Statistics 16_est_parameters.pdf	Michael Hallstone, Ph.D. hallston@hawaii.edu

Lecture 16: Estimating Parameters (Confidence Interval Estimates of the Mean)

Some Common Sense Assumptions for Interval Estimates

- The variable used is appropriate for a mean (interval/ratio level). **(Hint for exam: no student project should ever violate this nor have to assume it. Your data set will have this sort of variable.)**
- The data comes from a random sample**. (Hint for exam: all student projects violate this assumption.)**

- If the sample size is greater than 30 (n>30) use Z distribution. Statistical theory says that if the population is known to be normal you can use Z when regardless of sample size, but you should ignore theory in this case. In practice if the population is known to be normal and the sample size is small, not around 30, it is better to use the t distribution instead of Z -- it's more conservative. If n<30 and population is unknown use t distribution. If n<30 we ALWAYS assume population is normal. *So in plain English when n is <30 we assume that the test variable is normally distributed in the population. If you test "mean age" then you assume age is normally distributed in the population.* **(Hint for exam: if your n<30 you will make this assumption!)** *This is the only way that the sampling distribution of means is normally distributed (when n<30).*

Introduction

First of all let me say that this lecture, along with the one on the Central Limits Theorem are probably the most important lectures in the whole course. If you understand these two lectures you will understand the whole basis of inferential statistics – that is how we are able to infer from sample data to a population. If you understand these two lectures, you will be able to "figure out" the basis of the remaining statistical techniques we'll study in this course.

When we wonder about the social world we are actually wondering about characteristics of a population. Thus, we are starting to formulate **research questions**. For example we may wonder:

- What is the mean age of people in prison in the US?

- What is the mean number of days needed to complete "successful" drug treatment program for released prisoners?

- What is the mean monthly income of clients served by a public health clinic?

- What is the mean time (in minutes) spent in line at customer service line at a C&C City Hall office?
- What is the mean age of clients served at one branch of a YMCA agency?

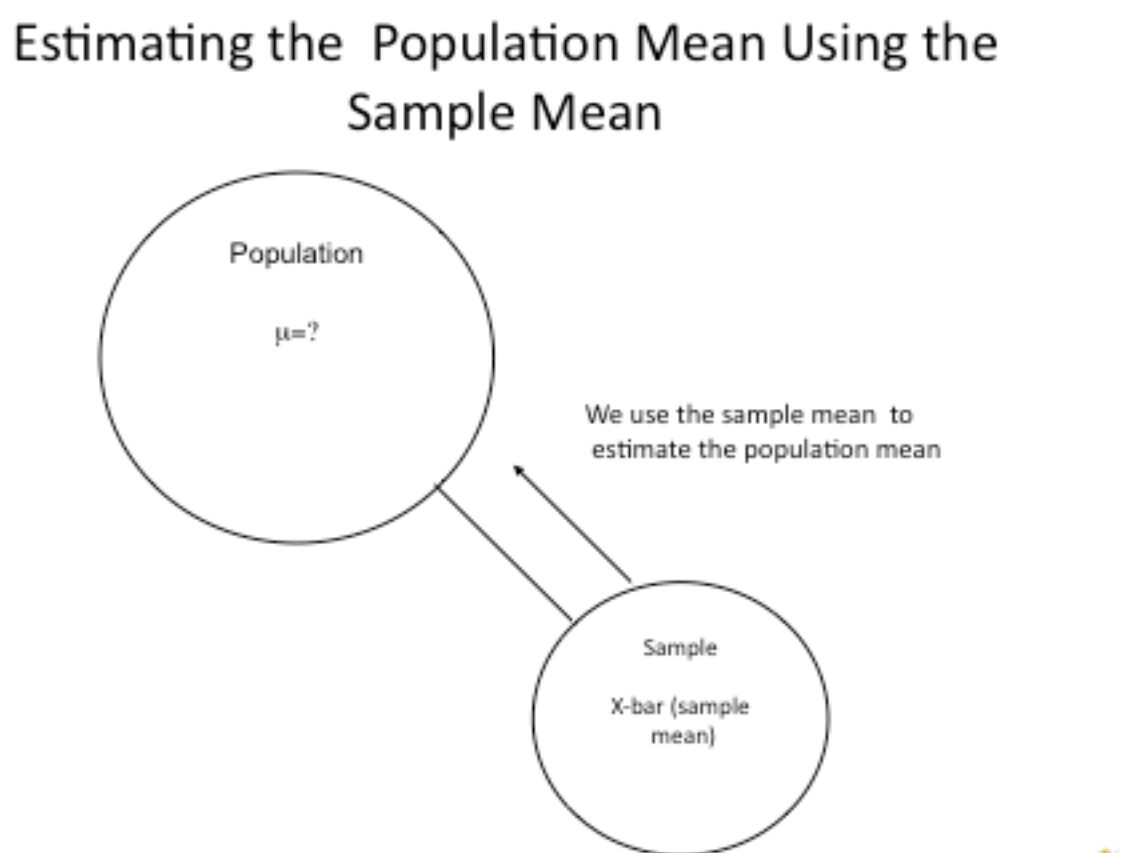
When we think about it a bit deeper all of these questions really wonder about characteristics of a population. These are the sorts of questions researchers ask, study, and write papers (although to be fair their questions are rarely as simplistic as only looking at the population mean).

But, if we had the data, all the research questions above could be answered using the technique taught in this lecture. This lecture and the following lectures (one sample hypothesis tests & two sample hypothesis tests) will allow us to answer questions regarding the population mean or means.

<u>The WHOLE REASON you are in this class is so that you can learn to formulate research questions and answer them using statistics!</u> Since we are just learning statistics we will start with very simplistic research questions.

Goal for this lecture: estimate the population mean

The whole gig in statistics is to estimate some characteristic of the population based upon data from a sample. (The population is what we are interested in studying!) Remember all the way back to the first lecture in this class and that funny circle diagram?



<u>In this lecture we will learn a technique called "confidence interval estimates" or "interval estimates" where we will learn to provide an estimate of the population mean using an interval</u>. In plain English, we will learn to come up with a spread of values (an interval) that estimates the population mean. So for example we will eventually make an estimate like this: I am 95% confident that the population mean is between 25.1 years of age and 29.5 years of age. See how we make an interval estimate of the population mean? We are saying that we are 95% confident that the population mean age falls between those two numbers. That is where we will get to by the end of this lecture.

Some terms and concepts needed for background

To estimate the population parameter we need some number in the sample that allows us to ***<u>estimate</u>*** the population. (That # is a statistic! ="a statistic is a number that describes a sample characteristic" this is a quote from the very first lecture – see how statistics is higher order learning!?)

<u>Estimator</u> = "any statistic used to estimate a parameter (population characteristic)"

$\bar{x}$ (sample mean) is estimator of μ (population mean).

