HW3: https://github.com/Panda-nny/SDS315_HW3

Danny Pan
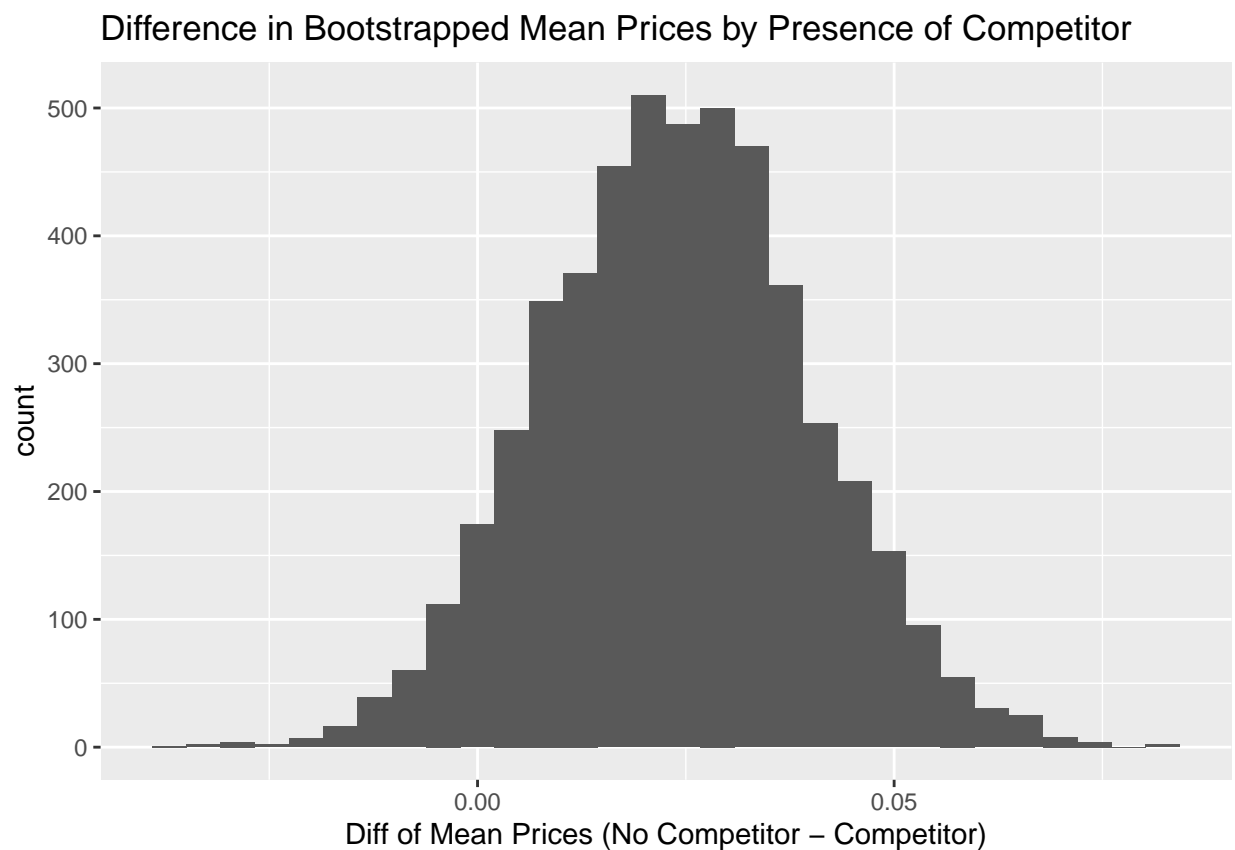
2025-02-09

Problem 1:

1.A)

Claim: Gas stations charge more if they lack direct competition in sight.

Evidence:

