# ONLY TWO MORE!!

Learning Unit 7 – Confidence Intervals and Sample Size, Day 3 and 4

Your Mean Professor Colton

# LU 7, Days 1 and 2 - Outline

Learning Unit 7 – Confidence Intervals and Sample Size

Means with Known Sigma (i.e. know population standard deviation)

- Point Estimate
- Confidence Interval Estimate
- Margin of Error
- Interpreting CI
- Minimum Sample Size

Means with unknown Sigma (i.e. only know sample standard deviation or sample data)

- Point Estimate
- Confidence Interval Estimate
- Margin of Error
- Interpreting CI

# Confidence Intervals Again!

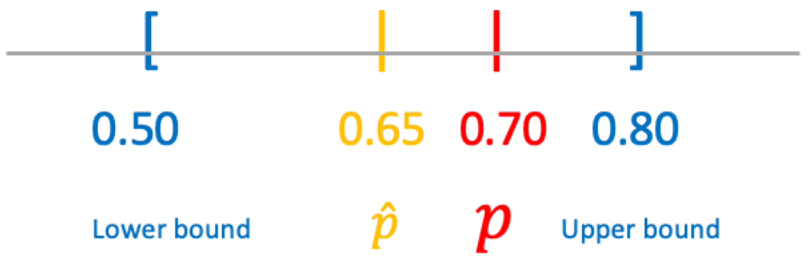
## Estimating Parameters

### Point Estimates

- Using a statistic to estimate a parameter (for means we use $\hat{p}$ or $\bar{x}$ to estimate $p$ *or* $\mu$, respectively.

### Interval Estimates

- Give a range for what we think the population parameter is.

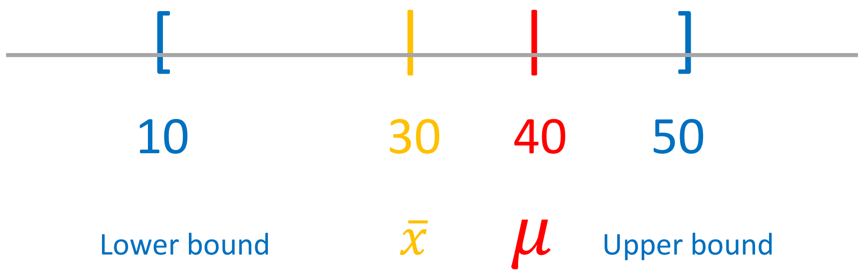



### Proportions

- CI for **proportions** deal with categorical variables (yes / no).

### Means

- CI for **means** deal with quantitative variables (numeric).



# Sampling Distribution and Central Limit Theorem (CLT) of $\bar{x}$

## RECALL!

Central Limit Theorem

- Let $\bar{x}$ be the mean of observations in SRS of size $n$ from a population with mean $\mu$ and standard deviation $\sigma$.
- If we take a large enough sample, then
  - The mean of $\bar{X}$ is equal to the mean of the population, $\mu$
$$\mu_{\bar{X}} = \mu$$
  - The standard deviation is equal to $\sigma$ divided by the square root of $n$
$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$
  - And this distribution will be approximately Normal!
$$\bar{X} \sim Normal(mean = \mu, SD = \frac{\sigma}{\sqrt{n}})$$

- Further, as the sample size n increases, the sampling distribution of $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ approaches a standard normal distribution ($\mu = 0, \sigma = 1$), regardless of the shape of the population distribution.

# CLT for $\bar{x}$ - Assumptions

- Just like with our distribution assumptions (Binomial and Poisson), each scenario <u>must meet some conditions / assumptions</u> in order for the <u>results of the CLT to apply</u> to the *distribution of sample means*!

<u>CLT Assumptions</u>

- **Randomization Condition:** The data values must be sampled randomly.
- **Large Enough Sample Condition: There is no one-size-fits-all rule (although we will have some specific guidelines for this class).**
  - If the population is Normally distribution, then ANY sample size would meet this condition (even $n = 1$).
  - If the population is <u>unimodal and symmetric</u>, even a <u>fairly small sample is okay</u>, maybe $10 \leq n \leq 30$.
  - If the population is <u>strongly skewed</u>, we need a <u>large sample</u> to allow the use of a normal model to describe the distribution of sample means.
    - For this class, if $n \geq 30$, then we can say this <u>is large enough</u> even if we have a <u>really skewed population</u>!

# Final Confidence Interval for $\mu$

Now the goal is to find an estimate for the unknown population mean $\mu$!

*** We are going to assume that the population standard deviation **σ** is known!

1 Mean Z Interval

* Same Critical Value as with a 1 Proportion Z Interval

C.I. = Point Estimate ± Margin of Error, MOE = CV * SE

$= \bar{x} \pm Z^* \sigma_{\bar{x}}$

$= \bar{x} \pm Z^* \frac{\sigma}{\sqrt{n}} \quad \rightarrow \quad \left(\bar{x} - Z^* \frac{\sigma}{\sqrt{n}}, \bar{x} + Z^* \frac{\sigma}{\sqrt{n}}\right)$

Don't forget to **Check the Conditions**

✓Randomization Condition
Need to have a random sample

✓Large Enough Sample Condition
Normal population OR
$n \geq 30$

Same interpretation too!!

I am % confident that the true/population parameter + context is between (lower bound) and (upper bound).