We need an **<u>unbiased estimator</u>** = if the expected value of a statistic equals the population mean then that statistic is unbiased. (Sample means, sample percentages, and sample variances are all unbiased estimators, *<u>but in this lecture we shall only be using sample means to estimate population means.</u>*

Formal Definition of an Interval Estimate and Point Estimate

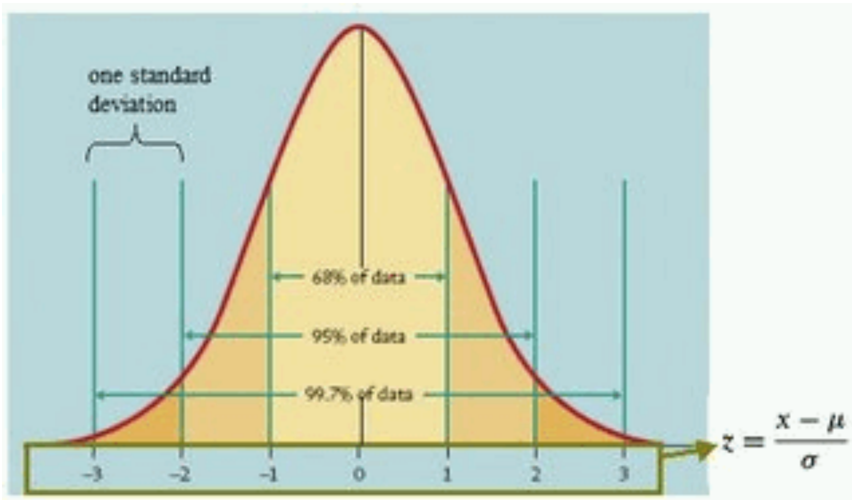
A ***point estimate*** = a single number used to estimate a population parameter. If you use the sample mean alone (one number) to estimate the population that is a point estimate. We will cover point estimation in other lectures.

An ***interval estimate*** = a spread of values used to estimate a population parameter and process of using these spread of values is called ***interval estimation.*** *We actually use the point estimate to do interval estimation.* In fact, point estimate is adjusted for sampling error to create interval estimate. In plain English we will use the sample mean (point estimate) to create a spread of values (interval estimate) that should include the population mean. Again, and interval estimate will end up "saying" something like "I am 95% confident that the population mean is between ____ (a number) and ___(a number)."

The Logic Behind Confidence Interval Estimates Comes From the Central Limits Theorem

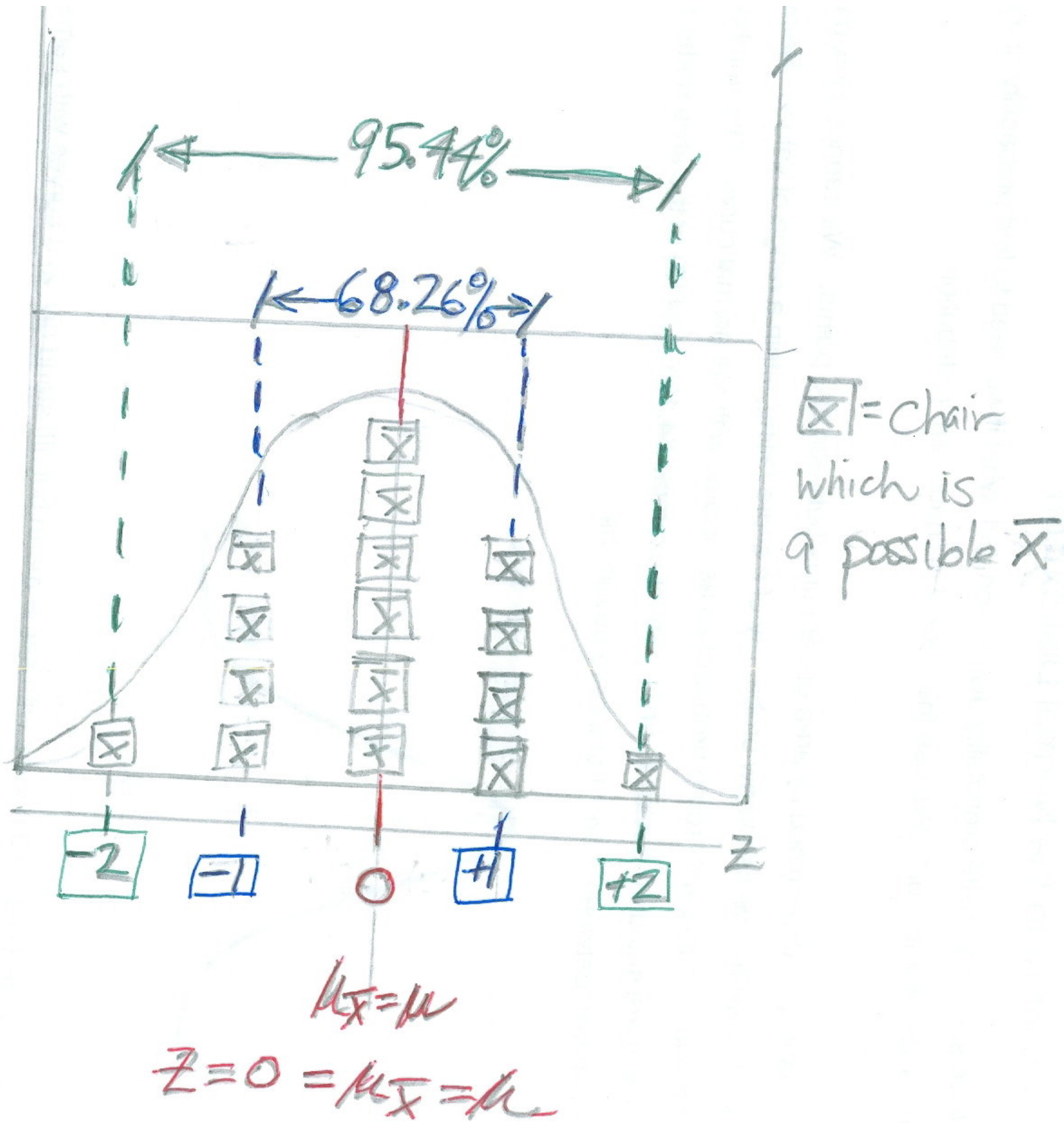
Using the logic of the central limits theorem we have seen that the sample mean is a pretty good estimator of the population mean. The chance of drawing a sample mean that is "close" to the population means is pretty good. Recall the following example modified from the lecture on the Central Limits Theorem:

We do not know which sample mean we will draw but (if the sampling distribution of means is normally distributed) the probability of our sample mean falling ±1 SD from the "true or real" population mean is 68.26%. Or the probability of our sample mean falling ± 2 SD from the mean is 95.44%. Remember the pictures from z-scores?



source: http://www.sci.sdsu.edu/class/psychology/psy271/Weeks/psy271week06.htm

Here is another picture of the same basic idea:



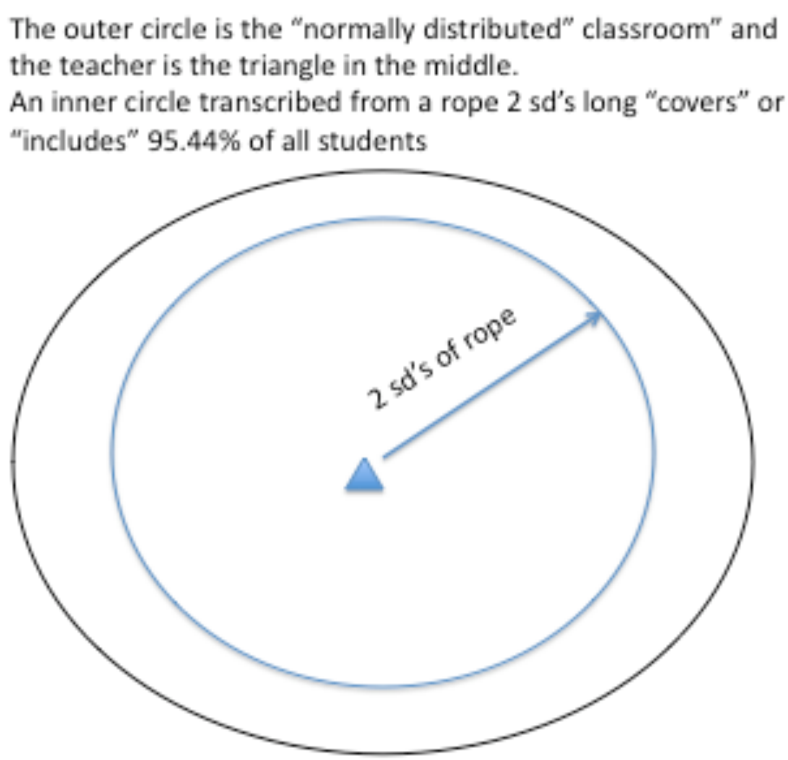
Interval Estimation and Confidence Intervals

Go back to sampling distribution of means and Central Limits Theorem. When the sample size is greater than 30 or when the population itself is normally distributed, <u>the sampling distribution of means follows a normal distribution</u>, clustered around the population mean. Applying what we know about the probabilities associated with a normal distribution, 95.4% of the time the sample mean will fall within plus or minus 2 standard deviations from the mean.

