BSTAT 3322: Homework 5

by Craig W. Slinkman

October 12, 2015

# Q1. Republican primary politics

CBS Poll of Republican presidential candidates of October 11, 2015, 10:30 AM posy=ted on the internet at

<http://www.cbsnews.com/news/poll-trump-still-leads-carson-in-second/>

reports

Donald Trump continues to lead the field nationally in the race to become the Republican nominee for president. Twenty-seven percent of Republican primary voters support Trump, giving him a six point lead over his closest competitor, neurosurgeon Ben Carson (21 percent).

Later in the article CBS gives the sample size data:

This poll was conducted by telephone October 4-8, 2015 among a random sample of 1,251 adults nationwide, including 1,038 registered voters. Data collection was conducted on behalf of CBS News by SSRS of Media, PA. Phone numbers were dialed from samples of both standard land-line and cell phones.

Denote as the proportion of votes who favor Donald Trump. Denote the proportion of votes who support Ben Carson. You task is to compute a 95% confidence interval for using the bootstrap. Your report must address both the procedure and the interpretation of the confidence interval for expressed in percentages not proportions.

Your answer is worth 50 points in this question. In your answer you must satisfy the following requirements:

|  |  |
| --- | --- |
| Requirement | Points |
| Data source | 2 |
| Appropriate plot (using ggplot2 ) | 3 |
| Point estimate based on sample statistic | 2 |
| Analysis of bootstrap bias and remedial measures | 3 |
| Normal QQ-plot of bootstrap sample statistic and analysis | 5 |
| Selection of appropriate type of confidence interval (normal or bca) | 5 |
| Bootstrap confidence interval (10) | 10 |
| Interpretation of the point estimate and the confidence interval | 15 |
| Total | 40 |

Note that you must only do the computations but must indicate why you need to do the computations. Requirements without any explanations will result in a requirement score of zero.

Note that you must only do the computations but must indicate why you need to do the computations. Requirements without any explanations will result in a requirement score of zero.

## Q2. Cheating on caffene content in cola

Caffeine is added to Cola products to give "energy" to the drinkers cola drinks. The source of the Cola is a syrup that is used by the bottler when the soft drink is bottled.

Because of the nature of the cola market the bottlers traditionally have been locale independent business. Supper Colas has 23 bottlers. Two of these are owned by Super Colas and the rest are bottles are independently owned locale bottlers.

Super Cola is one of these bottlers. The most expensive ingredient in the cola bottling process is the syrup. This is because the price of the syrup includes not only material and labor costs but also support and advertising costs. This make the syrups the bottler uses very expensive.

The syrup is added by a special machine that adds the syrup to the bottles. The average amount of syrup can be altered by changing settings on the machine that dispenses the syrup into the bottles.

Because the syrup is expensive some locale bottlers are tempted to set the machine to lower the mean amount of syrup in each bottle. This effects the quality of the product and threatens the uniformity of the product on a national scale. It also lowers Super Cola's profits.

To counter this, Super Colas employs a Marketing Research team to sample the market by randomly purchasing Supper Cola from retail outlets. Then the amount of Syrup added to the market is measured by measuring the amount caffeine in the bottle by using Gas chromatography–mass spectrometry (GC-MS) is an analytical method that combines the features of gas-chromatography and mass spectrometry to identify different substances within a test sample. Because the bottles are labels with specific codes Super Cola can tie each observation to the bottler who made the sampled cola.

In order to improve the quality of Super Cola tries various strategies and experiments. One experiment they have run is to assess the effective of a quality promotional where they give owner promotional gifts for acceptable caffeine levels. They measure this by averaging the caffeine contents of the Super Cola bottle. They then compare the mean caffeine level for Super Cola against the Program factor. The question they need to address is did there managerial promotion program result in an increase of the caffeine levels?

Let be the mean of the caffeine levels for cola produced when the promotion program is in effect. Let be the mean level of caffeine when no promotion program is in effect. Super Cola wants to estimate the quantity .

## Drawing a sample To draw a sample load the R-script *drawSample.r* found in the *BSTAT3322-H05* folder on GitHub. Load this folder into your working directory for your project. Run the furniture script. You should find a file named *SuperCola.csv*.

## Variables

The variables in the *SuperCola* data file are given in the table below.

|  |  |  |
| --- | --- | --- |
| Variable | Type | Description |
| Bottler | Factor | The bottler who bottles sample bottle. |
| Owned | Factor | Is the bottler one of the Super Cola owned bottling facilities? |
| Program | Factor | Was this bottler bottled during the time the Bottler was participating in the quality promotion program? |
| Caffeine | Quantitative | The measurement of the caffeine found in the samples in mg/l bottle standardized to 12 ounces. |

## Requirements

The requirements for this analysis are given in the table below.

|  |  |
| --- | --- |
| Requirement | Points |
| Data source | 2 |
| Appropriate plot (using ggplot2 ) | 3 |
| Analysis of plots | 10 |
| Point estimate based on sample statistic | 2 |
| Analysis of bootstrap bias and remedial measures | 3 |
| Normal QQ-plot of bootstrap sample statistic and analysis | 5 |
| Selection of appropriate type of confidence interval (normal or bca) | 5 |
| Bootstrap confidence interval | 10 |
| Interpretation of the point estimate and the confidence interval | 10 |
| Analysis of the effectiveness of the quality promotion program | 10 |
| Total | 60 |

Note that you must only do the computations but must indicate why you need to do the computations. Requirements without any explanations will result in a requirement score of zero.

Each section must start with a short explanation of explanation of what aspect of the data your are investigating, followed by the computer analysis, and then your statistical conclusions.