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!

(3/3 points)

Review of the Central Limit Theorem

In this lab, you will use a **simulation** to better understand the Central Limit Theorem. Let's start by remembering the key features of the **Central Limit Theorem**.

1a) In this lab, we will draw samples to answer the following question: *What percentage of the time are college students happy?* How does the **Central Limit Theorem** predict our answer to this question will change as sample size increases?

- (•) As sample size increases, our sample means should become less variable and be closer to the true mean. ✔
- ( ) Increasing sample size will have no impact on the value of our sample means.
- ( ) As sample size increases, the sample means we draw will all be equal to the population mean.
- ( ) As sample size increases, our sample means should become more variable and less accurate estimators.

1b) What does it mean to increase the **sample size** in a simulation?

- ( ) It means to run more iterations on the data.
- (•) It means to draw more individuals in each of our samples. ✔
- ( ) It means to draw from a larger population.
- ( ) It means to draw more samples from our population.

1c) What should be true about our **sampling distributions** as we increase our sample size?

- ( ) The means should decrease and the standard errors should remain about the same.

- The means and the standard errors should increase.
- The means and the standard errors should decrease.
- The means should remain about the same, but the standard errors should decrease. ✔

You have used 1 of 1 submissions

Lab Preparation

In this lab you will be working with data from the UT Student Survey.

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.
3.Type **`survey <- StudentSurvey`**. This will assign the data to your Workspace.

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.

Feel free to use the script from the week's PreLab, which you can modify for use in this Lab.