Stats in Practice #9, Population Proportions

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/~knrumsey/cdi_sample.csv')`

2. Poverty rates

In 2000 (the year this CDI dataset is from), the average poverty rate was about 10%. For each of the $n = 100$ counties in this dataset, check to see if the poverty rate is higher than average by typing:

```
high_poverty <- (cdi$PercentBelowPoverty > 10)
```

This gives a list of TRUE/FALSE values. We can get the sample percentage of counties above the national poverty rate by typing:

```
n <- 100 #Sample size
phat <- sum(high_poverty)/n #Sample proportion
```

1. Report the value of $\hat{p}$. Make sure you understand exactly what this value means.
2. For a 80% confidence level, find the confidence multiplier $z_\star$ using the table. Also find it exactly in R by typing: `qnorm((1-C)/2, lower.tail=FALSE)` where C is the appropriate confidence level.
3. Calculate a 95% CI for p, the true proportion of counties in the dataset which are higher than the national average.
4. Interpret this interval in words.