# Using Calc!

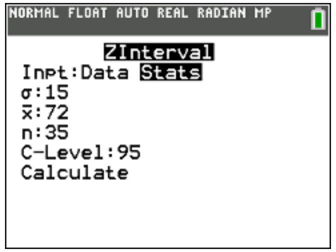
**Setup**

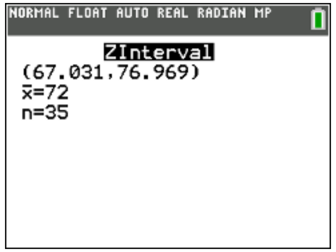
Suppose the number of points scored by your favorite basketball team is normally distribution with known population standard deviation **σ** = 15 points.

The mean from a random sample of 35 games was 72 points. **Calculate** and **interpret** the corresponding *95% confidence interval*!

**GOAL**: Find the Confidence Interval!

1. ZInterval
    a) Input = Stats
    b) **σ** = population standard deviation
    c) $\bar{x}$ = sample mean
    d) n = sample size
    e) C-Level = Confidence level (as a decimal or whole number, both work)




*95% CI = (67.301, 76.969)*

Interpret results:

*We are 95% Confident that the true mean points scored per game is between 67.031 and 76.969 points.*

# Another LCQ

**Setup**: Lets assume the population of ACT scores is normally distribution with known population standard deviation **σ** = 3.5 points. From a random sample of 15 students, there was a sample mean score of 24.

1) Check the conditions for a 1-Mean Z Interval.

2) Calculate the 85% Confidence Interval

3) Interpret this interval.

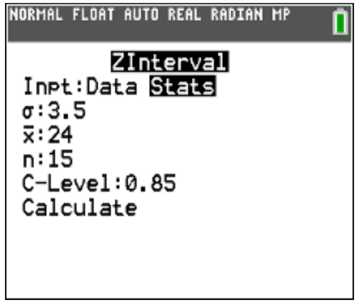
# Another LCQ

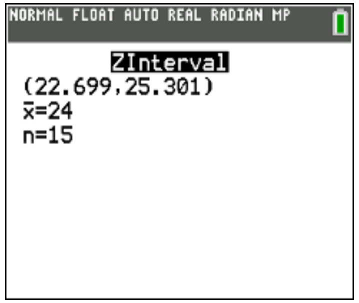
**Setup**: Lets assume the population of ACT scores is normally distribution with known population standard deviation **σ** = 3.5 points. From a random sample of 15 students, there was a sample mean score of 24.

1) Check the conditions for a 1-Mean Z Interval.

*Normal population distribution (so sample size doesn't matter), random sample —> all conditions met!*

2) Calculate the 85% Confidence Interval = *(22.699 , 25.301)*




3) Interpret this interval.

*Common answers:*

a) *We are 85% confident the ACT score of students is between 22.699 and 25.301 -> MISSING parameter (true mean)*

b) *We are 85% confident that the true population mean is 22.699 and 25.301 -> MISSING context (ACT score of students)*

c) *We are 85% confident that true/population mean of Students scored on ACT between 22.69 and 25.301*

*Option C is PERFECT!! Has all the parts we need!*

# One more LCQ

**Setup**: Suppose an instructor wishes to predict the average time needed to take the final exam. A random sample of 45 students shows a mean of 1.4 hours. The population standard deviation is known to be 0.25 hours.

1) If appropriate, calculate and interpret the 97% Confidence Interval.

2) If I increase the sample size to 60, find the margin of error of the new CI. Is it smaller or larger than the MOE in (1)?

3) If we were mistaken and the actual population standard deviation is 0.75 hours. With a sample size of 45, will a 97% CI be wider or narrower than the result in (1)? Find the new interval.

# One more LCQ

**Setup**: Suppose an instructor wishes to predict the average time needed to take the final exam. A random sample of 45 students shows a mean of 1.4 hours. The population standard deviation is known to be 0.25 hours.

REMEMBER! If it says "if appropriate, we HAVE TO check the conditions!

SHOW YOUR WORK! For calculating an interval, that means writing the name of the procedure and the inputs!

1) If appropriate, calculate and interpret the 97% Confidence Interval.

*Check conditions: Random sample and n = 45 ≥ 30, both conditions are met!*

*Calculation: Zinterval -> Stats, st dev = 0.25, mean = 1.4, n = 45, C-Level: 0.97 -> (1.3191, 1.4809)*

*a) I'm 97% confident that the mean of the time it takes to take the exam is between 1.3191 and 1.4809 hours -> MISSING true*

*b) We are 97% confident that the sample mean of the time to take the test is between 1.3191 and 1.4809 hour. -> WRONG! We are trying estimate the POPULATION mean, we already know what the SAMPLE mean is (DON'T write SAMPLE!)*

*c) CORRECT!!! I'm 97% confident that the true mean of the time it takes to take the exam is between 1.3191 and 1.4809 hours*

2) If I increase the sample size to 60, find the margin of error of the new CI. Is it smaller or larger than the MOE in (1)?

*After increasing our sample size to 60, 97% CI = (1.33, 1.47).*

*New MOE = Width / 2 = (UB – LB) / 2 = (1.47 – 1.33) / 2 = 0.055, which is a smaller interval than the original MOE. We knew it would be smaller even before calculating the new interval!*

*Bonus, interpretation: We are 97% confident the true mean of the time to take the exam is between 1.33 and 1.47 hours.*

3) If we were mistaken and the actual population standard deviation is 0.75 hours. With a sample size of 45, will a 97% CI be wider or narrower than the result in (1)? Find the new interval.