Difference in Bootstrapped Mean Prices by Presence of Competitor

Conclusion: According to our bootstrapped samples, 95% of the difference in means between gas prices of stations with no competitors vs. with competitors fall within the interval [-0.0067, 0.0553]. Since 0 is within the interval, there is no concrete evidence that gas stations charge more if they lack direct competition in sight.
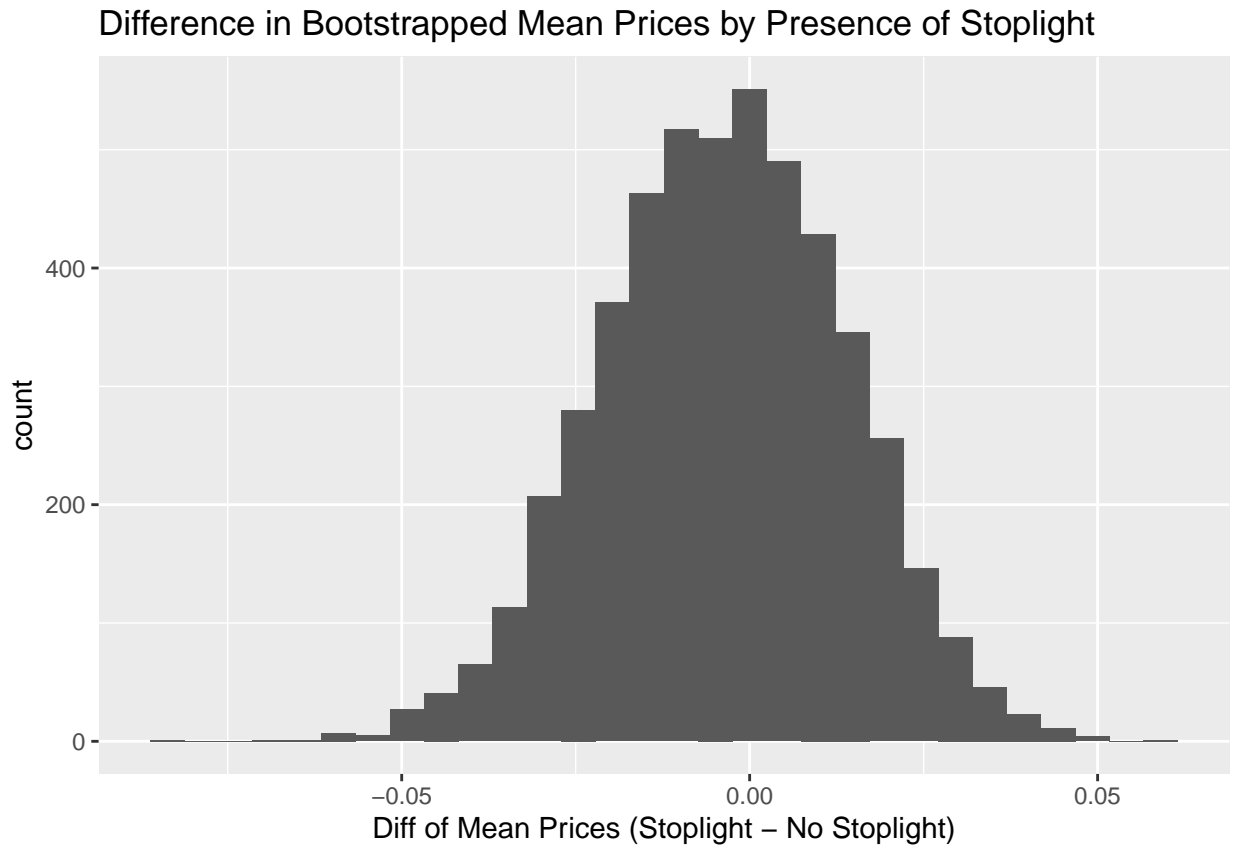
1.B)

Claim: The richer the area, the higher the gas prices.

Evidence:

Distribution of Gas Prices vs. Household Income

Conclusion: The scatterplot of gas prices vs. area income for the given sample suggests a weak relationship between the two variables. According to the r^2 value, only around 15.69% of the variation in gas prices can be explained by differences in the station's median ZIP code income.

1.C)

Claim: Gas stations at stoplights charge more.

Evidence:

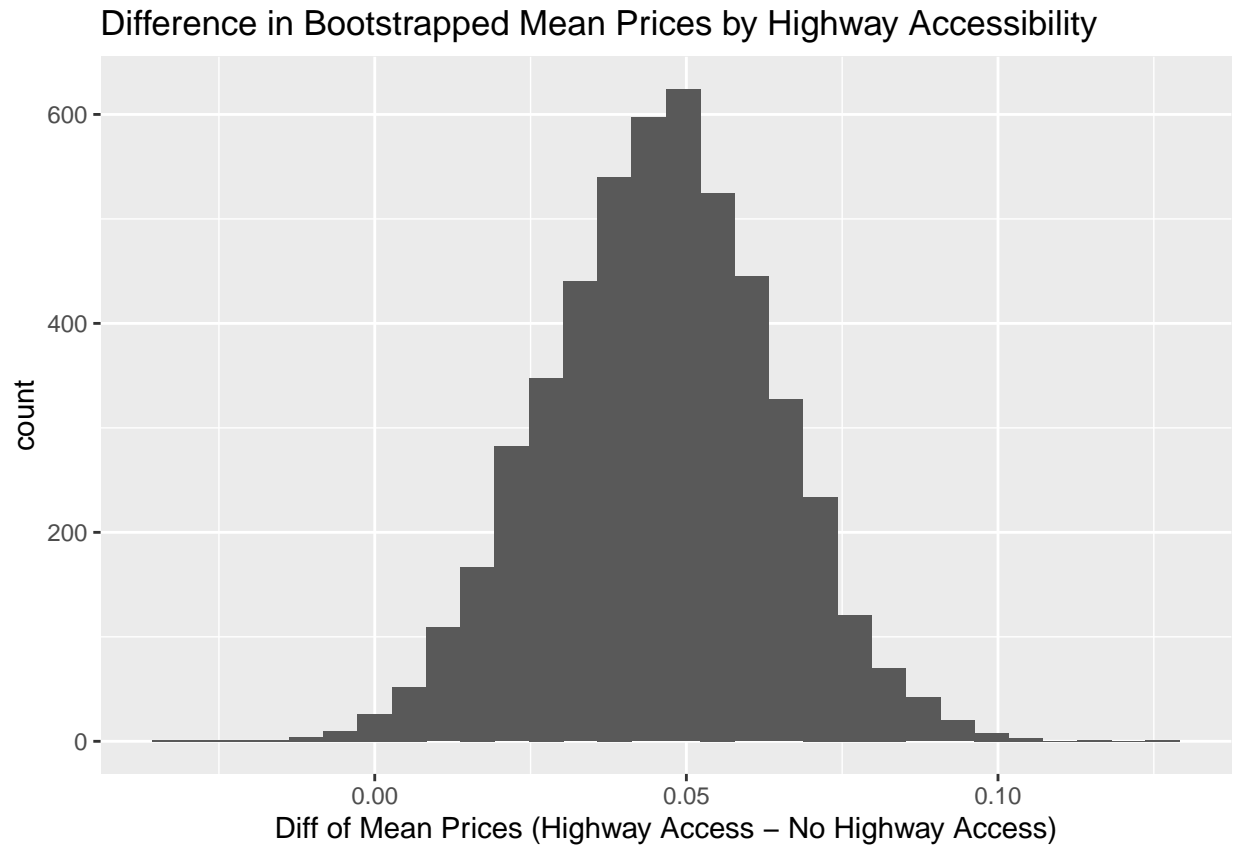### Difference in Bootstrapped Mean Prices by Presence of Stoplight

Conclusion: According to our bootstrapped samples, 95% of the difference in means between gas prices of stations by a stoplight vs. stations not by a stoplight lies within the interval [-0.0380, 0.0298]. Since 0 is within the interval, there is no concrete evidence that gas stations charge more if they are at a stoplight.

1.D)

Claim: Gas stations with direct highway access charge more.

Evidence:

Difference in Bootstrapped Mean Prices by Highway Accessibility

Conclusion: According to our bootstrapped samples, 95% of the difference in means between gas prices of stations with direct access to a highway vs. with no direct access to a highway fall within the interval [0.0099, 0.0811]. Since 0 is not within the interval, evidence suggests that with 95% confidence, gas prices are higher for stations with direct access to highways.

1.E)

Claim: Shell charges more than all other non-Shell brands
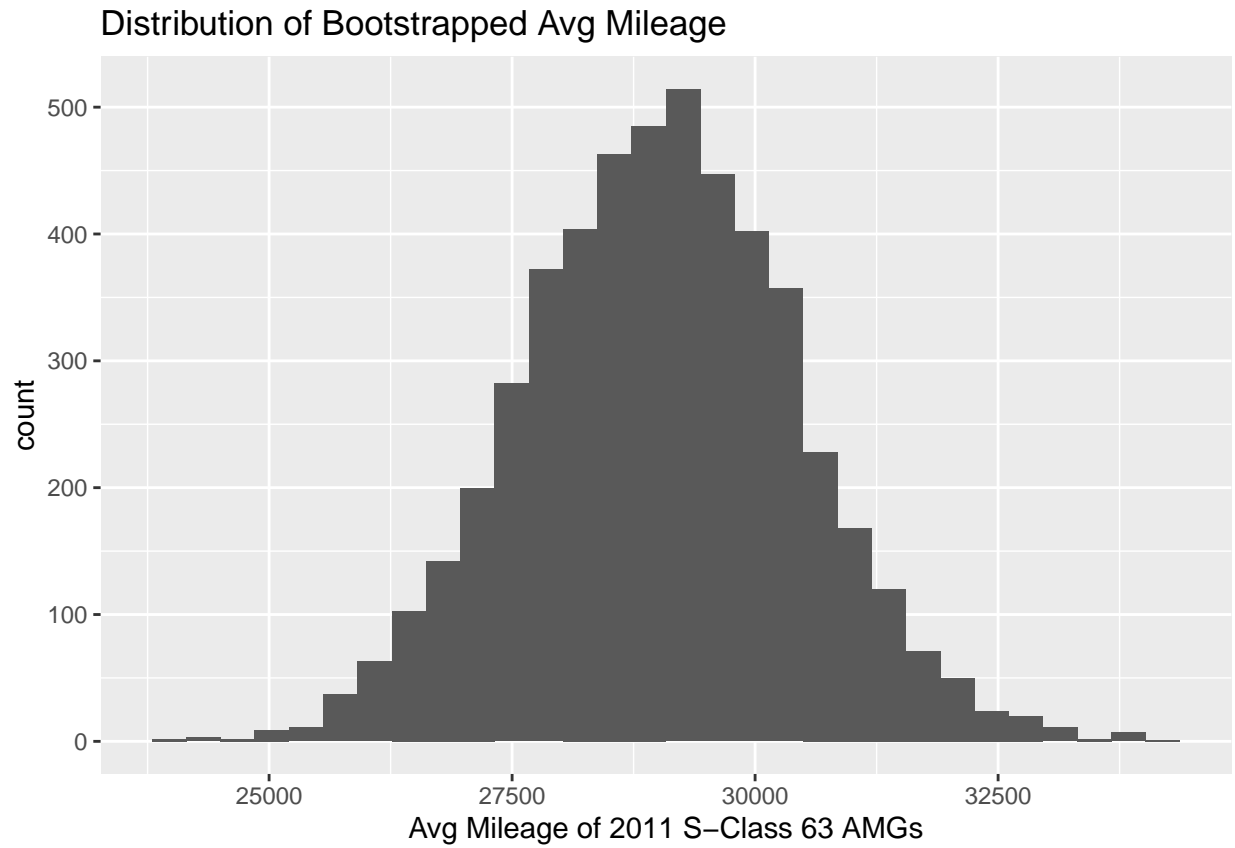
Evidence:

Proportion of Bootstrapped Samples that Charged Highest

Not Shell (99.37%) — Shell (0.63%)
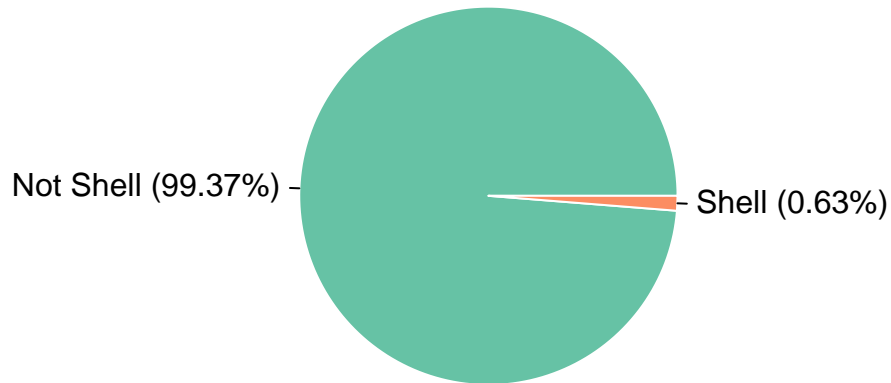
Conclusion: According to our bootstrapped samples, Shell was the gas station that had the highest average prices only around 0.11% of the time. This strongly suggests that Shell does not charge higher than every other brand.
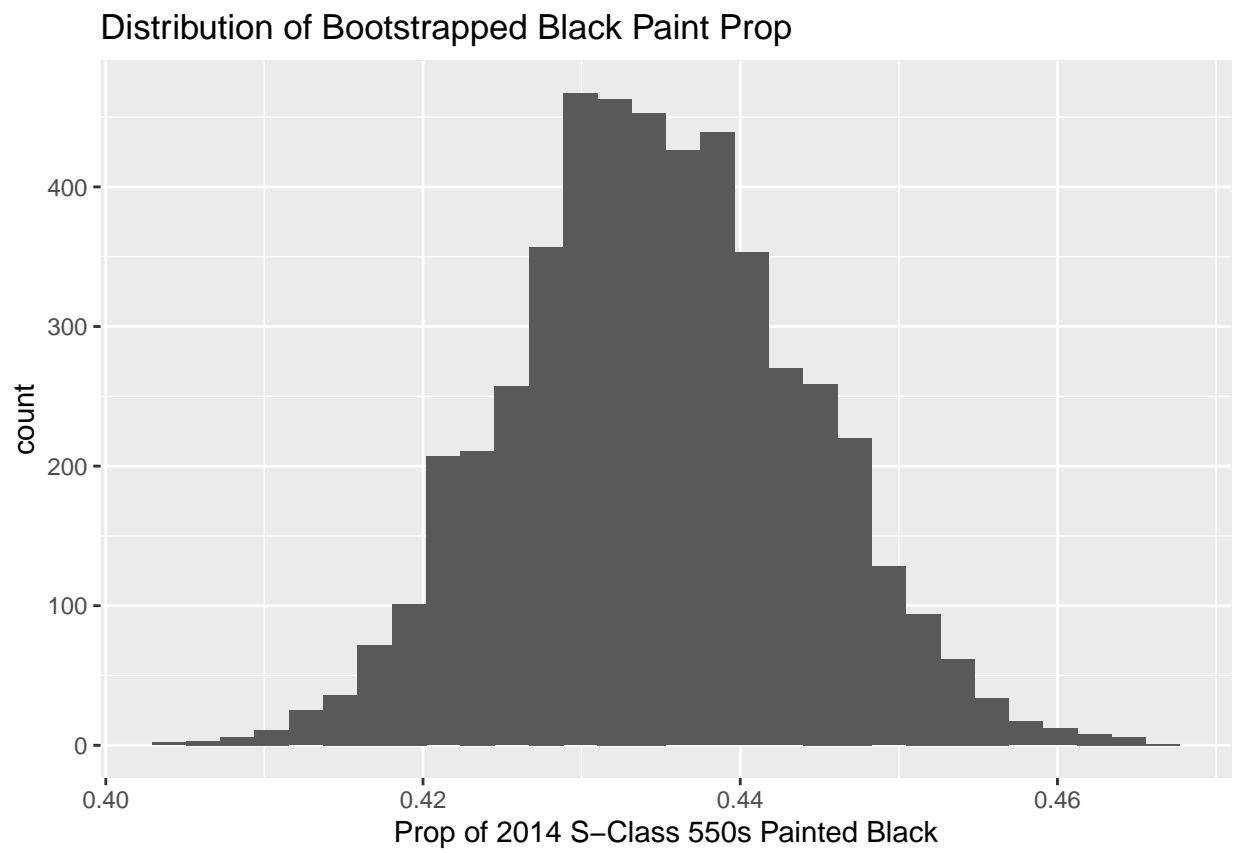
Problem 2:

2.A)

Distribution of Bootstrapped Avg Mileage

Upon evaluating bootstrapped samples for the average mileage of 2011 S-Class 63 AMGs, the true average lies within the interval [26256,31851] with a 95% confidence.

2.B)

## Distribution of Bootstrapped Black Paint Prop

Upon evaluating bootstrapped samples for the proportion of all 2014 S-Class 550s that were painted black, the true proportion lies within the interval [0.417,0.453] with a 95% confidence.

Problem 3:

3.A)

Question: Is there evidence that one show (between "Living with Ed" and "My Name is Earl" consistently produces a higher mean Q1_Happy response among viewers?

Answer:

### Difference in Bootstrapped Mean Happiness Rating (Scale of 1–5)
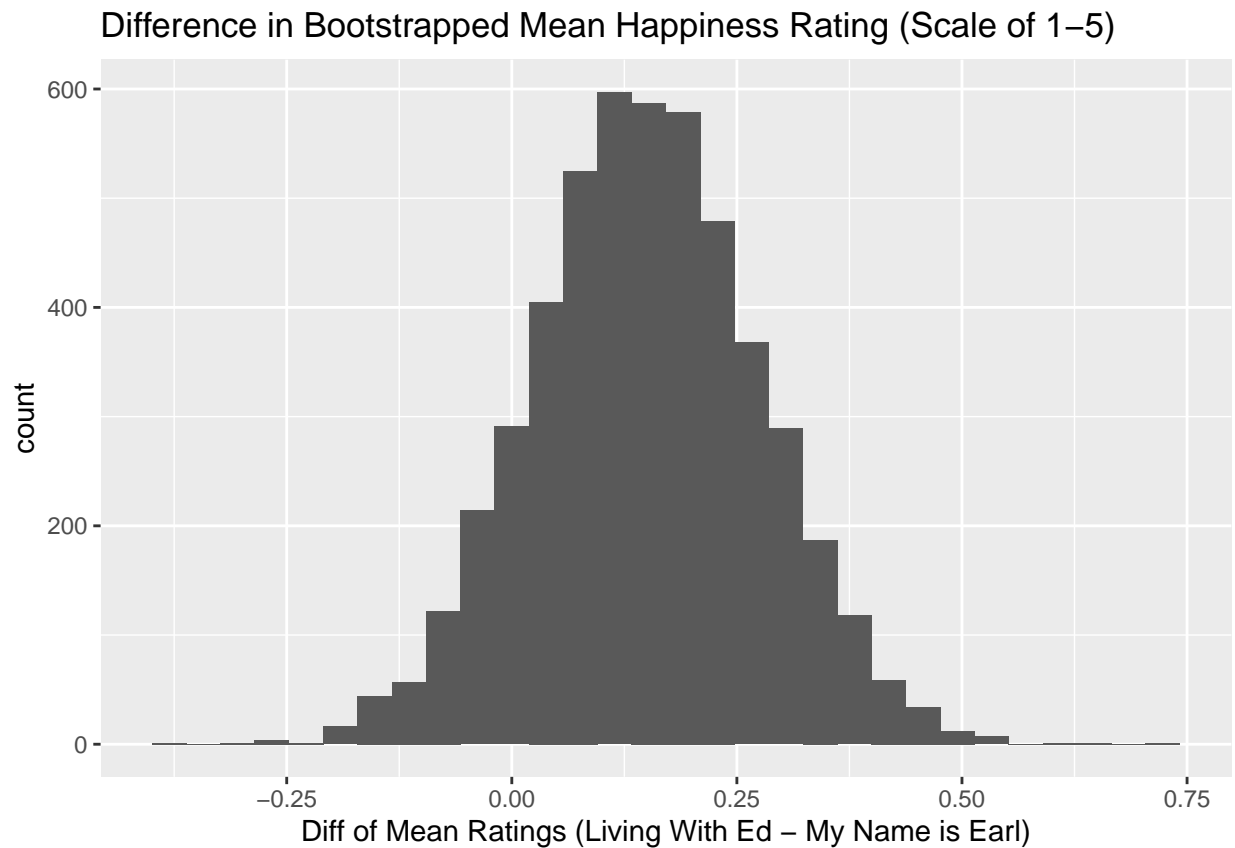
The histogram of bootstrapped (simulated) samples of differences in mean happiness ratings for the shows "Living with Ed" and "My Name is Earl" suggest that the true difference in ratings lie within the interval [-0.093,0.395] with a 95% confidence. Since 0 is within this interval, theres no strong evidence suggesting that one show is consistently rated higher than the other in the happiness metric.

