

MATH 145: Biostatistics

Assignment #6

1. Go to the [StatKey](#) website. In the upper middle part of the page, select "CI for Single Mean, Median, St. Dev." There is pre-loaded data when you first arrive at the ensuing page, data on the prices of used Ford mustangs (cars). Take a few minutes orienting yourself to what you see, select "Show Data Table," and note the locations of these numbers on the dot plot you see under "Original Sample."
 - (a) In the app, with the "Bootstrap Dotplot" initially cleared, click "Generate 1 Sample." Look at the dotplot and summary statistics under "Bootstrap Sample" as well as the single dot that appeared on the "Bootstrap Dotplot".
 - i. Describe, as if it was done by a person using a bag and slips of paper, what process was followed to generate that dot.
 - ii. The same data is available to you in RStudio with the **Lock5withR** package loaded, in the Price column of the data frame `MustangPrice`. Give a command that would simulate the process you described using bags and slips.
 - (b) Back in the StatKey app, you can hit "Generate 1000 Samples" several times to fill out the Bootstrap Dotplot.
 - i. What alteration(s) to your RStudio command above is required in order to store in a data frame called `manyBstrapXbars` thousands of numbers like those you just generated as dots on your bootstrap dotplot? Be specific about the new command.
 - ii. Assuming you've done the last part, what command gives you an approximate standard error, similar to the std. error displayed in StatKey in the upper right corner of the bootstrap dotplot?
- [2.] [Work for this one should not be handed in.] Starting again from the [StatKey](#) website, select "CI for Single Proportion." The app you are taken to has as its pre-loaded data set one called "Mixed Nuts." In "Original Sample", you can see the sample was a collection of mixed nuts of size $n = 100$, with 52 of them being peanuts. So, the sample proportion is $\hat{p} = 0.52$.
 - (a) With the "Bootstrap Dotplot" initially cleared, click "Generate 1 Sample." Look at the information under "Bootstrap Sample" as well as the single dot that appeared on the "Bootstrap Dotplot".

- i. Describe, as if it was done by a person using a bag and slips of paper, what process was followed to generate that dot.
 - ii. Give an R command that would simulate the process you described using a bag and slips. [You do not need to construct a raw data set to accompany the original sample, in this instance; not, at least, if you use `rflip()`.]
- (b) Back in the StatKey app, you can hit "Generate 1000 Samples" several times to fill out the Bootstrap Dotplot.
 - i. What alteration(s) to your RStudio command above is required in order to store in a data frame called `manyBstrapPhats` thousands of numbers like those you just generated as dots on your bootstrap dotplot? Be specific about the new command.
 - ii. Assuming you've done the last part, what command gives you an approximate standard error, similar to the std. error displayed in StatKey in the upper right corner of the bootstrap dotplot?
- 3. From the StatKey menu page, select "CI for Difference in Proportions." The app here has pre-loaded data called "Use Text Messages (by Age)." In "Original Sample", you see that $\hat{p}_T = 0.87$ (696 of $n_T = 800$ respondents) is the sample proportion among teens that *text*, while for adults it is $\hat{p}_A = 0.72$ (1621 of $n_A = 2252$ respondents). We wish to estimate $p_T - p_A$, and our point estimate, also displayed under "Original Sample", is $\hat{p}_T - \hat{p}_A = 0.15$.
 - (a) With the "Bootstrap Dotplot" initially cleared, click "Generate 1 Sample." Look at the information under "Bootstrap Sample" as well as the single dot that appeared on the "Bootstrap Dotplot".
 - i. Describe, as if it was done by a person using bags and slips of paper, what process was followed to generate that dot.
 - ii. Give R commands that would simulate the process you described using a bag and slips. [You probably do need to construct a raw data set to accompany the original sample in this instance.]
 - (b) Back in the StatKey app, you can hit "Generate 1000 Samples" several times to fill out the Bootstrap Dotplot.
 - i. What alteration(s) to your RStudio command above is required in order to store in a data frame called `manyBstrapPhatDiffs` thousands of numbers like those you just generated as dots on your bootstrap dotplot? Be specific about the new command.

- ii. Assuming you've done the last part, what command gives you an approximate standard error, similar to the std. error displayed in StatKey in the upper right corner of the bootstrap dotplot?
- [4.] [Work for this one should not be handed in.] From the [StatKey](#) menu page, select "CI for Difference in Means." The app comes pre-loaded with data called "Exercise Hours (Male - Female)." I have demonstrated the use of RStudio to bootstrap on the difference of means $\bar{x}_1 - \bar{x}_2$ using this very example, both in an [abbreviated version here](#) (see the last page), and in a [wordier manner here](#). Using the StatKey app as appropriate, work toward a solid understanding of those RStudio commands/results.
- Read Chapter 5 of the Ellenberg book. Then submit answers to the [Socrative questions](#).
 - From the Exercises for Section 4.1 (hand in those which are underlined): Do Problems [4.5](#), [4.6](#), [4.7](#), 4.9, 4.11, 4.13, [4.14](#), 4.15 parts (a), (c) and (d), [4.16](#), 4.17,, 4.19 parts (a) and (b), [4.21](#), [4.22](#), [4.24](#), [4.25](#), and [4.28](#).