Another way of looking at it (from book): if you were to take 1000 samples then 954 of them would fall within 2 SDs from the μ (population mean). If that doesn't make sense you didn't understand the central limits theorem. (If you do not understand it, you need to go back and re-visit the lecture, readings, find out where you are lost and come get some help from me!)

Another way to look at it is to use (Liem's) example of the teacher standing in the middle of the room. Pretend the teacher is standing in the middle of the room and all the students are 'normally distributed' from the teacher. If the teacher were to let out 2 SDs length of rope, then have a student hole the end of the rope and walk in a circle around the teacher, the then 95.44% of the students would be "in the 2 SDs circle."

In the diagram below the outer circle is the "normally distributed" classroom" and the teacher is the triangle in the middle. An inner circle transcribed from a rope 2 SDs long "covers" or "includes" 95.44% of all students

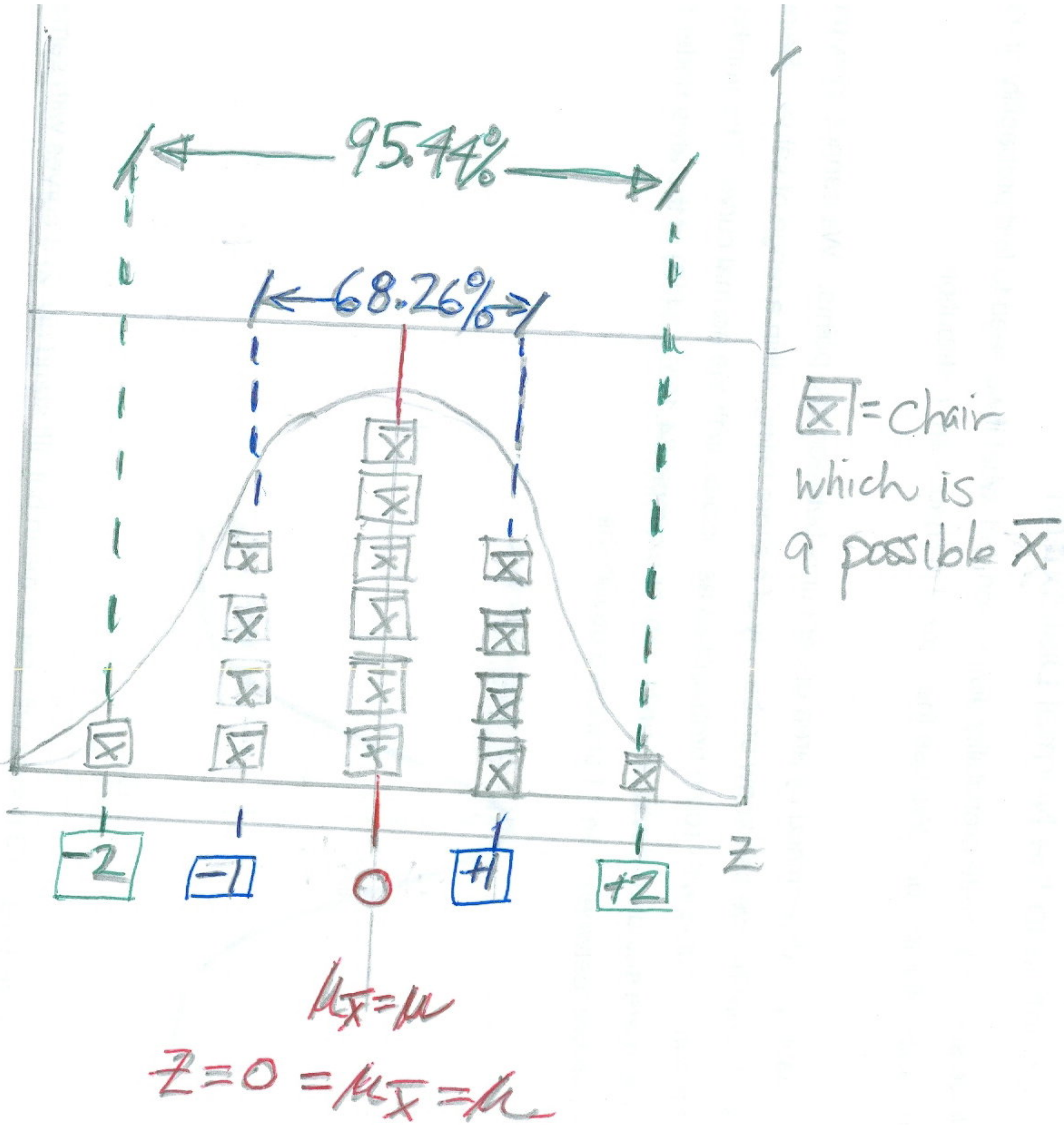


Another way to look at this. Pretend we use a rope with a length of 2 SDs and make a circle around the teacher. Pretend that 95% of the students in the class will be within his rope circle. That means that there's a 95% chance that the teacher will be within 2 SD of any student. It also means that there is a 95% chance any student chosen at random will be "2 SDs or closer" to the teacher. So "the distance" question works both ways -- for the teacher or any student.

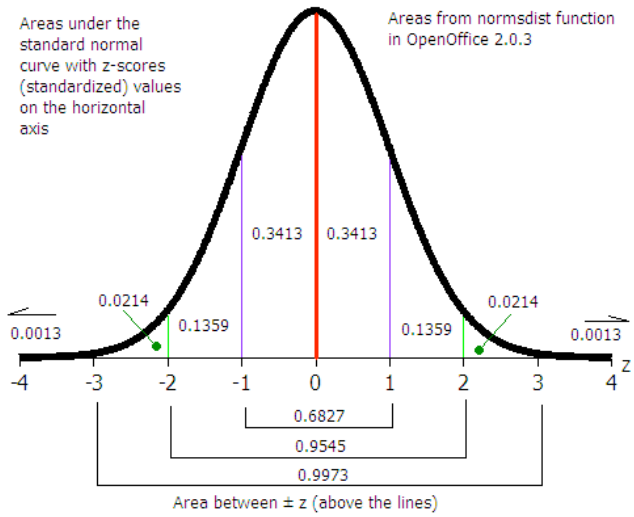
Chairs arranged in the shape of a normal curve example.