3.B)

Question: Is there evidence that one show between "The Biggest Loser" and "The Apprentice: Los Angeles" consistently produces a higher mean Q1_Annoyed response among viewers?

Answer:

### Difference in Bootstrapped Mean Annoyance Rating (Scale of 1–5)
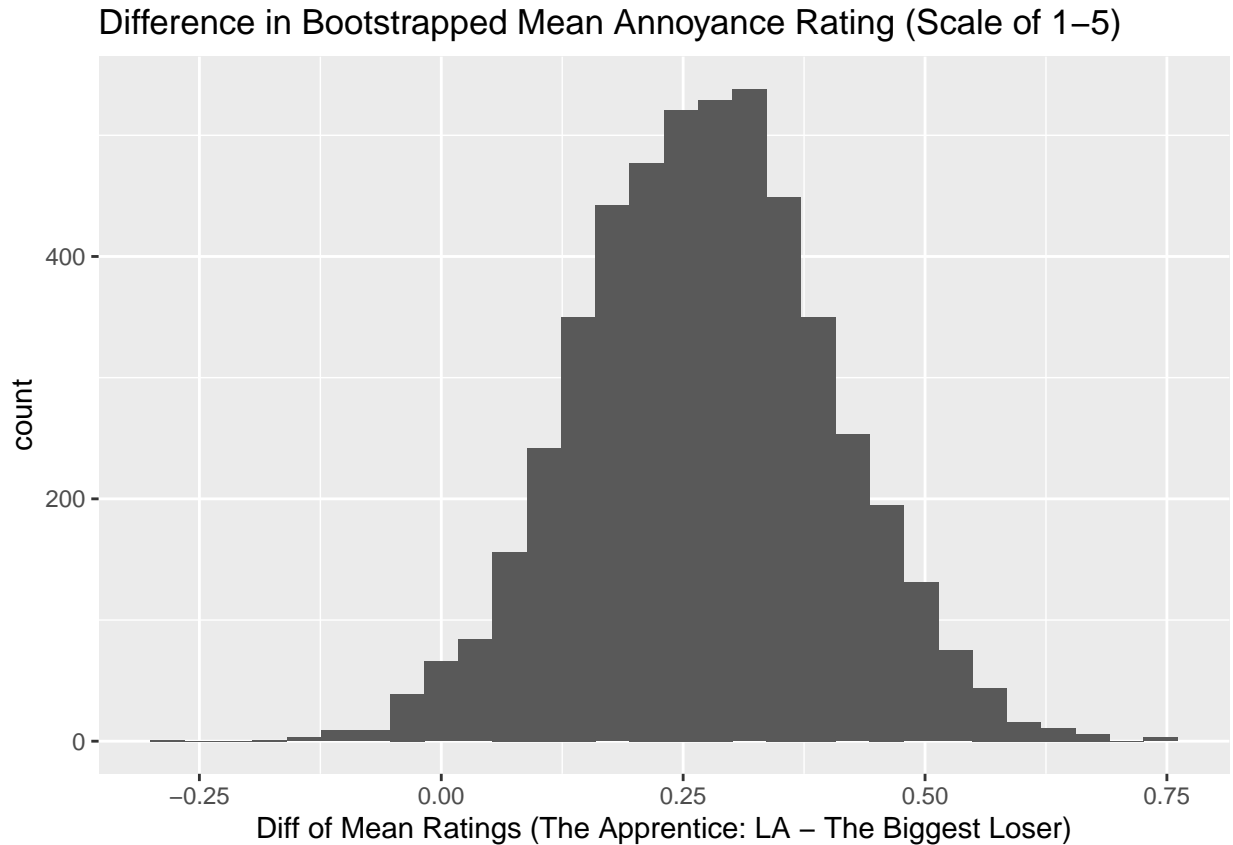
The histogram of bootstrapped samples of differences in mean annoyance ratings for the shows "The Apprentice: LA" and "The Biggest Loser" suggest that the true difference in ratings lie within the interval [0.016,0.525] with a 95% confidence. Since 0 is not within this interval, evidence suggests that "The Apprentice" is consistently rated higher than The Biggest Loser" in the annoyance metric.