*Wider!* $MOE = Z^* \frac{\sigma}{\sqrt{n}}$, *increasing the numerator of the SE (and keeping the sample size and CL the same) will increase the MOE and this a wider interval!*

*New Calculation: Zinterval -> Stats, st dev = 0.75, mean = 1.4, n = 45, C-Level: 0.97 -> (1.1574, 1.6426)*

# Finding the Minimum Sample Size for Means - Calculation

- Same strategy: Start with the formula for Margin of Error and rearrange to solve for *n*:

$$MOE = Z^* \frac{\sigma}{\sqrt{n}} \Longrightarrow n = \left(\frac{Z^*\sigma}{MOE}\right)^2$$

- Solve for *n*!
  - Since this formula is for the *minimum* sample size, if any decimal occurs, ALWAYS round up to the next largest whole number.

- The error (MOE) is often identified by the term "within”.

Example

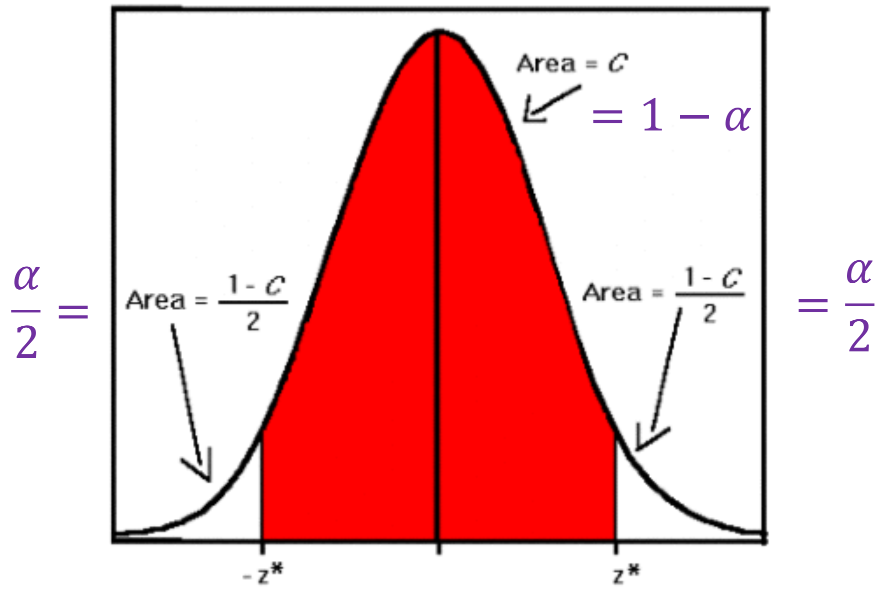
- Determine the sample size necessary to construct an 95% confidence interval for the average credit card balance. Suppose the standard deviation is $215 and we want to estimate the true mean balance within $50.
  - *95% C -> Z* = 1.96,* $n \geq \left(\frac{Z^*\sigma}{MOE}\right)^2 = \left(\frac{1.96 \times 215}{50}\right)^2$ *= 71.03 -> 72, Need a sample size of at least 72 to estimate the mean credit card balance within $50 of the true mean balance with 95% Confidence.*

# Alpha $\alpha$

- This is an idea that becomes more relevant in the next unit, but we will introduce it here.

<u>Alpha Level for a CI</u>

- Recall % Confident represents the middle C% of the Z or t distribution.
- Alpha level is just the complement probability!
    - Ex) If we have a 95% CI, alpha level $\alpha = 1 - 0.95 = 0.05$
- The area for $\alpha$ is split equally between the upper and lower tails!



# A New CI for a New Unknown

- We've looked at how to create a CI for a population mean:

$$\text{C.I.} = \bar{x} \pm Z^* \frac{\sigma}{\sqrt{n}}$$

- In doing so, what did we have to assume?? Population standard deviation **σ** was KNOWN!
- Is this practical?? NO!! It's a veryyy big assumption, unlikely to be known.

- What do we do when we don't have this information? Any ideas???
  - If we known the value of **σ**, we can **estimate** it with our **sample standard deviation**!!
  - This intuitively makes sense! It is essentially the same thing we are doing with the population mean too. We use our sample mean as a point estimator!
    - And the population mean is ultimately what we are after, so we then give an interval around the point estimator, the sample mean.
    - With the population standard deviation, we aren't necessarily interested in this quantity, but we still need it! So just having a single point estimator suffices!

# Final Confidence Interval for $\mu$, unknown $\boldsymbol{\sigma}$!

The goal is still to find an estimate for the unknown population mean $\mu$!

*** But now we have to **estimate** the population standard deviation $\boldsymbol{\sigma}$ is with our sample standard deviation $s$!

- So we have a new procedure! And specifically, a new critical value and standard error!
- It is now based on the ***t*-distribution** rather than the standard normal distribution Z.

1 Mean T Interval

This is wayyy more common than a Z Interval!