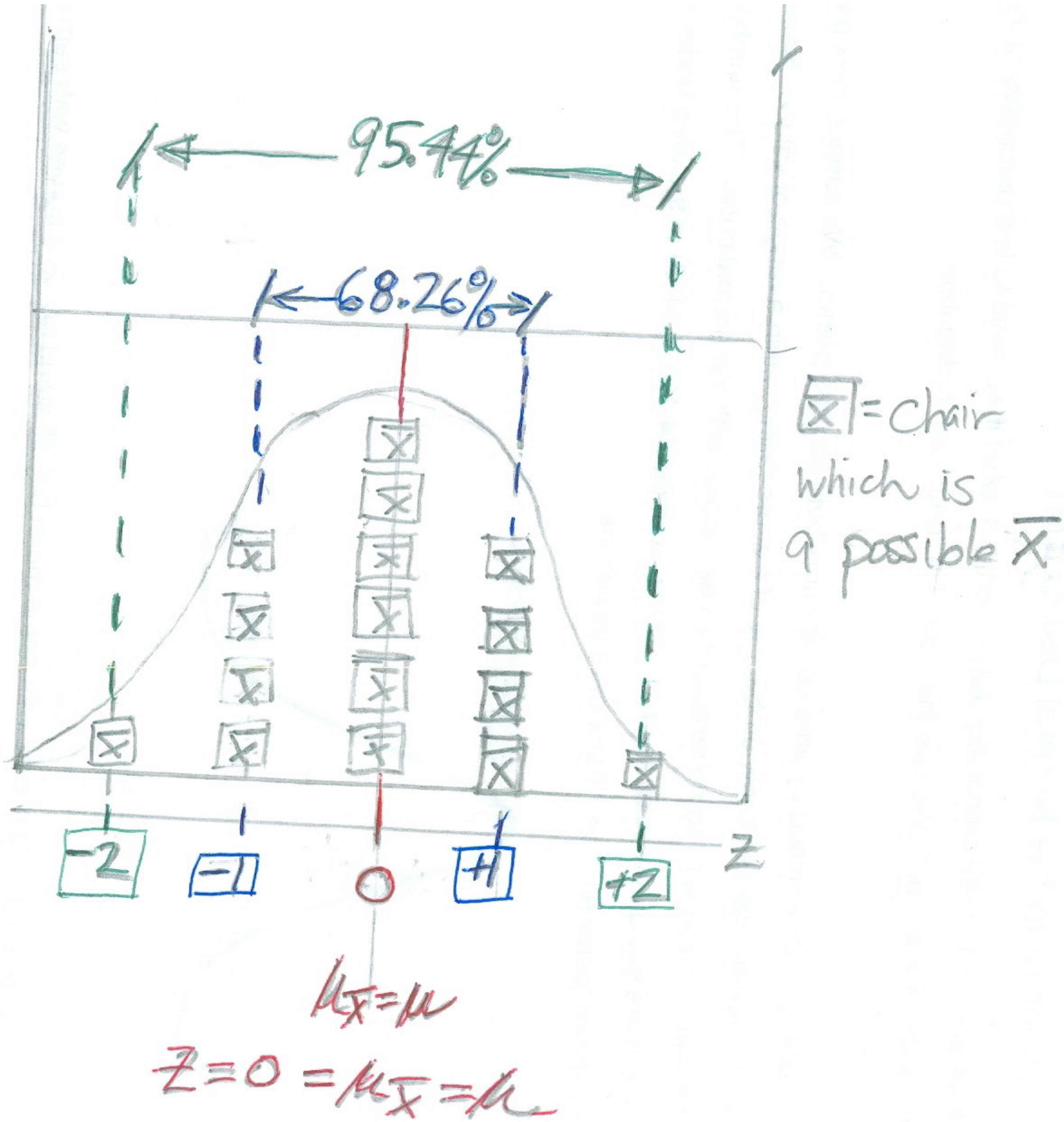
Pretend you were in a classroom and you could arrange the chairs in the room to be in the form of a normal distribution. Pretend each seat represents an individual sample mean. As a whole, all of the seats in the classroom (each seat is an individual sample mean) represent the "sampling distribution of means." (Recall that the sampling distribution of means = all the possible sample means that are possible give your population size and your sample size.)

I'm sorry UHWO does not have software for "drawing pictures on the computer," but below is my cheap hand drawn picture of the example of chairs in a classroom representing the sampling distribution of means



Here is another picture that you can use to represent the same idea with some imagination. In the diagram below, there are rows of seats at -3 SD, -2 sd (green line), -1 sd (purple line), 0 sd (red line), +1 sd (purple line), +2 sd (green line), and +3 SD. So the row of seats on the red line are exactly equal to the population mean.


Areas under the standard normal curve with z-scores (standardized) values on the horizontal axis
Areas from normsdist function in OpenOffice 2.0.3
0.3413
0.3413
0.0214
0.1359
0.0013
0.0214
0.1359
0.0013
-4
-3
-2
-1
0
1
2
3
4
z
0.6827
0.9545
0.9973
Area between ± z (above the lines)



So in our classroom, each of the seats represent a possible sample mean we could draw from the sampling distribution of means. As you can see 68.26% of all sample means fall no further away from the population mean than plus or minus one SD (or z-score unit). Conversely, 95.44% of all sample means fall no further away from the population mean than plus or minus TWO SDs (or z-score units).

Will our sample mean be close or far from the population mean?

When we take ONE sample from a population, we do not know which sample mean we will get! Will it close to the population mean or will it be far away from the population mean?

We do not know! However, we can make probability statements that suggest that it is very unlikely that we will get a sample mean that is “far away” from the population mean. Consider the previous statements in reverse: If 68.26% of all sample means fall no further away from the population mean than plus or minus one SD, then there is only a 31.74% chance that OUR sample mean will be further away from the population mean than plus or minus one SD.

Conversely, if 95.44% of all sample means fall no further away from the population mean than plus or minus or TWO SDs, then there is only a 4.56% chance that OUR sample mean will be further away from the population mean than plus or minus two SDs. ***So there is a less than 5% chance that any sample mean selected at random from the sampling distribution of means falls MORE than 2 SDs away from the population mean.***