3.C)

Question: For the show, "Dancing with the Stars", what proportion of American TV watchers would we expect to give a response of 4 or greater to the "Q2_Confusing" question?

Answer:

Bootstrapped Mean Proportions of High Confusion Ratings for "Dancing With the Stars"
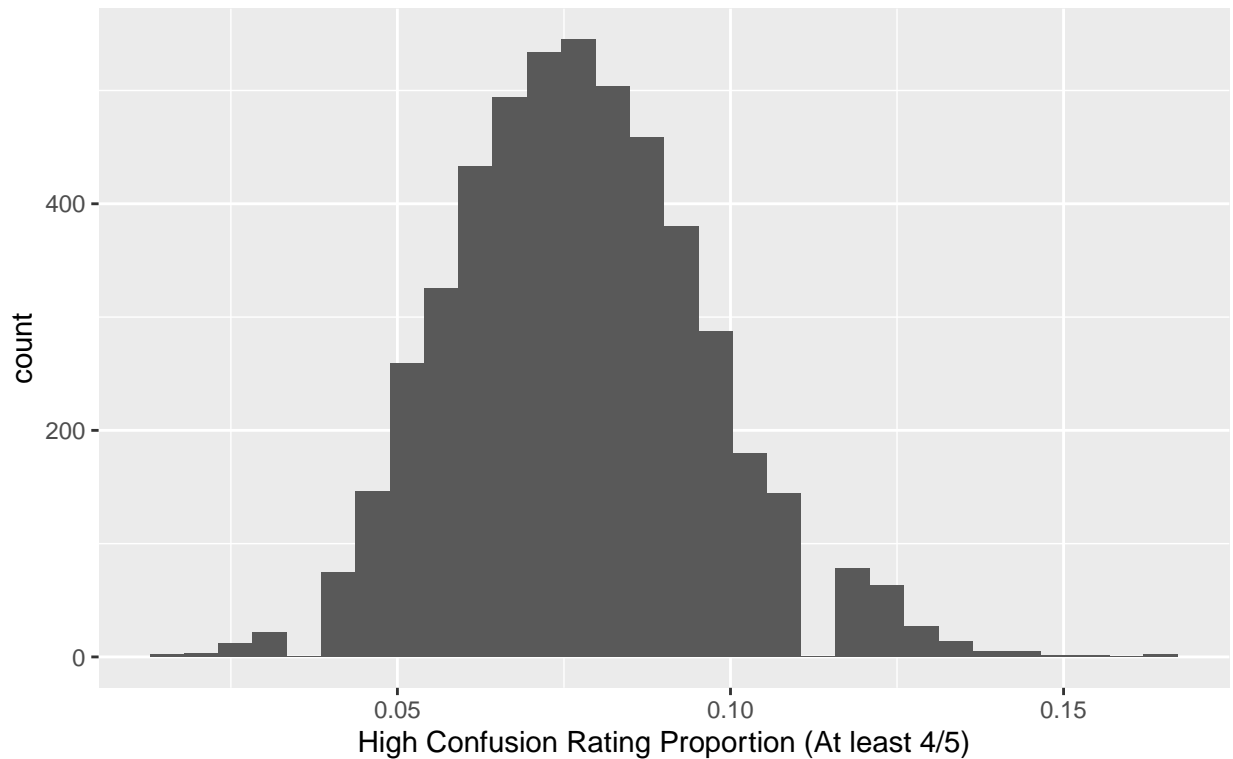
The histogram of bootstrapped samples of proportions of mean high ratings for "Dancing With the Stars" suggest that the true proportion of mean high ratings for the confusion metric lies within the interval [0.044,0.116] with a 95% confidence. Accordingly, on average, somewhere from 4.4% to 11.6% of viewers will find the show to be confusing.

