Pre-Lab 3: Post Student-Survey Data

Students at The University of Texas at Austin answered a set of questions for us at the beginning of the semester and then again at the end. We'll use this data to compare different groups, and to explore what has (or has not) changed over time for these students. (Please note that in the United States, the labels of "freshmen," "sophomore," "junior," and "senior" designate whether a student at a four-year university or college is in their first, second, third, or fourth year respectively.)

Primary Research Questions

1. Who is happier at the beginning of the semester: under-classmen or upper-classmen?
2. Does student happiness change from the beginning of the semester to the end?

R Tutorial Videos

Pre-Lab
Pre-Lab due May 03, 2016 at 17:00 UTC

Lab
Lab due May 03, 2016 at 17:00 UTC

Problem Set
Problem Set due May 03, 2016 at 17:00 UTC

- Week 4: Hypothesis Testing (Categorical Data)

(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.
2. Type `library(SDSFoundations)`. This will automatically load the data for the labs.
3.Type `post <- PostSurvey`. 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 follow the steps in the "Importing a Data Frame" R tutorial video, and use the PostSurvey.csv file. (Right-click and "Save As.") Make sure to **name** the dataframe "post" when importing.

1. Open RStudio.
2. Click on the "Import Dataset" button at the top of the workspace window. Choose *"from text file."*
3. Click on the location of the PostSurvey.csv file you just downloaded.
4. Click on the PostSurvey.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 the dataset?

214 ✔

Answer: 214

214

1b. What is the classification of the first male student? (Make sure your spelling matches the variable outcome as spelled in the dataframe.)

Freshman ✔

Answer: Freshman

1c. Of the first 10 students in the dataset, what percentage live on campus? (Report without the "%" sign.)

50 ✔

Answer: 50

50

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

You have used 1 of 1 submissions

(6/6 points)

Check the Variables of Interest

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

2a. Which variable tells us whether a student is a **lower-classman** (freshman or sophomore)?

The variable name in the dataset is classification ✔

Answer: classification , which is a categorical ✔

Answer: categorical variable.

2b. Which variable tells us how **happy** students were at the **beginning** of the semester?

The variable name in the dataset is happy ✔ **Answer:** happy , which is a quantitative ✔ **Answer:** quantitative variable.

2c. Which variable tells us how **happy** students were at the **end** of the semester?

The variable name in the dataset is post_happy ✔

Answer: post_happy , which is a quantitative ✔

Answer: quantitative variable.

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. We will use an **independent t-test** to help us compare the happiness of the under and upper-classmen. Why?

- We are comparing two means, so it must be an independent t-test.
- We want to see the change in happiness as students go from freshman to senior year.
- We want to determine if there is any kind of relationship between lower and upperclassmen.
- We want to compare the happiness of two different populations of students. ✔

3b. We will use a **dependent t-test** to help us determine whether happiness levels changed over the semester. Why?

- We want to see the change in happiness as students go from freshman to senior year.
- We are looking for a change over time for the same group of students. ✔
- We want to determine if happiness is dependent on what month of the year it is.
- We are looking at a difference between means, so it must be a dependent t-test.

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

You have used 1 of 1 submissions