C.I. = Point Estimate ± Margin of Error,     MOE = CV * SE

$= \bar{x} \pm t^* \sigma_{\bar{x}}$

$= \bar{x} \pm t^* \frac{s}{\sqrt{n}}, \quad df = n - 1$

$\rightarrow \left(\bar{x} - t^* \frac{s}{\sqrt{n}}, \bar{x} + t^* \frac{s}{\sqrt{n}}\right)$

df = Degrees of freedom

Don't forget to **Check the Conditions**

✓Randomization Condition

Need to have a random sample

✓Large Enough Sample Condition

Normal population OR

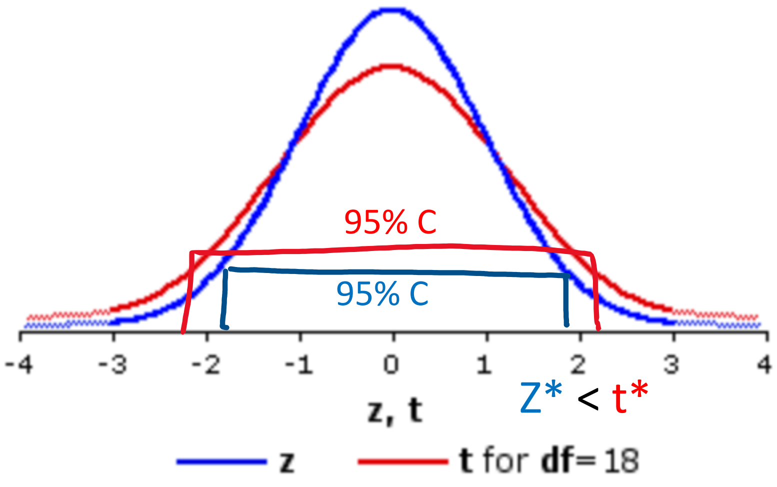
$n \geq 30$

Still same interpretation!!

I am % confident that the true/population parameter + context is between (lower bound) and (upper bound).

# CIs for Known vs Unknown σ

- Because we have to **estimate** an <u>extra parameter</u>, there is <u>inherently more variability in a *t*-interval</u>!
    - This makes sense! Sample means vary from sample to sample, so do the standard deviations!
- So to account for that extra level of <u>uncertainty</u>, ***t*-intervals** produce <u>wider intervals</u> compared to Z-intervals (for the same confidence level and sample size).

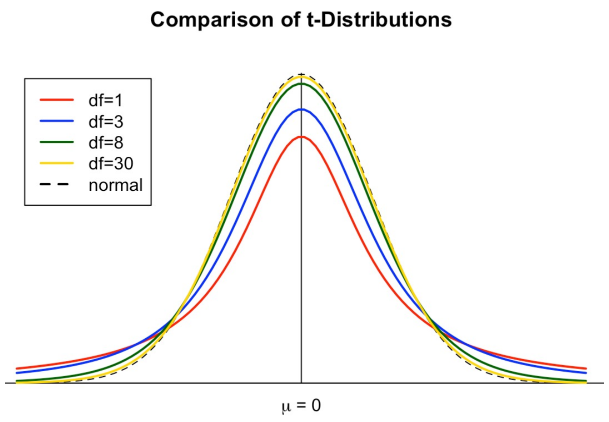



http://vassarstats.net/textbook/ch9pt3.html

# t-Distribution

We are just going to briefly mention some ideas about the t-Distribution.

https://www.geo.fu-berlin.de




<u>Shape and Parameters</u>

- Symmetric and unimodal distribution (slightly resembles a bell shape).
- The t-distribution has heavier tails than the Z distribution!
  - This means there is more probability near edges and likewise less probability in the middle!
- Indexed by the degrees of freedom, *df = n – 1 = sample size – 1*
  - This tells us exactly <u>which</u> t-distribution we are talking about!

<u>Why do we need to use t??</u>

- Recall, as the sample size <u>*n* increases</u>, the sampling distribution

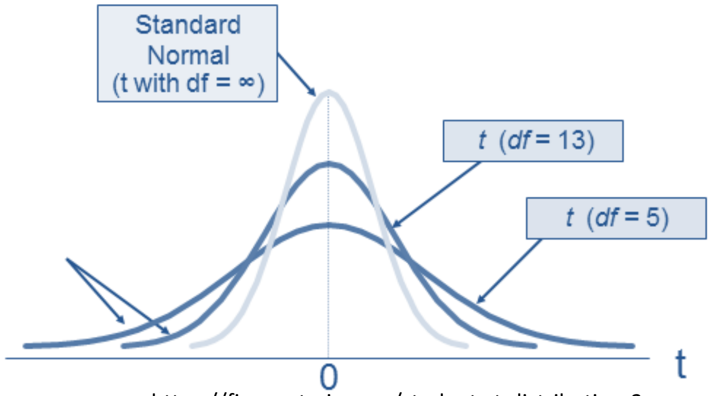
$$\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \rightarrow Z$$

- Well now, because we have to substitute *s* for $\sigma$, this becomes

$$\frac{\bar{x}-\mu}{s/\sqrt{n}} \rightarrow t$$

<u>Interesting Tidbit</u>

- As the df goes towards infinity, t becomes Z!




https://financetrain.com/students-t-distribution-2

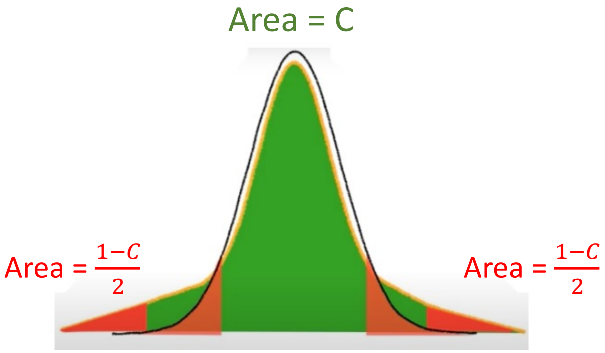
Doper applet!

