# Lab Confidence

Abhi Thanvi, Paul Holaway

July 12th, 2022

## Contents

# Lab Confidence

## Welcome

Just like learning a new spoken language, you will not learn the language without practice. Labs are an important part of this course. Collaboration on labs is **extremely encouraged**. If you find yourself stuck for more than a few minutes, ask a neighbor or course staff for help. When you are giving help to your neighbor, explain the **idea and approach** to the problem without sharing the answer itself so they can figure it out on their own. This will be better for them and for you. For them because it will stick more and they will have a better understanding of the concept. For you because if you can explain it to other students, that means you understand it better too.

## The Idea of this Lab

I can confidently say that Confidence Interval is a huge chunk of statistic that falls under the statistical inference section. Through this lab, you will be exploring the application and calculation of Confidence Intervals using `R`. Understanding Confidence Interval (CI) is crucial to understand what we do for the time that is left for us. I am not going to huge description for today's lab as you probably understand the importance of this topic and I also want to give you time to do this lab and maybe ask questions for project. LET'S GO!

**"Confidence Interval is like giving an estimated range for answer, which you aren't 100% sure about" - Woke Abhi**

## Problem 1: Z vs. t, You Tell Me

**Question 1:** You are trying to find the 98% confidence interval from a population whose size is 100. You know the Standard Deviation too! Do you use Z-Interval or t-interval?

**Answer:** (Student Response Here) You can use Z-interval.

**Question 2:** You are trying to find the 85% confidence interval from a population whose size is 1000. You do not know the Standard Deviation. Do you use Z-Interval or t-interval? If you choose Z-interval, explain why (Hint: Something we learned in the last few classes)?

**Answer:** (Student Response Here) You can use Z-interval or t-interval in this case. Using CLT, we can assume that our sample distribution is normal, thus we could use Z-interval even if we do not know the standard deviation.

**Question 3:** You are trying to find the 95% confidence interval from a population whose size is 10. You do not know the Standard Deviation. Do you use Z-Interval or t-interval?

**Answer:** (Student Response Here) You can use t-interval.

**Question 4:** You are trying to find the 90% confidence interval from a population whose size is 5. You do know the Standard Deviation. Do you use Z-Interval or t-interval?

**Answer:** (Student Response Here) You can use Z-interval.

## Problem 2: Age Is Just an Interval

**Question 1:** Abhi wants you to create a 90% and 98% confidence interval for the average age for a batter in MLB. Again, I don't know Abhi's fascination with age of people, but let's just ignore that. The data for this is in the `MLB_Batters` data set. Create the two confidence intervals.

**Answer:**

```
#Imports
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.7     v dplyr   1.0.9
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
batters <- read.csv("~/Desktop/DPISu22/Data Sets/MLB_Batters.csv", stringsAsFactors=TRUE)
#Common Setup Work
xbar = mean(batters$Age)
n = length(batters$Age)
sigma = sd(batters$Age)*sqrt((n-1)/n)
#90% CI Work
c1 = (1-0.90)/2
Z1 = qnorm(c1,lower.tail = FALSE)
CI_1 = xbar + c(-1,1)*Z1*sigma/sqrt(n)
#98% CI Work
c2 = (1-0.98)/2
Z2 = qnorm(c2,lower.tail = FALSE)
CI_2 = xbar + c(-1,1)*Z2*sigma/sqrt(n)
#Print Answers
CI_1
```

```
## [1] 28.36029 28.82008
```

```
CI_2
```

```
## [1] 28.26504 28.91533
```

**Question 2:** With your group, discuss what confidence interval mean? Write the answer to what your second confidence interval means in context from question 1.

**Answer:** (Student Response Here) We are 98% confident that the true average age of MLB batters is between the age of 28.26504 and 28.91533.

## Problem 3: Shoes!

**Question 1:** Abhi is trying to convince Paul that people own more than 3 pairs of shoes on average. Help Abhi do so by creating a 95% confidence interval for the average number of shoes that students own. Use the `hello` data set. Remember that this is a sample of UIUC STAT107 students.

**Answer:** (Either Z or t is fine here.)

```
#Imports
hello <- read.csv("~/Desktop/DPISu22/Data Sets/hello.csv", stringsAsFactors=TRUE)
#Common Setup Work
xbar = mean(hello$Shoes)
n = length(hello$Shoes)
c = (1-0.95)/2
#Z CI
Z = qnorm(c,lower.tail = FALSE)
sigma = sd(hello$Shoes)*sqrt((n-1)/n)
CI_Z = xbar + c(-1,1)*Z*sigma/sqrt(n)
# t CI
df = n - 1
t = qt(c,df,lower.tail = FALSE)
s = sd(hello$Shoes)
CI_t = xbar + c(-1,1)*t*s/sqrt(n)
#Print Answers
CI_Z
```

```
## [1]  7.564998 10.372502
```

```
CI_t
```

```
## [1]  7.552353 10.385147
```

**Question 2:** With your group, discuss what confidence interval mean? Write the answer to what your confidence interval means in context from question 1.

**Answer:** (Student Response Here) We are 95% confident that the true average number of shoes for UIUC STAT107 students is between 7 and 10 (approximately) pairs.

## Problem 4: Narrower or Wider

**Question 1:** Assuming the standard deviation remains the same. If I increase my confidence level from 90% to 95%, would my confidence interval be wider or narrower? In other words, will the range of my interval be larger or smaller?

*The answer should be contextualized or in your own words. Providing numbers or math is not recommended*

**Answer:** (Student Response Here) The confidence interval will be have to be wider because if "confidence" increases, I need to widen the range of the interval to account for sampling error. In other words, if the interval is widened I can be more confident therefore increasing my Confidence level forces the CI to widen.

**Question 2:** Assuming the standard deviation and CI remains the same. If I increase my sampling size, would my confidence interval be wider or narrower? In other words, will the range of my interval be larger or smaller?

*The answer should be contextualized or in your own words. Providing numbers or math is not recommended*

**Answer:** (Student Response Here) The confidence interval will be have to be narrower because as my sample size increases my sampling error or standard error decreases. This shrinks the range (or the extreme/endpoint values) of our confidence interval. Therefore, as the sample size increases, the confidence interval becomes narrower.

**Question 3:** After finishing the entire lab. With your group, discuss and report how to change the width of your confidence interval. Keep in mind, we cannot change the population and it's standard deviation.

*Hint: Think about how Confidence Level and Sample Size affect the specific parts of Confidence intervals?*

**Answer:** (Student Response Here) Width of the confidence interval depends on Confidence Level, the higher the confidence the wider the interval, the lower the confidence the narrower the interval. Secondly, Standard/Sampling Error, the higher the sample size, the lower the Standard Error thus narrower the interval. You can deduce what happens with the opposite.

## Project Questions

Feel free to work on your project if there is any time left after the labs. Paul and I are here to answer any questions during the second half of the lab times to answer mainly project related questions, but general questions are more than welcome too. Feel free to discuss among your group about any project ideas or help each other out. Remember collaboration is promoted, plagiarism is not! :)

## Submission

Once you have finished your lab...

1. Go to the top left and click `File` and `Save`.
2. Click on the `Knit` button to convert this file to a PDF.
3. Submit **BOTH** the `.Rmd` file and `.pdf` file to Blackboard by 11:59 PM tonight.