Chapter 1 (3): Confidence Interval and Bootstrapping

September 7, 2016

1. Previously in Stat 217…

- A subset of MockJury (with Unattractive and Average levels in `Attractiveness` variable) was used to test the hypothesis that there is no discrimination between these groups.
- Two methods of testing was used
- Non-parametric test: permutation test by shuffling the groups
- Parametric test: two independent sample t-test
- Their p-value is very similar (~0.033), giving the decision to reject the null hypothesis and conclude that there is indeed difference between sentence years between these groups.
- Scope of inference: non-causal to the sample

2. Confidence Interval

Confidence interval gives an of estimated range of values for the population parameter under some confidence levels. The meaning of "confidence level" is if people replicate the studies and construct confidence intervals, the proportion of such intervals that contain the true value of the parameter will match the given confidence level.

What is the meaning of confidence in the statement "[Pew Research Center] is 95% confident that from 19 to 35 percent of American adults from 18-24 years old have used an online dating site or mobile dating app."?

If more and more samples of 18-24 years old American adults are collected and 95% confidence intervals for the percentage of them who have used an online dating site or mobile dating app are created, 95% of them will cover the true percentage.

Hypothesis testing gives us a yes/no answer (reject/fail to reject) to a statement. It can't give us an **estimate** about the parameter of interest.

Ex. Which claim gives you more information?

- ○ **On average, an American female is higher than a Chinese female.**
- ◉ **An average American female is from 1 to 3 inches higher than an average Chinese female.** ✔

Submit | Clear

In this activity we're going to take one last look at the MockJury data. Here we're going to look at two different ways to build a confidence interval for the difference in true mean sentence lengths between the average and unattractive women. Once again, our two methods are (1) non-parametric (bootstrap) and (2) parametric (theoretical distributions).

3. Method 1: Bootstrapping

Bootstrapping is a **non-parametric** method.

What is a non-parametric method?

- ( ) **It involves a known theoretical distribution (e.g., normal, t, etc.)**
- (•) **The reference distribution is created by simulation.** ✔

Submit Clear

The purpose of bootstrapping is to simulate the replication of the experiment by re-sampling our data with replacement. The following video shows you an example of two bootstrap re-samples. Watch and answer the following questions.

Choose all that apply

- [x] **The size of a re-sample is the same as the original sample** ✔
- [ ] **Re-sample is exactly the same as the original sample.**
- [x] **Re-sample can only take values from the original sample.** ✔
- [x] **Some values may appear more than once, while others don't appear at all.** ✔

Submit Clear

If you use a deck of 8 cards to simulate sample with replacement, what should you do?

- ( ) **Continuously deal the cards until it's empty.**
- (•) **Deal one card, put it back, then repeat until you do it 8 times.** ✔

Submit Clear

R

In R, `resample()` function in the `mosaic` package will create a bootstrap resample. Look at the following code and output and confirm what the `resample()` do.

```
require(mosaic)
resample(c(1, 2, 3, 4, 5, 6, 7, 8))
```

```
## [1] 3 7 4 8 8 1 5 8
```

Apply to MockJury data set

As usual, we need to load necessary packages and modify the original `MockJury` data set to include only Average and Unattractive female defendants. The following code will help if you want to try the code by yourself

```
# Load packages
require(heplots)
require(mosaic)
require(beanplot)

# Load the data set
data(MockJury)
# Remove Beautiful group
MockJury2 <- MockJury[MockJury$Attr != "Beautiful", ]
MockJury2$Attr <- factor(MockJury2$Attr)
```

To create many bootstrap resamples (say, 1000) we need to instruct R to do the following steps. Recall that we are interested in the difference in mean sentence lengths between Average and Unattractive groups in `Attr`.

1. Create a variable `bootDiff` that contains 1000 empty slots to save the result.
2. Repeat the following process 1000 times
    a. Using bootstrap to create one resample from the original data.
    b. Find the difference in means of the resample and save it to its associated position in `bootDiff`.

The above algorithm suggests using `for` loop in R. This is the equivalent code

```
# number of bootstrap resamples we want
B <- 1000

# empty matrix to save the results
bootDiff <- matrix(NA, nrow = B)

for (b in 1:B) {
    resample.MockJury <- resample(MockJury2)
    bootDiff[b] <- diffmean(Years ~ Attr, data = resample.MockJury)
}
```

> Note: Whatever commands inside the loop has to be put inside the “block” with two curly brackets.