# New Critical Value

<u>Critical Value – t Star</u>

- Same exact logic as with Z* and Confidence, just a new distribution! Now t*
- The t-scores that mark the middle %C of the t-distribution with n-1 degrees of freedom!
- Values are symmetric, so can just find one!

$$t^* = invT\left(area = \frac{1-C}{2}, df = n-1\right)$$




t-distribution (*for different df*)

https://www.youtube.com/watch?v=QoV_TL0IDGA

<u>Mini LCQ</u>

Find the Critical Values for the following Confidence Levels with *n = 33*:

a) 90% Confident -> *t* = ??*
b) 95% Confident -> *t* = ??*

Find the Critical Values for the following Confidence Levels with *n = 20*:

a) 90% Confident -> *t* = ??*
b) 95% Confident -> *t* = ??*

Find the % Confidence based on the following Critical Value:

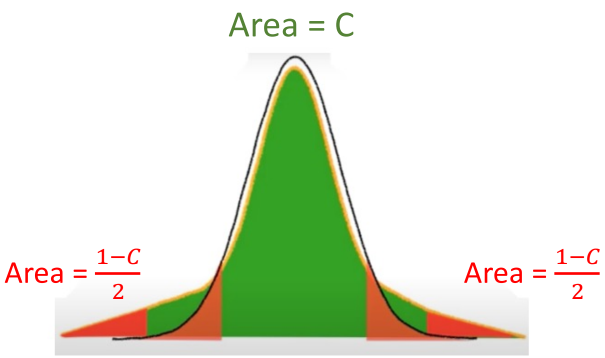
c) t* = 1.281, df = 40 -> *C = ??*

# New Critical Value

<u>Critical Value – t Star</u>

- Same exact logic as with Z* and Confidence, just a new distribution! Now t*
- The t-scores that mark the middle %C of the t-distribution with n-1 degrees of freedom!
- Values are symmetric, so can just find one!

$$t^* = invT(area = \frac{1-C}{2}, df = n-1)$$




t-distribution (*for different df*)

https://www.youtube.com/watch?v=QoV_TL0IDGA

** Technically this expression gives you the negative t*, but when finding the MOE we want to use the positive value

<u>Mini LCQ</u>

Find the Critical Values for the following Confidence Levels with *n = 33*:

a) 90% Confident -> *t* = invT(area =* $(\frac{1-0.9}{2}$*, df = 33 – 1) = -1.6938*

b) 95% Confident -> *t* = invT(area = (0.025, df = 32) = -1.6938*

Find the Critical Values for the following Confidence Levels with *n = 20*:

a) 90% Confident -> *t* = invT(area = 0.05 + 0.90, df = 19) = 1.729*

b) 95% Confident -> *t* = invT(area = 0.025 + 0.95, df = 19) =* 2.093

Same CL! But different t* because of df!!!

** Could just find the positive t* as well

Find the % Confidence based on the following Critical Value:

c) t* = 1.281, df = 40 -> *C = cdf(lower = -1.281, upper = 1.281, df = 40) = 0.79*

# Using Calc!

**Setup**

Lets assume the population of SAT scores is normally distribution with unknown population standard deviation.

From a random sample of 6 students, there was a sample mean score of 1190 and sample standard deviation of 205.91 points. **Calculate** and **interpret** the corresponding *95% confidence interval*!

**GOAL**: Find the Confidence Interval!

TInterval

- Option 1) Input = Stats

a) $\bar{x}$ = sample mean
b) Sx = sample standard deviation
c) n = sample size
d) C-Level = Confidence level (as a decimal or whole number, both work)

- Option 2) Input = Data
  - Enter raw data in $L_1$

a) List = $L_1$
b) Freq = 1
c) C-Level = Confidence level (as a decimal or whole number, both work)

| Score |
|---|
| 1300 |
| 1200 |
| 1190 |
| 1050 |
| 1500 |
| 900 |
| |
| **Mean = 1190** |
| **SD = 205.91** |

Interpret results:
??

# Using Calc!

**Setup**

Lets assume the population of SAT scores is normally distribution with unknown population standard deviation.

From a random sample of 6 students, there was a sample mean score of 1190 and sample standard deviation of 205.91 points. **Calculate** and **interpret** the corresponding *95% confidence interval*!

**GOAL**: Find the Confidence Interval!

TInterval

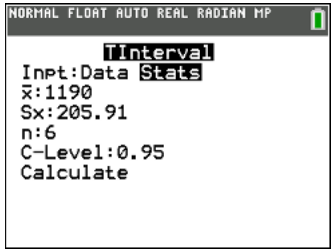
- Option 1) Input = Stats

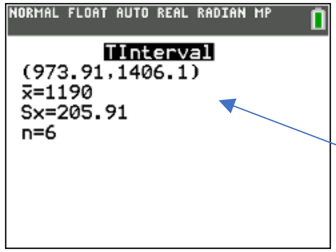
a) $\bar{x}$ = sample mean

b) Sx = sample standard deviation

c) n = sample size

d) C-Level = Confidence level (as a decimal or whole number, both work)




Same intervals! (maybe a little roundoff error from $\bar{x}$ *or s)*

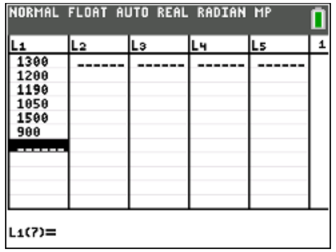
- Option 2) Input = Data
    - Enter raw data in $L_1$