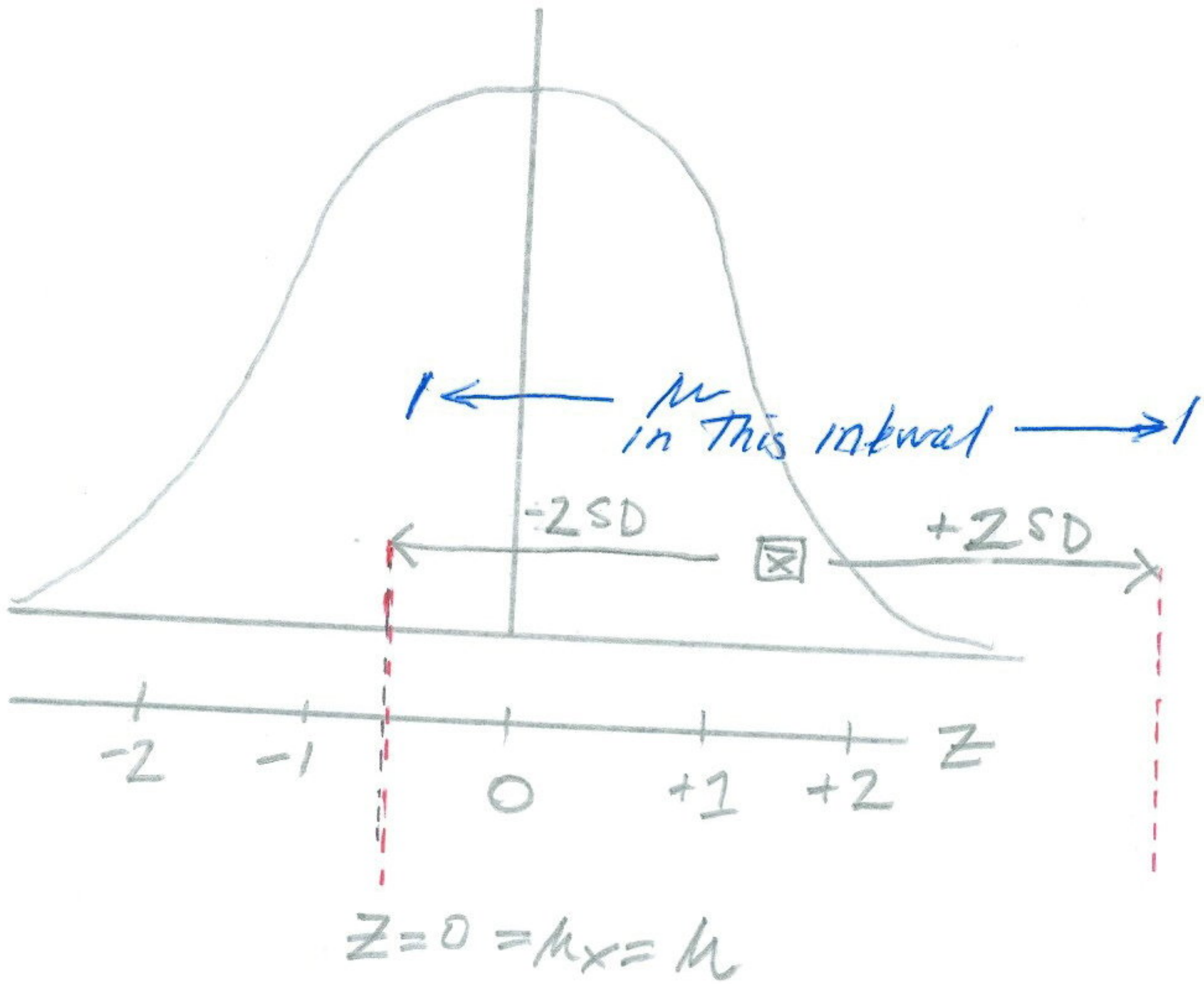
So if we took our sample mean and added 2 SDs to it and subtracted 2 SDs from it, then there is about a 95% chance that our estimate would include the true population mean.

This is exactly what we do with interval estimates.

Going back seats in a classroom representing the sampling distribution of means example

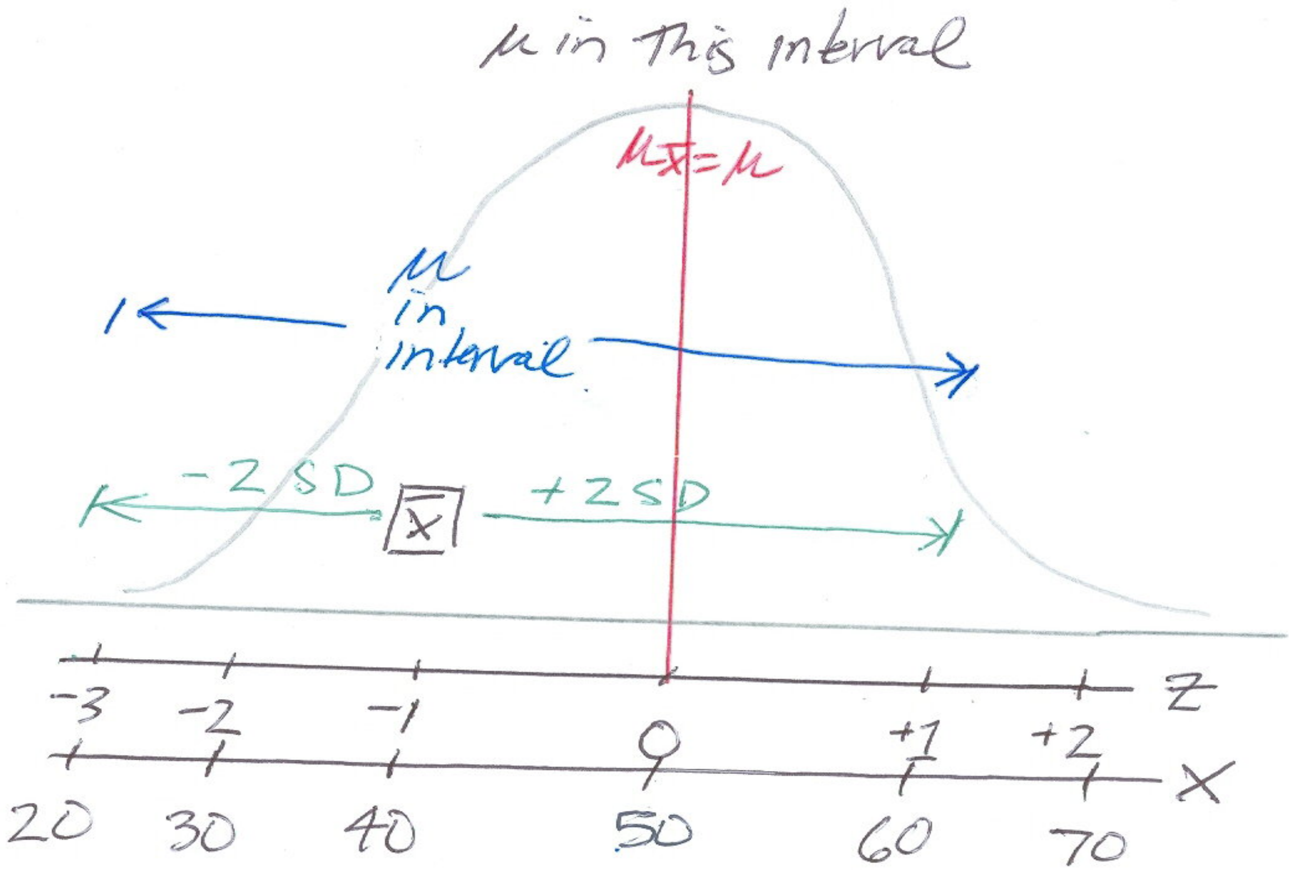
Recall our example above where we pretend that all of the seats in a classroom represent an individual sample mean in the sampling distribution of means.

What if we took our sample mean and added 2 SDs of rope to it and subtracted 2 SDs of rope from it. That interval will cover 95.44% of the area under the curve. That means that there is a 95.44% chance that this interval will include the population mean. Look at the hand drawn picture below.



