

Confidence Interval of Proportion

Te Rutherford

January 29, 2015

Today

- ▶ We will look at confidence interval and what it means.

What is confidence interval? and why do we care?

- ▶ Confidence interval gives you the uncertainty of our estimate.
- ▶ When you say that a proportion is around 0.3 - 0.5, we want to make this more rigorous and formal.
- ▶ We want to say the we are 90% confident that the true proportion is in the interval.

Intuition

- ▶ We want to say that we are 90% confident that the true proportion is in $0.4 - 0.6$.
- ▶ If you make the interval bigger, then you can be more confident.

What do you mean by confidence?

- ▶ A 95% confidence interval means that if you run this survey/study/experiment many times and compute 95% confidence interval, then 95% of the intervals will cover the true proportion.
- ▶ Crap, this is so confusing. Let's look at an example.

Example

- ▶ Suppose we are curious about the proportion of Brandeis students that had personal tutoring during junior in high school.
- ▶ We have to conduct a survey to estimate the proportion.
- ▶ Each survey conductor computes the estimated proportion and the 95% confidence interval.
- ▶ Look at the picture I drew on the board.

How do we compute 95% confidence interval?

- ▶ Obviously, we use R.
- ▶ Suppose we are looking at the new medication data from last time.
- ▶ If a new medication works on 300 patients but does not work on 150 patients, the rate of medication being effective is $300 / (300 + 150)$
- ▶ BTW, rate is now called proportion because Manny asked for it. But it is really interchangeable.

Confidence interval in R

- ▶ We use `bcanon` function to compute confidence interval
- ▶ If we want $(100 - \alpha)\%$ confidence interval, then the lower bound is at the $\alpha/2$ percentile, and the upper bound is at the $100 - (\alpha/2)$ percentile.

```
library(bootstrap)
medication_data = c(rep(0,150), rep(1,300))
b= bcanon(medication_data, nboot= 10000, theta=mean)
b$confpoints
```

```
##      alpha bca point
## [1,] 0.025    0.6222
## [2,] 0.050    0.6289
## [3,] 0.100    0.6378
## [4,] 0.160    0.6422
## [5,] 0.840    0.6867
## [6,] 0.900    0.6933
## [7,] 0.950    0.7022
```


Problem 7.54 in the handout

- ▶ The proportion of people satisfy MSDS = $11/100$
- ▶ 150 people have been surveyed.
- ▶ First step: we want to create a data vector using `rep` command.
There are $150 * 11/100$ successes and $150 * 89/100$ failures.

```
msds_data = c(rep(0, 134), rep(1, 16))
```

Problem 7.54 in the handout (continued)

```
b= bcanon(msds_data, nboot= 20000, theta=mean)
b$confpoints
```

##	alpha	bca point
## [1,]	0.025	0.06000
## [2,]	0.050	0.06667
## [3,]	0.100	0.07333
## [4,]	0.160	0.08000
## [5,]	0.840	0.12667
## [6,]	0.900	0.13333
## [7,]	0.950	0.14667
## [8,]	0.975	0.16000

- ▶ For the confidence level we set $(100 - 5\%)$, the lower bound is $\frac{5}{2}$ th percentile and the upper bound is $100 - \frac{5}{2}$ percentile.
- ▶ So the 95% confidence interval is $[0.06, 0.16]$. We are 95% confident that the true proportion is in this interval.

Specify the percentile you want to begin with

```
b= bcanon(msds_data, nboot= 20000, theta=mean, alpha = c(0.025, 0.975))  
b$confpoints
```

```
##      alpha bca point  
## [1,] 0.025      0.06  
## [2,] 0.975      0.16
```

- ▶ By putting in `alpha = c(0.025, 0.975)`, we only get the percentiles that we want.
- ▶ If you don't put this in, you might not get the percentiles you want.