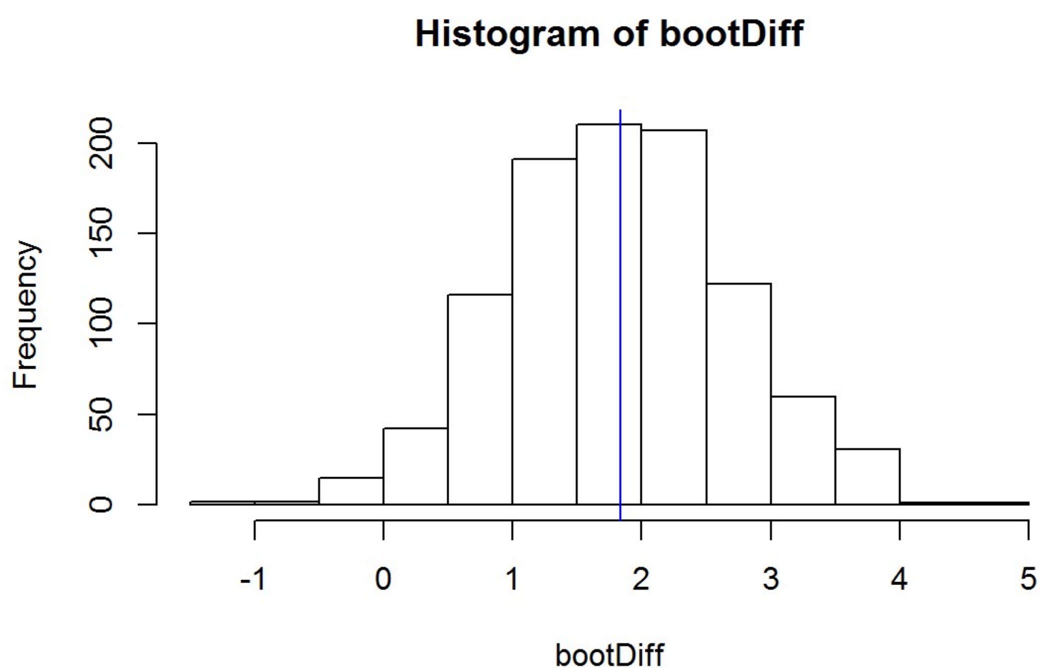
1. What variable will contain 1000 differences in means from 1000 resamples?

bootDiff

Submit Clear

Let's take a look at the histogram of 1000 bootstrap differences in means.

```
obsDiff <- diffmean(Years ~ Attr, data = MockJury2)
hist(bootDiff)
abline(v = obsDiff, col = "blue")
```



Compare the R code and the plot. What is the blue line in the plot? Why the bootstrap distribution is centered at the value of the blue line?

The observed difference in mean sentence years between Unattractive and Average groups (1.8). The distribution centers at that value because we bootstrap from the observed sample.

In order to create a confidence interval from this set of bootstrap statistics, we need to pick the appropriate percentiles from the distribution of the bootstrap statistics to include correct percent of values in the middle of the distribution.

Ex. If we want to create a 95% CI, what percentiles we would pick?

- ○ **0 and 95**
- ○ **5 and 100**
- ◉ **2.5 and 97.5** ✔
- ○ **5 and 95**

Submit Clear

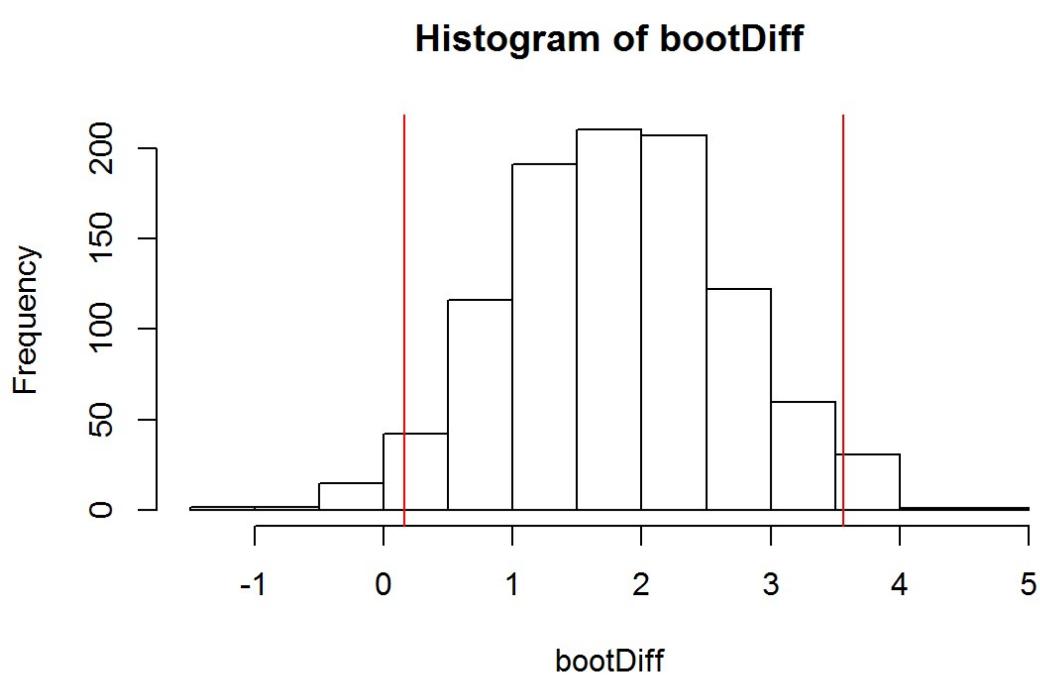
`qdata` (also in `mosaic`) will give us the confidence interval with the defined percentiles (**written in decimals**)

```
qdata(bootDiff, c(0.025, 0.975))
```

```
##        quantile     p
## 2.5%  0.1605945 0.025
## 97.5% 3.5648728 0.975
```

