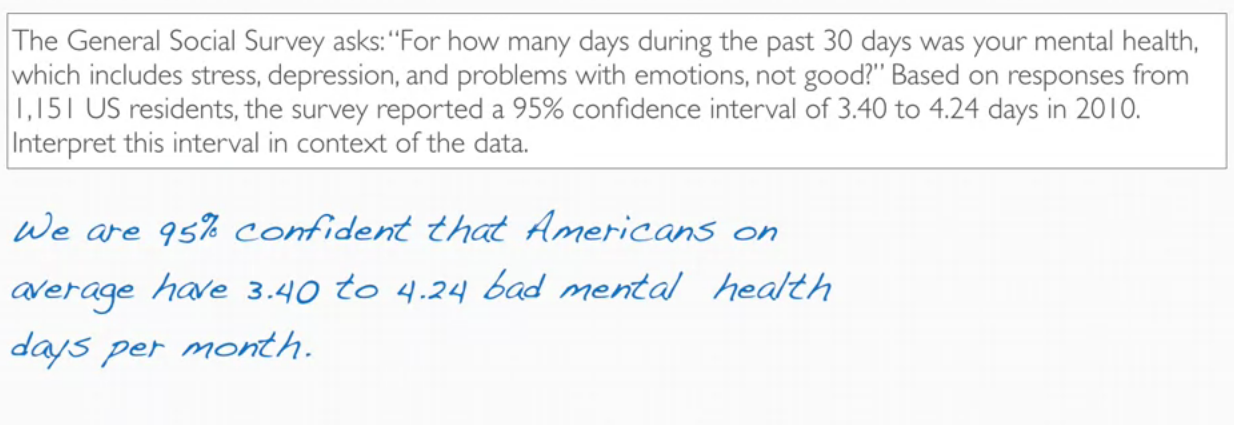
The margin of error is (3.83 – 3.53) / 2 = 0.3 / 2 = 0.15 hours.

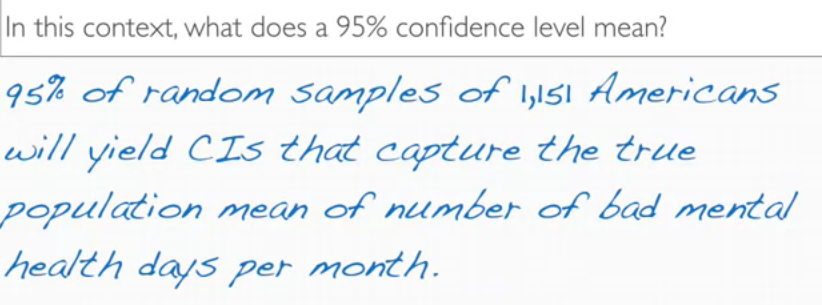


Since 55..13 was minimum we will round it to **56**



the reation between **n** and **me** is inverse and exponential as well as we have to take the square of the number by which we want our **me** to be decreased.





Which of the following is a condition that needs to be met to calculate a confidence interval for a population mean using methods that rely on the Central Limit Theorem?

A) The population distribution must be nearly normal.

B) At least 10% of the population must be sampled.

C) The sampled observations must be independent with respect to the variable in question.

**Correct**

The population distribution does not necessarily need to be nearly normal, as the CLT will hold if the sample size is large even if the population distribution is skewed. At most, not at least, 10% of the population must be sampled. The success-failure condition is useful for categorical variables, not numerical. So the only choice that is correct is “The sampled observations must be independent with respect to the variable in question.”

D) There should be at least 10 successes and 10 failures.