Problem 4:

4.A)

Question: Is the revenue ratio is the same for the treatment and control groups of Ebay's experiment?

Approach: The following histogram displays the distribution of differences of mean revenue ratios for the control vs. treatment group before and after the experiment occurred, according to bootstrapped (simulated) samples. If the difference is consistently positive, then evidence suggests that Ebay's paid advertising initiative does generate extra revenue.

Difference in Bootstrapped Mean Revenue Ratios
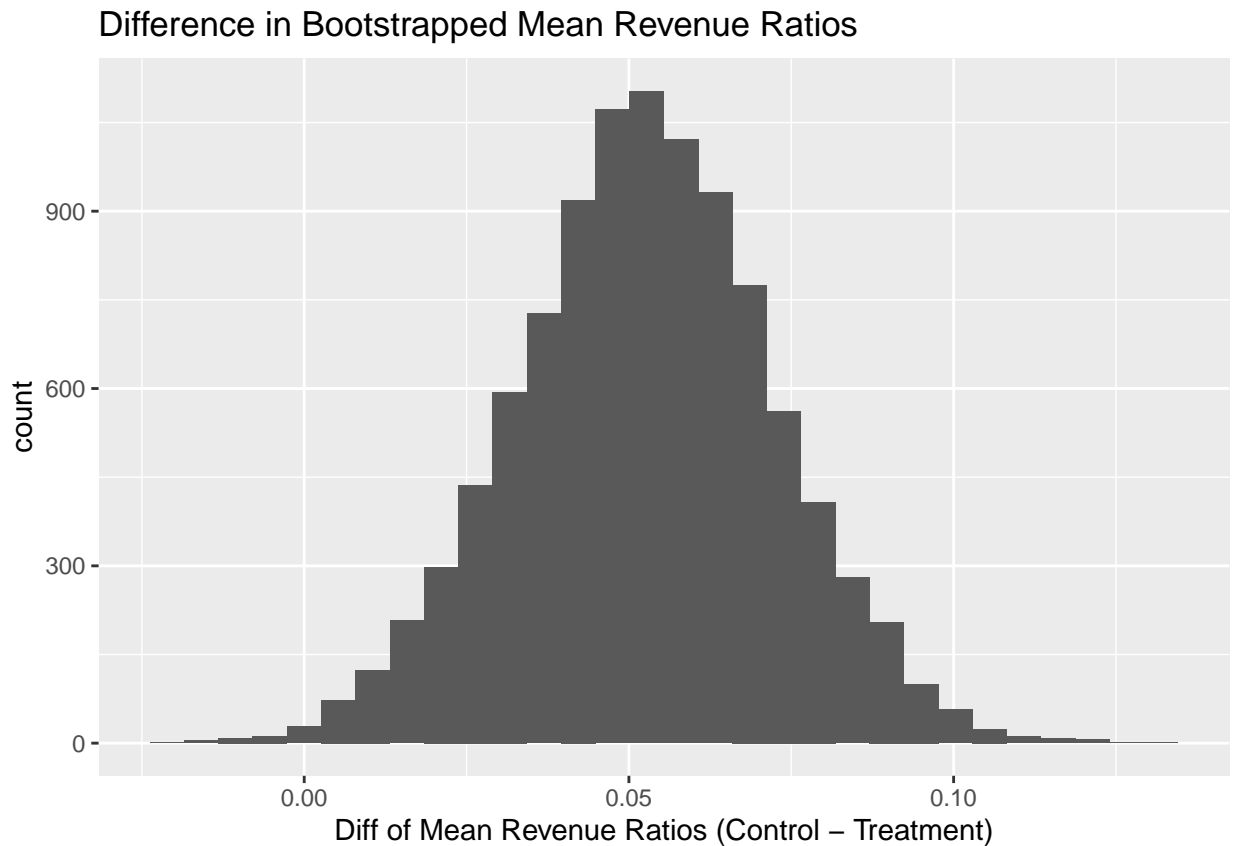
Results/Conclusion: According to the histogram of bootstrapped samples comparing the control vs. treatment group, the true difference of mean revenue ratios lie within the interval [0.013,0.090], with a 95% confidence. Since the lower end of the interval is greater than 0, this suggests that Ebay's control group generally generated more revenue proportionately than the treatment group where their advising was stopped, thus it's advertising initiative on Google is successful.