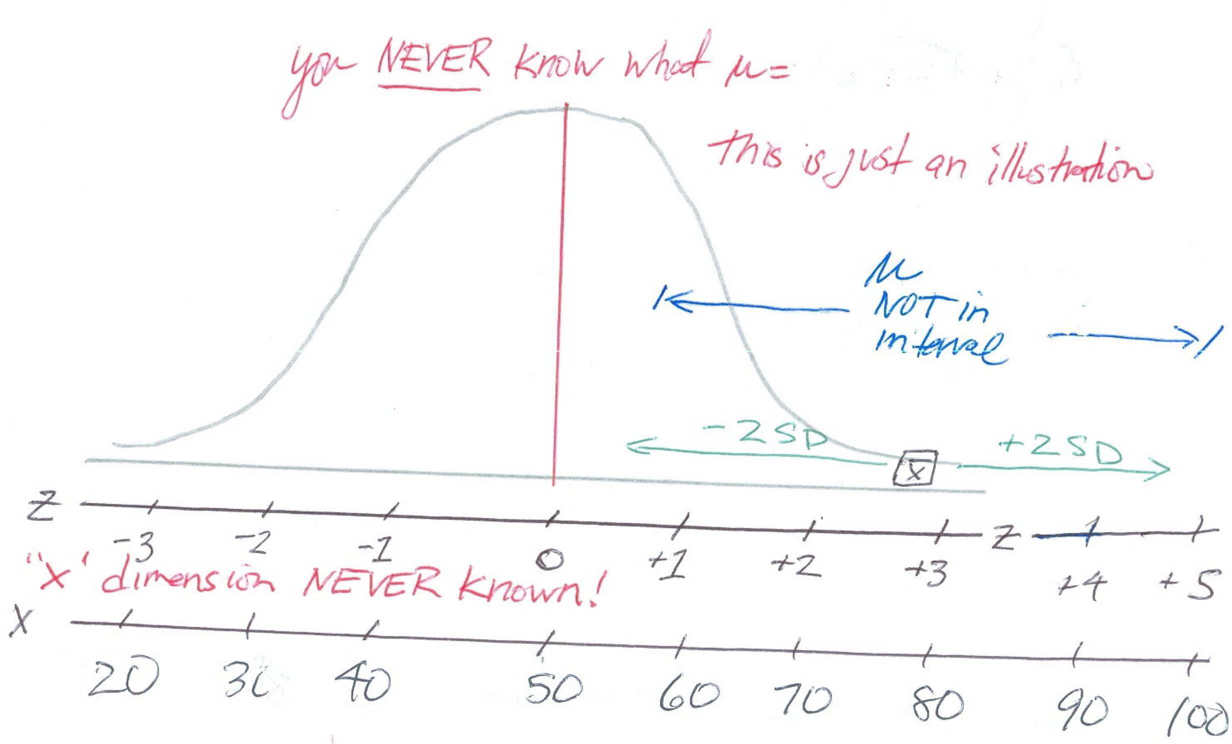
So in our picture above the red dotted lines represent the end-points of each strand of rope (2 SDs in length). You can see that this interval includes the population mean. ***In fact 95.44% of the time your interval will include the population mean if you take your sample mean and add 2 SDs of rope to it and subtract 2 SDs of rope from it!***

Below is yet another example of taking a sample mean at random from the sampling distribution of means and adding and subtracting 2 SDs of rope.


μ in this interval
$\mu_{\bar{x}} = \mu$
μ in interval
-2 SD
$\boxed{\bar{x}}$
+2 SD
-3 -2 -1 0 +1 +2 Z
20 30 40 50 60 70 X

4.56% of the time the population mean is further than 2 SDs away…

Remember the idea from above: 95.44% of the time your interval will include the population mean if you take your sample mean and add 2 SDs of rope to it and subtract 2 SDs of rope from it. Well the logic of this statement also suggests that 4.56% of the time your interval will NOT include the population mean (if you take your sample mean and add 2 SDs of rope to it and subtract 2 SDs of rope from it). You can be 95% confident that most time it will but less than 5% of the time it will not. In the picture below we our sample mean was more than 2SDs away from the population mean; thus the population mean is not included in the interval. (By the way you can ignore the red writing about x dimension etc for now.)



So would you feel comfortable making an estimate of the population mean with a 95.44% chance of being right? If so you would have to accept a 4.56% chance of being wrong. Well, this is what we do in statistics when we use interval estimates. We select a level of confidence (that has a corresponding chance of being wrong). Generally speaking in statistics the "standard of the industry" is that we will accept a 5% or 1% chance of being wrong because that gives us a 95% or 99% chance of being right – which aint bad!

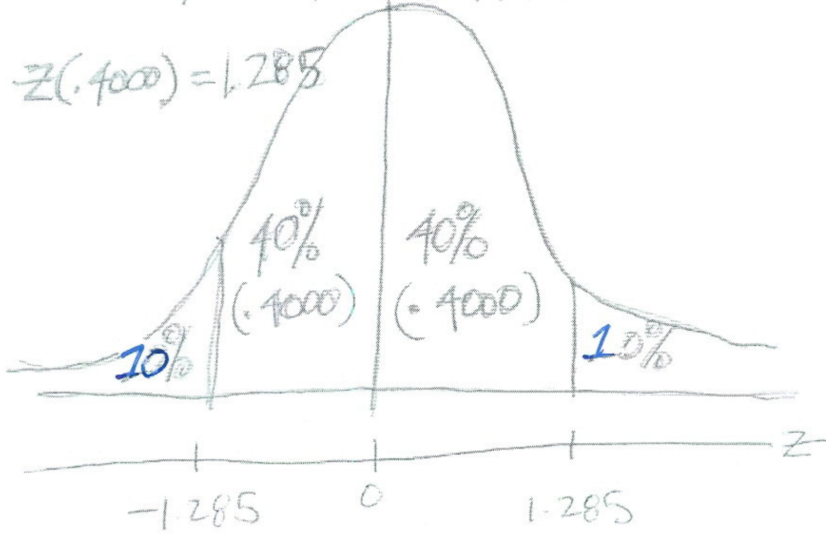
Interval Estimates: more confidence = more SDs

Interval estimates are also called "confidence interval estimates." You decide how much confidence you would like to have and that is how many SDs of rope you have to add and subtract. If you are 80% confident you have a 20% chance of error; 90% confident a 10% of error; 60% confident a 40% chance of error and so on.

Note that the more confidence you want the more rope you have to let out. Saying the same thing in statistics jargon, “Notice that as the level of confidence increases the width of the confidence interval does also.”

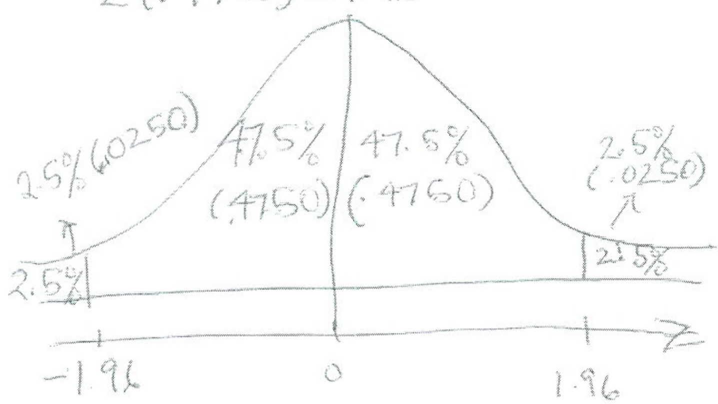
80% confident = 1.285 SD's

Z(.4000) = 1.285



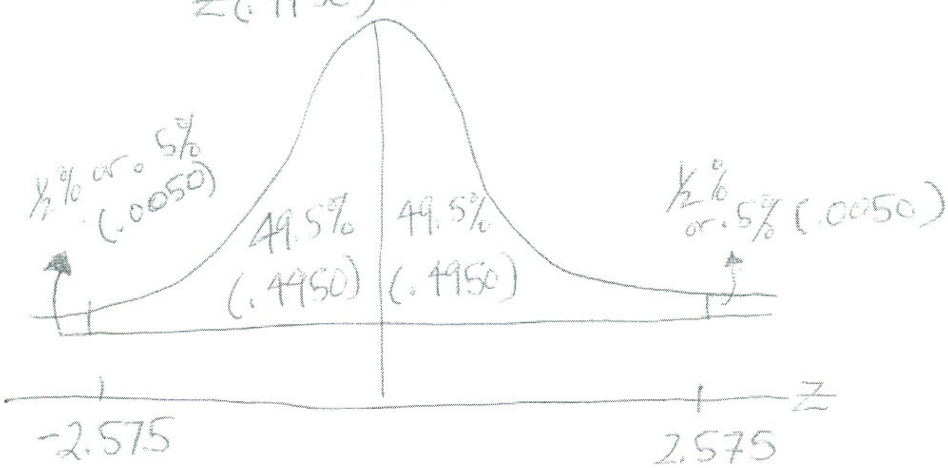
95% confident = 1.96 SD's

Z(.4750) = 1.96



99% confident = 2.575 SD's

Z(.4950) = 2.575



Formulas for Confidence Interval Estimates

Towards the very end of this lecture there is a page with all of the various formulas for confidence intervals. (Some books have cool flow chart that tell you which formula to use (for example see Sanders and Smidt , Figure 7.8 on page 258, sixth edition). But you essentially use the formula that corresponds to the information you have. This will become more clear as you practice problems.

