

# Stats in Practice #8, Confidence Interval

*Kellin Rumsey*

4/22/2019

## 1. The dataset

The CDI dataset contains demographic information for the 440 most populated counties in the United States. Data for a random sample of  $n = 100$  of these counties can be found on the course webpage in a file called `cdi_sample.csv`.

### Load the dataset

- Open Rstudio and open a new script.
- Load the dataset by typing: `cdi = read.csv('http://math.unm.edu/~knumsey/cdi_sample.csv')`

## 2. Total serious crimes

The variable total serious crimes represents a count of all violent crimes recorded in each county during a year. Get this variable in R by typing:

```
crimes <- cdi$TotalSeriousCrimes
```

1. Make a histogram of the crimes variable (using the `hist()` function). Do you think it is safe to use t-procedures for this data?

We can make the data look much more normal by using total serious crimes *per capita*. To do this, divide by population as follows.

```
crimes_pc <- cdi$TotalSeriousCrimes/cdi$Population
```

2. Make another histogram of this data. Do you think it is safer to use t-procedures now? **Include this plot in your writeup.**
3. Calculate the mean and standard deviation of crimes per capita.
4. For a 95% CI, find the approximate confidence multiplier  $t_*$  using the table. Also find it exactly by typing `qt((1-C)/2, df=?, lower.tail=FALSE)` where  $C$  is the appropriate confidence level.
5. Calculate a 95% confidence interval for  $\mu$ , the true average number of crimes per capita in US counties. (Use the exact value for  $t_*$ .)
6. Interpret this interval in words.