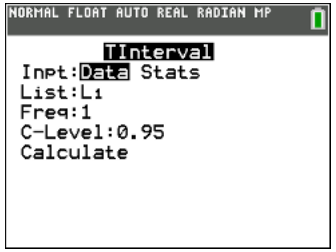
a) List = $L_1$

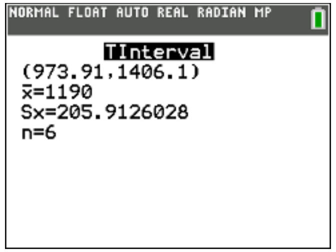
b) Freq = 1

c) C-Level = Confidence level (as a decimal or whole number, both work)

| Score |
| --- |
| 1300 |
| 1200 |
| 1190 |
| 1050 |
| 1500 |
| 900 |
| |
| **Mean = 1190** |
| **SD = 205.91** |




Interpret results:

*We are 95% Confident that the true mean SAT scores for students is between 973.91 and 1406.1 points.*

# LCQ

**Setup**: We know that number credit hours students at CSCC is normally distributed. From a random sample of 20 students, there was a sample mean 8.5 credit hours and sample sd = 2 credit hours.

1) (Only) Calculate the 90% Confidence Interval.

2) Admissions department tells us the true (population) standard deviation is 1.25 credit hours. (Only) Calculate the 90% Confidence Interval.

3) Out of the 20 students, 8 said they were taking a Statistics course. (Only) Calculate the 85% Confidence Interval.

# LCQ

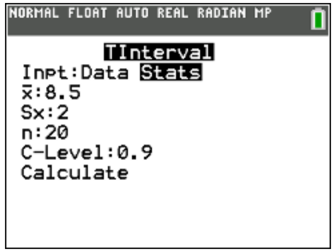
*** **Need to be able to recognize which type of interval to make**! To help, first think about the response variable (or the parameter of interest)!

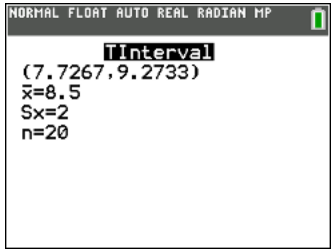
- Mean $\mu$ or proportion *p??* (Quantitative or Qualitative)
- Then think about what information you have, what you need, and what to type in calc!

**Setup**: We know that number credit hours students at CSCC is normally distributed. From a random sample of 20 students, there was a sample mean 8.5 credit hours and sample sd = 2 credit hours.

1) (Only) Calculate the 90% Confidence Interval.

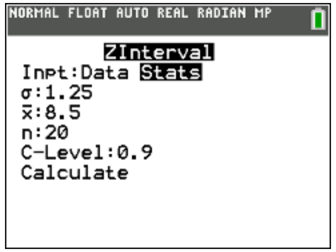
- *TInterval -> Stats, $\bar{x}$ = 8.5, Sx = 2, n = 20, C-Level= 0.9 -> (7.7267, 9.2733)*

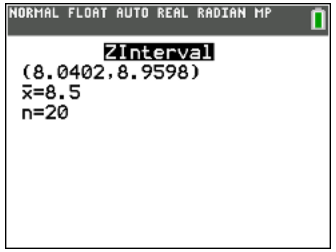



2) Admissions department tells us the true (population) standard deviation is 1.25 credit hours. (Only) Calculate the 90% Confidence Interval.

- ***NOT CORRECT**: TInterval -> Stats, , $\bar{x}$ = 8.5, Sx = 1.25, n = 20, C-Level = 0.9 -> (8.0167, 8.9833)*
- *We KNOW what the value of **$\sigma$** is now!! So we don't have to estimate it anymore and can therefore use a **Z Interval**!!*
  - *This will give us a more precise (better) interval for the same confidence level! So this would be the CORRECT type of interval*
- *ZInterval -> Stats, **$\sigma$** = 1.25 , $\bar{x}$ = 8.5, n = 20, C-Level = 0.9 = (8.0402, 8.9598)*



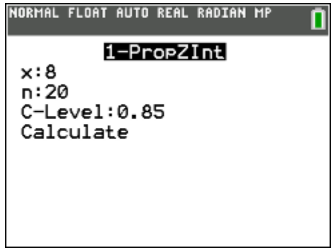



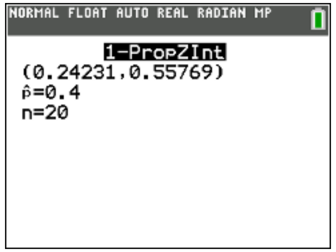

3) Out of the 20 students, 8 said they were taking a Statistics course. (Only) Calculate the 85% Confidence Interval.

- *We were given are **sample proportion**. -> NEED TO FIND PROPORTIONS INTERVAL!!!!!!*
- *The response variable has changed! We NO LONGER are after the amount of credit hours, rather whether or not students are taking statistics!*
  - *This is a YES/NO variable, i.e. **CATEGORICAL**!!*
- *So we CANT use mean formulas, we need to use **PROPORTIONS and 1PROPZINT**!!!*

*1-PropZInt -> x = 8, n = 20, C-Level = 0.9 = (8.0402, 8.9598)*

*1PropZInt!!! -> (0.24231, 0.55769) true proportion of CSCC students taking a stats course!*



# Finding the Minimum Sample Size for Means with Unknown $\boldsymbol{\sigma}$

- There is no procedure for this when $\boldsymbol{\sigma}$ is unknown
- We can’t substitute in a sample standard deviation because we haven’t taken a sample yet!
- So researchers have to make an assumption about the value of $\boldsymbol{\sigma}$ in order to do sample size calculations!
    - This can be a tricky task!!
    - Resulting estimates for minimum sample sizes can change drastically based on how much variability (i.e. the value of $\boldsymbol{\sigma}$) researchers assume is in the process they are studying!

