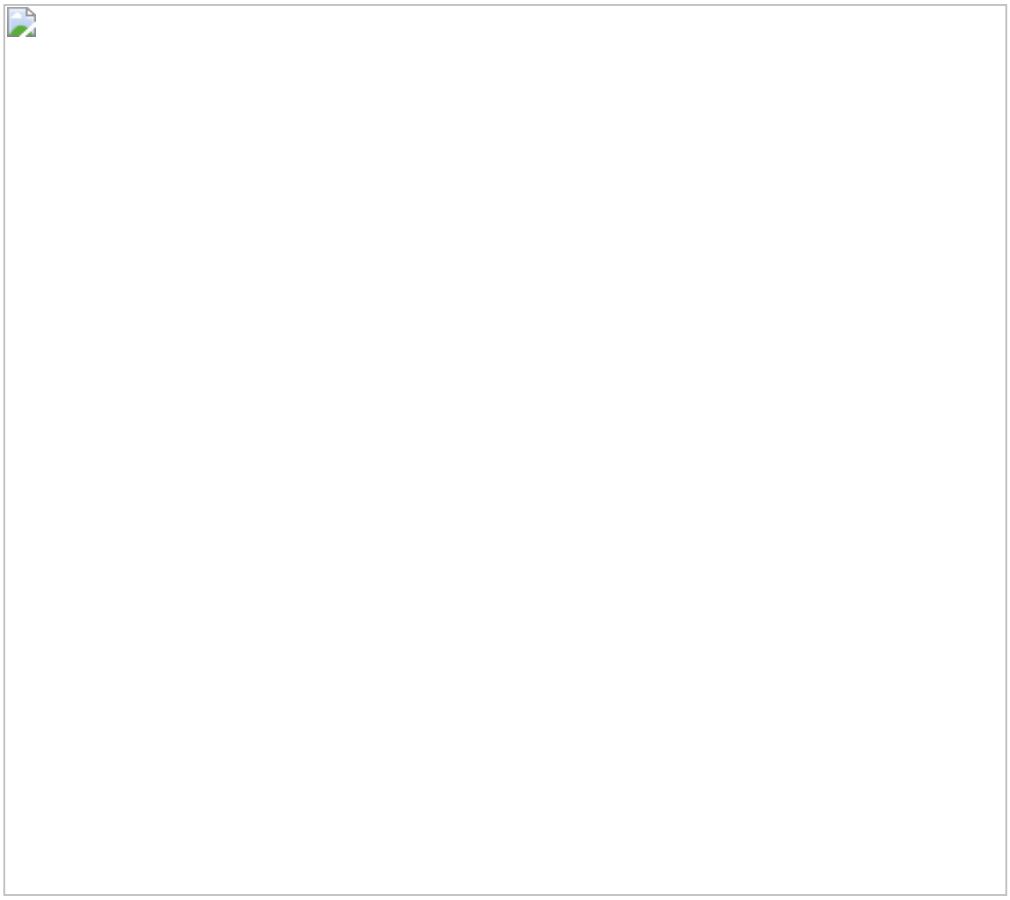
Reflect on the Question | Analyze the Data | Draw Conclusions

Pre-Lab 1: UT Student Survey Data



In this lab, we will examine how *sample* data can be used to discover the truth about a *population*. Our population data consists of data we collected from our statistics students here at The University of Texas at Austin. They told us several things about themselves, including how happy they are and the amount of time they study. We'll run a few simulations on this data to see if we can replicate what the Central Limit Theorem tells us about sampling. We are pretending that we don't know the "true" population parameters, but in fact we do!

Primary Research Question

How many letters long is the typical UT student's name? How does our estimate change as we increase the size of our sample?

(3/3 points)

Check the Data

Let's begin by examining our data in R.

1. Open RStudio. Make sure you've installed the SDSFoundations package (Version 1.4).
2. Type **`library(SDSFoundations)`**. This will automatically load the data for the labs. (Note that the console may produce a warning message when this command is run. If SDS Foundations has been properly installed, this message can be ignored.)
3.Type **`survey <- StudentSurvey`**. This will assign the data to your Workspace.
4. Look at the spreadsheet view of the data to answer the following questions.

Alternatively, you can use follow the steps in the "Importing a Data Frame" R tutorial video, and use the StudentSurvey.csv file. (Right-click and "Save As.") Make sure to **name** the dataframe "survey" when importing.

1. Open RStudio.
2. Click on "Import Dataset" button at the top of the workspace window. Choose *"from text file."*
3. Click on the location of the StudentSurvey.csv file you just downloaded.
4. Click on the StudentSurvey.csv file. Then, click Upload.
5. Look at the spreadsheet view of the data to answer the following questions.

1a) How many students are in this dataset?

379 ✔ **Answer:** 379

379

1b) How many of the first 10 students in the dataset had names longer than 5 letters?

5 ✔ **Answer:** 5

5

1c) How long is the name of the first student in the dataset who is happy less than 40% of the time?

7 ✔ **Answer:** 7

7

Click here for a video explanation of how to answer this question.

You have used 1 of 1 submissions

(2/2 points)

Check the Variables of Interest

Let's find the variables we need to answer the question.

2a) Which variable tells us how many letters are in each student's first name? The name of this variable in the dataset is:

name_letters ✔ **Answer:** name_letters

2b) What type of variable is this--categorical or quantitative?

quantitative ✔ **Answer:** quantitative

Click here for a video explanation of how to answer this question.

You have used 1 of 1 submissions

(2/2 points)

Reflect on the Method

Which method should we be using for this analysis and why?

3a) What makes something a **sampling** distribution?

◉ It is a distribution of sample *statistics*, such as a distribution of sample means. ✔

- It is a distribution of all the *possible values* in a population.
- It is a distribution of all the *observed values* in a sample.

3b) What does the **Central Limit Theorem** predict about a sampling distribution of means?

- The distribution looks more and more Normal as you draw larger samples.
- The sample means become less variable as your sample size increases.
- You will find the population mean at the center of the sampling distribution.
- All of these ✔

Click here for a video explanation of how to answer this question.

You have used 1 of 1 submissions