Are you sure about that?

Unit 3 - Lab 4

Directions: Follow along with the slides and answer the questions in **BOLDED** font in your journal.

Starting out

- During our previous lesson, we learned about creating informal confidence intervals based on sample medians.
- Today, we will be creating many intervals in an attempt to identify the *best* one.
- For this lab, you will need to load the `atus` data set. We will be working with the `age` variable again.
- **What is the median age in this data set?**

Getting one sample

- We first need to subset the data so we are only using the `age` variable. As in lesson 12, we can create a new data set called `justAge` to store this data.

```
justAge <- print(~age, data=atus)
```

- **What does the code `print(~age, data=atus)` tell RStudio to do?**

- Now we can create a sample of size 100 from the data set of just ages. Recall:

```
sample1 <- sample(justAge, size = 100,
                  replace = FALSE)
```

What's your sample median?

- Calculate the median age of your 100 sampled ages. **What was your sample's median age?**

- **How far off from the true median age is your estimate?**

- **Was your estimate too high or too low?**

Let's create one interval

- **Based on your answers from the previous slide, if you were to create an interval that includes the true median, how far away would you need to go in one direction?**
- For example, if my sample median was 50, I would need to go at least 3 numbers below 50 to include the true median of 47.
- But, I can't just go down 3. I also need to go up 3 to create my interval in both directions. Therefore, my actual interval would be from 47 to 53.
- **What values make up your interval?**

Calculating the range of your interval

- In order to see how wide the interval is, we can simply find the range of the upper and lower values. (e.g. if the interval was 47 to 53, the range = 53 - 47 = 6).
- **What is the range of your confidence interval?**

A quick calculation

- Note that another way to calculate the range is to find the difference between your sample's median and the true median, then multiply that value by 2.
- In RStudio, we could use the following code:

```
range1 <- abs(samp1Median - 47) * 2
```

- **Why is it necessary to include the syntax `abs()` here?**

Remember: Our goal is to come up with ONE confidence interval

- How do we know if the range we calculated from our first sample is the best?

- Obviously, we would like each range to be as small as possible and still include the true median.

- It's impossible to decide what the best interval would be with just one sample estimate, so we can create multiple samples and find the average ranges of all of them.

- **Record the code you would use to create 50 samples of size 100.** (Note: be sure to include how you would store the median value of each sample.)

Lots of ranges

- Now that you have 50 samples and their median ages, you can calculate the range for each sample's confidence interval.
- **Record the code you used to create a new column, called `range` in your data set.**

Now what do we do?

- We now have 50 different ranges for our confidence interval.
- **What is the average range of your 50 samples?**
- **Is this value bigger or smaller than you expected?**
- Compare your average range to your neighbor's. **Whose interval is better? Explain.**

Does it work?

- Basically, we just created a general rule for how far away an interval needs to go (in either direction) from a sample median in order to include the true median.

- Based on your rule (or range), if a sample produced a median age of 51, would your confidence interval include the true median age of 47?
- Do you think your confidence interval is good? Explain.