Generic formula for confidence interval estimates of the population mean

$$\bar{x} - z\,\sigma\bar{x} < \mu < \bar{x} + z\,\sigma\bar{x}$$

Notice how this formula is essentially "adding" and "subtracting" something from the sample mean and the population mean is in the middle. (You have to understand the number line and the < and > things. Remember that the "mouth" always want to "eat" the bigger number: 1<2, 5<10, 15<100.) The z is the amount of SDs "of rope" you add and subtract and corresponds to your desired level of confidence. The $\sigma\bar{x}$ is a measure of the SD of the sampling distribution of means and account for sampling error (as n grows, this number becomes smaller and smaller and thus has less and less an effect on the interval.) This should make sense to you. Remember way back when: if you had to make an estimate about the population of UH West Oahu students would you rather have a sample size of 10 or 100? All of you would choose 100 because you know that it would have less sampling error!

Examples of Estimating Parameters

Let's plug some number into the generic formula to get the hang of it.

When σ is known

First know that this "assumption that "σ is known" when we are seeking and estimate of μ is quite frankly CRAZY!!!!! If we know σ then we already know μ:

$\sigma = \sqrt{\dfrac{\sum (x-\mu)^2}{N}}$ = population standard deviation . Notices how μ is there!!!

For some strange reason our, and other books ignore this fact and give us two formulas for standard error when σ is known! (Actually, this can happen but it is very rare. What it is assuming is that there was a study done prior to the current one, and it found a population mean and population standard deviation. You are using your current study to "double check" to see if the population mean remains the same. <u>If you don't understand this do not sweat it. Just plug in the numbers.)</u>

So when (a previous) population standard deviation is known we have two formulas for standard error:

Infinite population when σ is known

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

Finite population

$$\sigma \bar{x} = \frac{\sigma}{\sqrt{n}} \text{ x } \sqrt{\frac{N-n}{N-1}}$$ N = population size n= sample size.

Estimating the population mean when σ is known and n>30 and infinite population

Infinite population

We will use the following formula

$\bar{x}$ - z $\sigma\bar{x}$ < μ< $\bar{x}$ + z $\sigma\bar{x}$ where $\sigma\bar{x} = \frac{\sigma}{\sqrt{n}}$

The tour company owner wants to project a budget!

The owner of a tour company wants to know what her mean monthly sales are. If she knows this then she will be better able to formulate a responsible budget for her company. That should make sense: you want to base your expenditure upon your income, so you need to know what your mean income is.

She takes a random sample of 100 days and finds a sample mean = \$10,000. Assume σ = \$100 n=100. Construct an interval estimate of the true mean monthly sales with a 90% confidence level.

$$\sigma\bar{x} = \frac{\sigma}{\sqrt{n}} = = \frac{100}{\sqrt{100}} = \frac{100}{10} = 10$$

z(90%)=1.645

$\bar{x}$ - z $\sigma\bar{x}$ < μ< $\bar{x}$ + z $\sigma\bar{x}$

10,000 - 1.645(10) < μ<10,000 + 1.645(10)

10,000 – 16.45 < μ<10,000 + 16.45

= 9,983.55< μ<10,016.45

We are 90% confident that the true pop. mean is within these values. In plain English relating the numbers to the question, *“We are 90% confident that the true mean monthly sales for the owner’s tour company are between \$*9,983.55*, and \$*10,164.50”

Notice if we increase the level of confidence the interval gets larger or more spread out

Let’s do a 99% confidence interval where everything else remains the same. The only thing changes is the z score.

z(99%)=2.575

$\bar{x}$ - z $\sigma\bar{x}$ < μ< $\bar{x}$ + z $\sigma\bar{x}$

10,000 - 2.575 (10) < μ<10,000 + 2.575 (10)

10,000 - 25.75 < μ<10,000 + 25.75

$9{,}974.25 < \mu < 10{,}025.75$

We are 99% confident that the true pop. mean is within these values. In plain English relating the numbers to the question*, "We are 99% confident that the true mean monthly sales for the owner's tour company are between $*9,9,74.25 *and $*10,025.75.

Again, the whole point of doing two confidence intervals in this example was to show that when you increase the confidence level, the z-score gets larger and thus the interval estimate has a larger spread of numbers.

***90%: between $9,983.55, and $*10,016.45**

99%: between $9,974.25 and $10,025.75

So as you increase confidence that your interval includes the population mean, but your interval is a greater spread of numbers. In this case the business owner should use the most conservative number: the lower estimate on the 99% interval or $9,974.25.

Estimating population mean when σ is unknown and n>30

When σ is unknown we have to estimate it with sample SD (s) and plug that into two formulas for standard error (note the carrot hat!!!). *Carrot hat means estimated value. We also always assume a random sample.*

Infinite population

$$\hat{\sigma}_{\bar{x}} = \frac{s}{\sqrt{n}}$$

Finite population

$$\hat{\sigma}_{\bar{x}} = \frac{s}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$ N = population size n= sample size.

Therefore a change in formula to signify that you are estimating the population standard deviation:

$$\bar{x} - z\,\hat{\sigma}_{\bar{x}} < \mu < \bar{x} + z\,\hat{\sigma}_{\bar{x}}$$

What is the mean age of people in prison in the US?

Pretend you are a prison administrator that has to design a health care system for the Federal prison system. You need to figure out which drugs to buy and how to negotiate drug prices with drug companies. You know that older prisoners will require different drugs than younger prisoners so you want to figure out the mean age of your prisoners. **Construct an interval estimate of the population mean age with a 95% confidence level.**

Pretend was a study that took a random sample of prisoners in the US and found:

$\bar{x}$ =30, s =10 n=100

$$\hat{\sigma}\bar{x} = \frac{s}{\sqrt{n}} = \frac{10}{\sqrt{100}} = \frac{10}{10} = 1$$

n>30 use z distribution z(95%)=1.96

30 – 1.96 (1) < μ<30+ 1.96 (1)

30 – 1.96 < μ< 30+ 1.96

28.04< μ< 31.96

Plain English: the prison administrator can be 95% confident that the mean age of the population of US prisoners is between 28.04 and 31.96 years.

Estimating population mean when σ is unknown and n<=30

Up until now we have been able to use the z distributions because we've known our populations are normally distributed and/or n>30. (see Central Limits Theorem) *What do we do when n is less than or equal to 30?????*

Assumptions when σ is unknown and n<=30

The population values MUST be known to be normally distributed. If they are not then you must have a sample size larger than 30 to continue! (If your project population values are unknkown and your sample size is less than 30 then just ***assume*** that the population is normally distributed.) You also must have data that comes from a random sample. Rather than the z distribution we use the t distribution (z is actually based upon t).