- So we would need to use the same formula presented earlier when $\boldsymbol{\sigma}$ is known! Refer back to those slides!

# When do I use which? *p* vs *Z* vs. *t*

Use **1 Proportion Z Interval** when…

- You are asked to find a Confidence interval for a Population **Proportion**
- Asked for a sample size for a Confidence Interval with particular width, etc.
  - Dealing with Proportions (i.e. a success or failure)

Use **T Interval** when...

- You are asked to find a Confidence interval for a Population **Mean**
  - This applies to estimating the population mean, not proportion
  - There is **no** “success” or “failure” being measured.
  - Generally always the population standard deviation will be unknown!

Use **Z Interval** only for…

- Finding a sample size of a confidence interval with a particular width/margin of error when looking at quantitative data. (The **only time** we use Z Interval Formulas even though proportions have Z’s)
- (Very uncommon) Finding a CI for a Population mean when the population standard deviation is known.

As always, think about what is being measured for each observation/individual!

**USE THE CORRECT NOTATION AND FORMULAS!!!!**

# Means vs. Proportions: When do I use which formulas?

Looking at a categorical variable that has a **"success" and "failure"**? (sample proportions)

Sampling Distributions:

- Mean: $\mu_{\hat{p}} = p$ aka the population proportion
- Standard Deviation: $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$

Confidence Interval: $\hat{p} \pm Z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Looking at a **quantitative** variable? (sample means)

Sampling Distributions:

- Mean: $\mu_{\bar{X}} = \mu$ aka the mean of the population
- Standard Deviation: $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$

Confidence Intervals:

$$\bar{x} \pm Z^* \frac{\sigma}{\sqrt{n}}$$

$$\bar{x} \pm t^* \frac{s}{\sqrt{n}}, df = n - 1$$

# Problem Session!!!

# Human Body Temperatures

- Human body temperatures are normally distributed
- Find a 95% confidence interval for $\mu$ given the random sample of 15 temperatures below.

| | | | | |
|---|---|---|---|---|
| 97.6 | 100.3 | 100 | 97.9 | 99.2 |
| 98.8 | 97.4 | 98 | 99.1 | 100.2 |
| 97.1 | 99.8 | 98.5 | 99.5 | 98.3 |

# Comparing CIs

- The confidence level of a CI signifies the confidence we have that the parameter actually lies in that interval
- Length of the CI indicates the precision of the estimate
- For the sample, the 90% CI for the mean human body temperature, μ, is between 98.3º and 99.3º
- For the sample, the 95% CI for the mean human body temperature, μ, is from 98.2º to 99.4º
- For the sample, the 99% CI for the mean human body temperature, μ, is between 97.9º and 99.6º
- As confidence level increases, the precision decreases

# Practice Problem #2, continued

A survey of 25 randomly selected customers found the mean age was 31.84 years and the standard deviation 9.84 years.

c) How many degrees of freedom does the t-statistic have?

d) How many degrees of freedom would the t-statistic have if the sample size had been 100?

e) Construct and interpret a 95% confidence interval for the mean age of all customers. Check to see if the assumptions and conditions for the confidence interval have been met.

f) How large is the margin of error?

g) How would the confidence interval change if you had assumed that the standard deviation was known to be 10.0 years?

# Practice Problem #2 Solution, continued

c) df = n – 1 = 25 – 1 = 24 d) df = n – 1 = 100 – 1 = 99

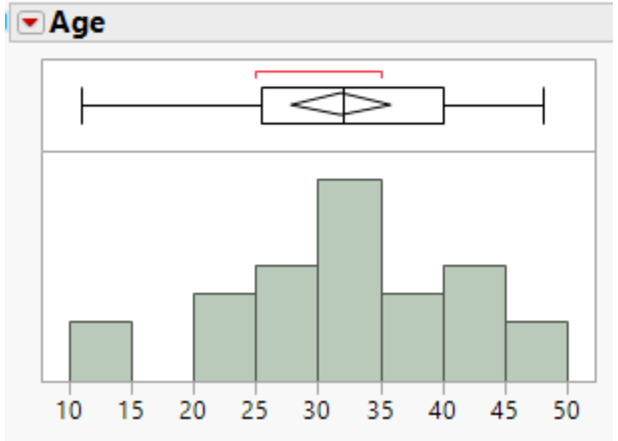
e) The data were randomly obtained and should be independent. The sample size is less than 10% of the population. Slightly left skewed, but unimodal, therefore, the t-inference procedure is appropriate.

I am 95% confident that the true mean age of customers is between 27.777 and 35.903 years.

f) $MOE = \frac{35.903 - 27.777}{2} = 4.063$

g) The confidence interval would be wider, as the standard error would be $\frac{10}{\sqrt{25}} = 2$, instead of $\frac{9.84}{\sqrt{25}} = 1.968$.

The new confidence interval would be (27.712, 35.968).




