**Statistics 200: Lab Activity for Section 3.3**

**This lab is the students’ first opportunity to truly create and use bootstrap samples. The TA needs to do some prep work for lab by placing a big stack of paper squares at the front of the classroom for students to take and use. The TA should also provide a few pencils in case students forget to bring a writing utensil. All TAs and LAs should be *very* familiar with bootstrap procedures and explanations in order to help students – GO TO LECTURE!!**

**Constructing Bootstrap Confidence Intervals - Learning objectives:**

* Describe how to select a bootstrap sample to compute a bootstrap statistic
* Recognize that a bootstrap distribution tends to be centered at the value of the original statistic
* Use technology to create a bootstrap distribution
* Estimate the standard error of a statistic from a bootstrap distribution
* Construct a 95% confidence interval for a parameter based on a sample statistic and the standard error from a bootstrap distribution.

**Activity 1: Create a bootstrap confidence interval for the average number times adults laugh in a day**

**This activity gets students actively making their own bootstrap samples. They must consider how they can randomly draw from the sample with replacement using the physical tools of index cards. Eventually, they should realize that they can do this by writing one of the data values on each scrap of paper, randomly drawing one, replacing it, and repeating until 6 values have been drawn. Remind students that it is fine for the same value to appear more than once!**

Let’s say we conducted a very small survey to estimate the average number of times adults laugh in a day. We asked 6 people to record how many times they laughed one day and got the following data values:

18, 24, 4, 30, 8, 42

Our goal in activities 1, 2, and 3 today are to carefully construct a 95% bootstrap confidence interval for the mean number of laughs per day.

1. What is the population parameter we are trying to estimate?

**Average number of times adults laugh in a day (μ)**

1. Calculate the sample estimate for this parameter. Use correct notation!

**x̅ = (18 + 24 + 4 + 30 + 8 + 42)/6 = 126/6**  **= 21**

At the front of the lab are stacks of cut index cards. You are to use them to create your own bootstrap samples. Before you go up to get them, figure out how you can use them to complete this task.

1. How many index cards will you need? **6**
2. What do you put on the cards? **Values of the sample that are given**.
3. How do you use them to create a bootstrap sample? **I take a sample of size 6 from these cards with replacement. In stat 200-compatible language: We put the six cards facedown in a pile, or have a classmate hold them, or shuffle them, or something along those lines. Then we will draw one card out at random and record the value written on it. We then replace the card and repeat until there are 6 values recorded.**
4. What statistic should you compute from each bootstrap sample? **Mean ( x̅ )**
5. Use your cards to create three bootstrap samples and give them below.
6. What were the bootstrap statistics calculated from each of your three bootstrap samples?

**Calculate the sample mean**

1. For each set of values below, determine whether it is a possible bootstrap sample from the original sample of laughs per day. If not, state why not.

**This part of the activity is meant to get students recognizing the key parts of a bootstrap sample – that it has the same number of values as the original sample, and that all values of the bootstrap sample must come from the original sample. If students have difficulty, ask them if both of those criteria are met. If so, great, it qualifies. If not, well, they can then see why not!**

**Original Sample: 18, 24, 4, 30, 8, 42**

|  |  |  |
| --- | --- | --- |
| **Sample** | **Possible bootstrap sample? Yes or No** | **If not, state why not.** |
| 24, 42, 30, 8, 18 | No | because sample size should be 6 |
| 30, 8, 30, 24, 42, 18 | Yes |  |
| 9, 24, 4, 18, 31, 8 | No | 9 is not in the sample |
| 30, 24, 8, 4, 18, 42 | Yes |  |
| 18, 24, 18, 24, 18, 18, 24 | No | because sample size should be 6 |
| 8, 8, 8, 42, 8, 8 | Yes |  |

**Activity 2: Use Statkey to make bootstrap CI for laughter**

**Now students are taking the bootstrap sampling out of the physical realm and going big time with technology. As an assistant in the lab, make sure you are very familiar with using StatKey for this activity.**

As you can tell from Activity 1, creating bootstrap samples by hand takes a long time. That’s why we have software create bootstrap samples for us! StatKey can create as many bootstrap samples and statistics as we want very quickly. In this activity we will use StatKey to do just this to create a bootstrap distribution for the number of laughs.

Open up [Statkey](http://www.lock5stat.com/StatKey/) and select Bootstrap confidence interval for a mean, median, Std.Dev. You’ll need to enter in the data for the laughing adults. To do this, click ‘Edit Data’, erase what’s in there, and enter the values from the original sample in Activity 1.

You can choose if you want to include a header row (name of variable) or not. Just be sure to select the correct option below your data.

1. Create a bootstrap distribution of 5,000 bootstrap statistics.
   1. what is the shape of the distribution?

**Symmetric and bell-shaped. (there’s a bit difference in the mean and median though which we can ignore I suppose)**

what is the standard error? **It is about 5.171, although actual values will differ somewhat from student to student.**

1. Use the standard error from part (b) above and the sample mean from activity 1 create a 95% confidence interval for the population parameter.

**x-bar – 2\*s to x-bar + 2\*x = (21-2\*5.171, 21+2\*5.171) = (10.658, 31.342)**

1. Interpret your interval estimate in context.

**We are 95% sure that the mean number of times an adult laughs in a day is between 10.7 to 31.3**

1. What is something we could do to make the interval narrower? **Increase the sample size**

**Activity 3: Mood of the nation**

**This is an opportunity for students to create a confidence interval for a more complicated parameter (difference in proportions). Students will need to extract the important numbers from the prompt and enter them correctly into StatKey, then create a bootstrap distribution to find the SE of their sample statistic (p-hat1 – p-hat2). It is very important that TAs and LAs know how to correctly do *all* these steps! Use StatKey before your lab to complete this activity.**

Gallup conducted a nationwide poll from January 4-8, 2017 to gauge public opinion on the question ‘will America be better off in 2020?’ 1,032 people took the survey. Exact numbers were not available, but assume that 416 people in the survey identified as democrats, of which 58 thought that America will be better off in 2020. Assume 502 survey participants identified as republican, and of these 427 thought America will be better off in 2020.

Our goal in this activity is to calculate a 95% confidence for the difference in proportion that think America will be better off when comparing republicans to democrats. Let group 1 be republicans and group 2 be democrats.

1. What population parameter are we trying to build an interval estimate for? Use correct notation: **p1 – p2, the population difference between republicans and democrats in the proportion thinking that America will be better off in 2020.**
2. Calculate the sample estimate for this quantity. Use correct notation.

**= 427/502 – 58/416 = 0.71**

1. Now we use Statkey to build a bootstrap confidence interval. Create a bootstrap distribution using at least 5,000 bootstrap samples.

What is the standard error for the original sample statistic?

**0.023 for me, but answers will vary somewhat (The standard error is read off the upper-right corner of the bootstrap distribution dotplot).**

1. What is the center of your bootstrap distribution? How does it compare to the sample estimate from the original sample?

**0.711. The original sample estimate is the same. Note this does not always happen. So always use the orginal sample statistic.**

1. Compute the 95% confidence interval for the population parameter.

**(0.711 – 2\*.023 to 0.711 +2.023) = (0.665 to 0.757)**

1. Interpret the confidence interval you computed in question 5.

**We are 95% sure that the true difference in proportion of people thinking America will be better off in 2020 is between 0.665 and 0.757 when comparing republicans to democrats.**