In the following plot, our 95% confidence interval is values between two red lines. The left one is at 2.5-th percentile and the right one is at 97.5-th percentile.

```
hist(bootDiff)
abline(v = qdata(bootDiff, c(0.025, 0.975))$quantile, col = "red")
```



What is the 95% confidence interval by bootstrap method?

(0.16, 3.56)

4. Method 2: t-test

4.1. Theory

Now we'll build a confidence interval for this difference in means using a parametric method: the two-sample t-based confidence interval. Again, this method requires these assumptions:

1. Independent observations.
2. Equal variances.
3. Normal distribution.

Please review Day 3 worksheet (https://vinhtantran.github.io/STAT217/D3-ttest/) for checking these assumptions for MockJury data.

The formula to find the confidence interval using two independent sample t-based is

$$\bar{x}_1 - \bar{x}_2 \pm t^*_{df} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \quad \text{where } s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$$

where $df = n_1 + n_2 - 2$, s_p is the pooled sample standard deviation. We can obtain the t^* multiplier using the following code. Recall we have 75 observations in two groups, so $df = 73$.

```
# 0.975 is the second percentile in 95% CI, 73 is the degree of freedom
qt(0.975, df = 73)
```

```
## [1] 1.992997
```

Let's compute the 95% confidence interval by hand. First, in order to calculate s_p , we need some of the output below.

```
favstats(Years ~ Attr, data = MockJury2)
```

```
##           Attr min Q1 median Q3 max     mean       sd  n missing
## 1      Average   1  2      3  5  12 3.973684 2.823519 38       0
## 2 Unattractive   1  2      5 10  15 5.810811 4.364235 37       0
```

What can you get from the above output?

1. n_1
   38
2. n_2
   27
3. s_1
   2.82
4. s_2
   4.36
5. t^*_{73}
   1.99
6. $\bar{x}_1$
   3.97

7. $\bar{x}_2$

5.81

Submit Clear

What is the command to compute s_p and CI in R? (See p.69 in textbook for solution)

s.p <- sqrt((n1 - 1)*s1^2 + (n2-1)*s2^2)/(n1 + n2 -2))

4.2. Lazy way

Similar to hypothesis testing, R has a "lazy" way to find the confidence interval. That's the same `t.test` function that we used last time.

```
t.test(Years ~ Attr, data = MockJury2, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  Years by Attr
## t = -2.1702, df = 73, p-value = 0.03324
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.5242237 -0.1500295
## sample estimates:
##     mean in group Average mean in group Unattractive
##                  3.973684                   5.810811
```

Note: The default confidence level of `t.test` is 95%. If you want to change it to say 90%, add the argument `conf.level = 0.9` to the function.

What is the confidence interval by t-based method?

(-3.52, -0.15)

Compare and contrast the confidence intervals here and the one found by bootstrapping. Why is the sign different? If we switch to have the same sign, which one is wider?

The sign is different because t.test uses opposite subtraction order from diffmean.
They are very similar, with the t-test method gives a bit wider interval.

5. Conclusion

1. Pick either confidence interval and interpret it in the context of the problem.

   We are 95% confident that the true difference in mean sentence years between Unattractive and Average groups is from 0.15 and 3.52 years.

2. What does this confidence interval tell us about the hypothesis test $H_0 : \mu_{attractive} - \mu_{average} = 0$ vs $H_a : \mu_{attractive} - \mu_{average} \neq 0$?

   Because 0 is not between the 95% confidence interval, we can reject the null hypothesis with significance level of 0.05. (Note that 0.95 + 0.05 = 1).

Take Home Messages

1. Interpreting a Confidence Interval

> We are _______ % confident that the true **parameter (in words) in context of the problem** lies between **lower CI** and **upper CI unit**.

2. Using CI for Hypothesis Testing

$H_0 : \mu_{unattractive} - \mu_{average} = 0$, so 0 is the null value

- If the null value is **within** the confidence interval, then the null value is **plausible**. This means we would **FAIL to reject** H_0 .
- If the null value is **NOT within** the confidence interval, then the null value is **NOT plausible**. This means we would **reject** H_0 .