**Summary Statistics**

| | |
|---|---|
| Mean | 31.84 |
| Std Dev | 9.8432718 |
| Std Err Mean | 1.9686544 |
| Upper 95% Mean | 35.903103 |
| Lower 95% Mean | 27.776897 |
| N | 25 |

**Confidence Intervals**

| Parameter | Estimate | Lower CI | Upper CI | 1-Alpha |
|---|---|---|---|---|
| Mean | 31.84 | 27.7769 | 35.9031 | 0.950 |
| Std Dev | 9.843272 | 7.685906 | 13.69349 | 0.950 |

# Choosing Sample Size for Means Example

The financial aid office wishes to estimate the mean cost of textbooks per semester for students at Miami University. For the estimate to be useful, it should be within \$25 of the true population mean. How large a sample should be used to be 90% confident of achieving this level of accuracy?

- Use \$100 as a reasonable estimate for $\sigma$

# Sample Size for Mean Solution

- Margin of error, MO = 25
- Since we want a 90% confidence interval, Z* = 1.645
- Use $\sigma$ = 100

$$n = \frac{\sigma^2 {Z^*}^2}{MO^2} = \frac{100^2 1.645^2}{25^2} = 43.29 \ldots \rightarrow \mathbf{44}$$

- We need to know the amount that at least 44 students paid for their textbooks to estimate the mean cost of textbooks per semester within $25 of the true mean cost of textbooks per semester with 90% confidence

# Still more of Practice Problem #2

A survey of 25 randomly selected customers found the mean age was 31.84 years and the standard deviation 9.84 years. The 95% confidence interval was (27.777, 35.903).

h) For the 95% confidence interval found in part (e), how large would the sample size have to be to cut the margin of error in half?

i) For the 95% confidence interval found in part (e), how large would the sample size have to be to cut the margin of error by a factor of 10?

j) How large of a sample size would you need to construct a 90% confidence interval with a margin of error of 2, assuming the population standard deviation is 10?

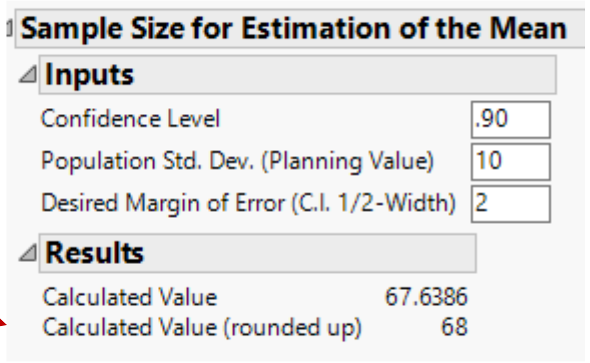
k) How large of a sample size would you need to construct a 99% confidence interval with a margin of error of 2, assuming the population standard deviation is 10?

# Still more of Practice Problem #2 Solution

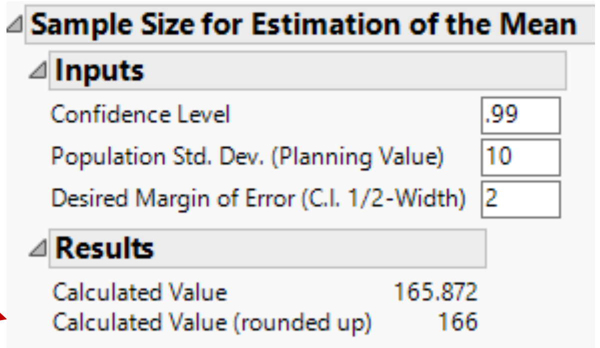
h) The sample size would need to be 4 times larger, n = 100

i) The sample size would need to be 100 times larger, n = 2500

j) The sample size would need to be 68




k) The sample size would need to be 166



# Practice Problem #3

A random sample of 30 gas stations in a region gives the following statistics:

$\bar{y}$ = \$4.49 s = \$0.29

a) Find a 95% confidence interval for the mean price of regular gasoline in that region.
b) Find the 90% confidence interval for the mean price of regular gasoline in that region.
c) If we had the same statistics from a sample of 60 stations, what would the 95% confidence interval be now?

# Practice Problem #3 Solution

a) (\$4.3817, \$4.5983) I am 95% confident that the true mean price of regular gasoline is between \$4.382 and \$4.598 per gallon.

b) (\$4.4000, \$4.5800) I am 90% confident that the true mean price of regular gasoline is between \$4.400 and \$4.580 per gallon.

c) 95% CI, n = 60, (\$4.4151, \$4.5649)

I am 95% confident that the true mean price of regular gasoline is between \$4.415 and \$4.565 per gallon.

# Practice Problem #5

Hoping to lure more shoppers downtown, a city builds a new public parking garage in the central business district. The city plans to pay for the structure through parking fees. For a random sample of 44 weekdays, daily fees collected averaged \$126, with a standard deviation of \$15.

a) What assumptions must you make in order to use this statistics for inference?
b) Find a 90% confidence interval for the mean daily income this parking garage will generate.
c) Explain in context what this confidence interval means.
d) Explain what 90% confidence means in this context.
e) The consultant who advised the city on this project predicted that parking revenues would average \$128 per day. Based on your confidence interval, what do you think of the consultant's prediction? Why?

# Practice Problem #5 Solution

a) That the sample was randomly obtained, and large enough sample
b) (\$122.20, \$129.80)
c) I am 90% confident that the true mean daily fees that will be generated by the new parking garage is between \$122.20 and \$129.80.
d) That if we were to obtain all possible samples of size 44, 90% of them would capture the unknown mean daily fee for the parking garage.
e) The consultant’s prediction was fairly accurate since \$128 is included in the confidence interval.