***t* distribution**

See the exercise at the end of this lecture which goes over how to read a t table and do some exercises with t distributions.

The t distribution is really a family of curves based upon different sized n that approximates a normal distribution when n approaches infinity (n>30 really). Actually "Z" is a part of t!

As such there is a new formula to signify that you are using the t distribution instead of z.

$\bar{x} - t_{\alpha/2}\hat{\sigma}\bar{x} < \mu < \bar{x} + t_{\alpha/2}\hat{\sigma}\bar{x}$

We use alpha/2 because t only deals w/ area on one side of curve.

In the t distribution the area goes out from the t score towards infinity not like z. Since there are different curves for different sample sizes, we must use df = n-1. Everything else about interval estimation using t is the same.

when σ is unknown and n<=30

(It is assumed that the data comes from a random sample and the population parameters are "normal.")

Brian Keaulana, Dennis Govea, Pua Mokuau, and other well respected lifeguards on the West Side train to hold their breath by running with rocks underwater. Since I want to be able to venture out into bigger surf, I start training with them. I want to know how long I will have to be able to hold my breath so that I can go out when Point Surf is going off. So I want to know what is the population mean for breath holding of those who go out in Point Surf. I take a random sample of 9 people who are capable of going out in Point Surf and find a sample mean of 66 seconds with a sample standard deviation of 9 seconds.

Construct an interval estimate of the true mean length of breath hold with a 99% confidence level.

t (df =8 & $\alpha/2$=.005)=3.355

σ_x (don't forget bar over x's and carrot hats)= s/sq.root of n= 9/3=3

x bar - $t_{\alpha/2}\sigma_x < \mu <$ x bar + $t_{\alpha/2}\sigma_x$ (don't forget bar over x's and carrot hats)

$66 - (3.355)\,3 < \mu < 66 + (3.355)\,3$

$66-10.065 < \mu < 66 +10.065$

$55.935 < \mu < 76.065$

What does this mean in plain English? I can be 99% confident that the population mean of breath holding is somewhere between 55 seconds and 76 seconds. If I don't want to drown, I should choose the more conservative of the two and plan on being able to hold my breath for at least 76 seconds if I plan to go out into bigger surf!

Formula Page for Interval Estimates

Some Common Sense Assumptions for Interval Estimates

You can cut and paste assumptions from all the lectures because they are text, but here they are again below. **HINT: do you have a random sample? And if your n<30 you have to assume something VERY special about your population:**

- The variable used is appropriate for a mean (interval/ratio level).
- The data comes from a random sample.
- If the sample size is greater than 30 (n>30) use Z distribution. Statistical theory says that if the population is known to be normal you can use Z when regardless of sample size, but you should ignore theory in this case. In practice if the population is known to be normal and the sample size is small, not around 30, it is better to use the t distribution instead of Z -- it's more conservative. If n<30 and population is unknown use t distribution. If n<30 we ALWAYS assume population is normal.

Generic formulas for confidence interval estimates of the mean

Infinite population when σ is known

$$\bar{x} - z\,\sigma\bar{x} < \mu < \bar{x} + z\,\sigma\bar{x} \quad \text{where} \quad \sigma\bar{x} = \frac{\sigma}{\sqrt{n}}$$

Finite population when σ is known

$$\bar{x} - z\,\sigma\bar{x} < \mu < \bar{x} + z\,\sigma\bar{x} \quad \text{where } \sigma\bar{x} = \frac{\sigma}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}}$$

N = population size n= sample size.

Infinite population when σ is unknown and n>30

$$\bar{x} - z\,\hat{\sigma}\bar{x} < \mu < \bar{x} + z\,\hat{\sigma}\bar{x} \quad \text{where } \hat{\sigma}\bar{x} = \frac{s}{\sqrt{n}}$$

Finite population when σ is unknown and n>30 and N is known

$$\bar{x} - z\,\hat{\sigma}\bar{x} < \mu < \bar{x} + z\,\hat{\sigma}\bar{x} \quad \text{where } \hat{\sigma}\bar{x} = \frac{s}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

N = population size n= sample size.

Infinite population when σ is unknown and n<=30 (assume population is normally distributed!)

$$\bar{x} - t_{\alpha/2}\hat{\sigma}\bar{x} < \mu < \bar{x} + t_{\alpha/2}\hat{\sigma}\bar{x} \quad \text{where } \hat{\sigma}\bar{x} = \frac{s}{\sqrt{n}}$$

Finite population when σ is unknown and n<=30 (assume population is normally distributed!)

$\bar{x}$ - $t_{\alpha/2}\hat{\sigma}\bar{x}$ < μ< $\bar{x}$ + $t_{\alpha/2}\hat{\sigma}\bar{x}$ where $\hat{\sigma}\bar{x} = \frac{s}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$ N = population size n= sample size.

T-Distribution or T-table Exercise

The purpose of this exercise is to learn how to read the table on the inside of your book's front cover (and in Appendix 4) called "Areas for t Distributions" and to do various exercises using that table.

Reading the Table

T-scores are really the same idea as z-scores. They refer to units of standard deviation. The only difference is each t-score is based upon the size of the sample (n).

The Central Limits Theorem says we can use a normal (or Z) distribution when our sample size (n) is greater than 30 (n > 30), but what do we do when our sample size is 30 or less (n < 31)? Thankfully we can use t distributions to accomplish similar tasks. The t distribution is really a family of distribution curves for smaller samples sizes that begins to approximate a normal distribution as n approaches 30. Just like Z distributions, t distributions are used to determine probabilities by using the area under the curve, albeit in a slightly different manner. Rather than measuring the center area of the curve (from the mean outwards), the t-table gives us areas in the tails (or ends) of the curve outwards to infinity. (If this confuses you look at the picture on the top of the table.) To use a t-table we need two pieces of information:

1) **degrees of freedom or *df* --** since the distribution we use is determined by the sample size we need to n to compute the **degrees of freedom or *df.* For one sample computations *df* = n - 1.**

2) we need to know the level of significance or area we are looking for in each tail of the curve. The areas under the curve (or levels of significance) are listed horizontally on the top of the table and the t-score is found in the interior of the table.

For example find the t-score for df=3 and with .10 in the tail. It is 1.638.

Here is a portion of the table:

df	**.10**	**.05**	**.025**	**.005**
1	3.078	6.314	12.706	63.657
2	1.886	2.920	4.303	9.925
3	**1.638**	2.353	3.182	**5.841**

1. What is the t score for 95% confidence interval with n=20, n=10, n=5? (*df*= n - 1)

2. What is the t score for 90% confidence interval with n=20, n=10, n=5? (*df*= n - 1)

3. What is the t score for 99% confidence interval with n=20, n=10, n=5? (*df*= n - 1)

Answers

1.For all problems put .025 in each tail (5% =.05 and .05/2=.025). n=20 df=19 t=2.09, n=10 df=9 t=2.26, n=5 df=4 t=2.77

2. For all problems put .05 in each tail (10% =.10 and .1/2=.05). n=20 df=19 t=1.729, n=10 df=9 t=1.83, n=5 df=4 t=2.13

3. For all problems put .005 in each tail (1%=.01 and .01/2=.005). n=20 df=19 t=2.86, n=10 df=9 t=3.250, n=5 df=4 t=4.60

Practice

Practice problems for this lecture can be found in “16b_practice.pdf”

Practice with SPSS

Practice problems for SPSS output can be found in “16c_SPSS.pdf”