Homework 4

Avik Malladi

2024-02-19

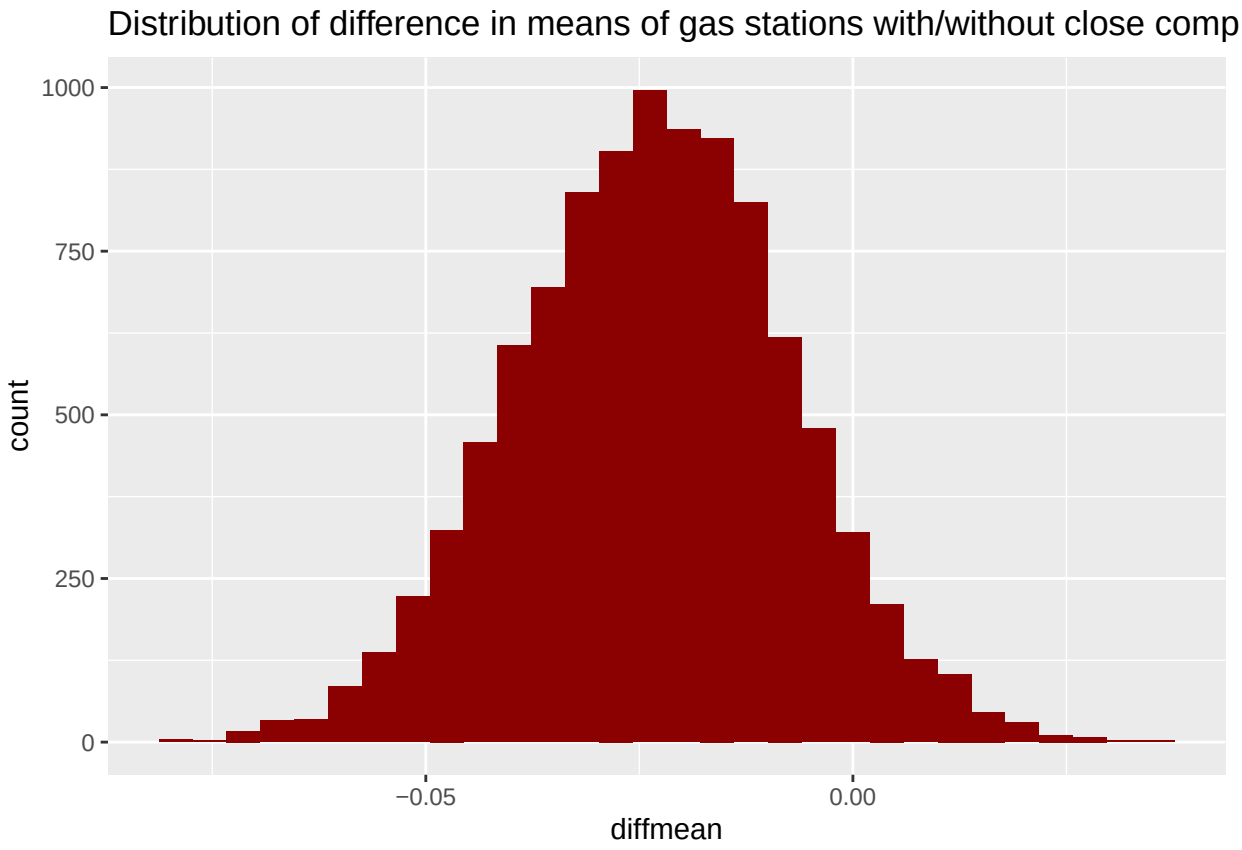
Name: Avik Malladi EID: arm7542 Github: https://github.com/avikMall/sds-315-homework-4

Problem 1

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

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

Evidence:



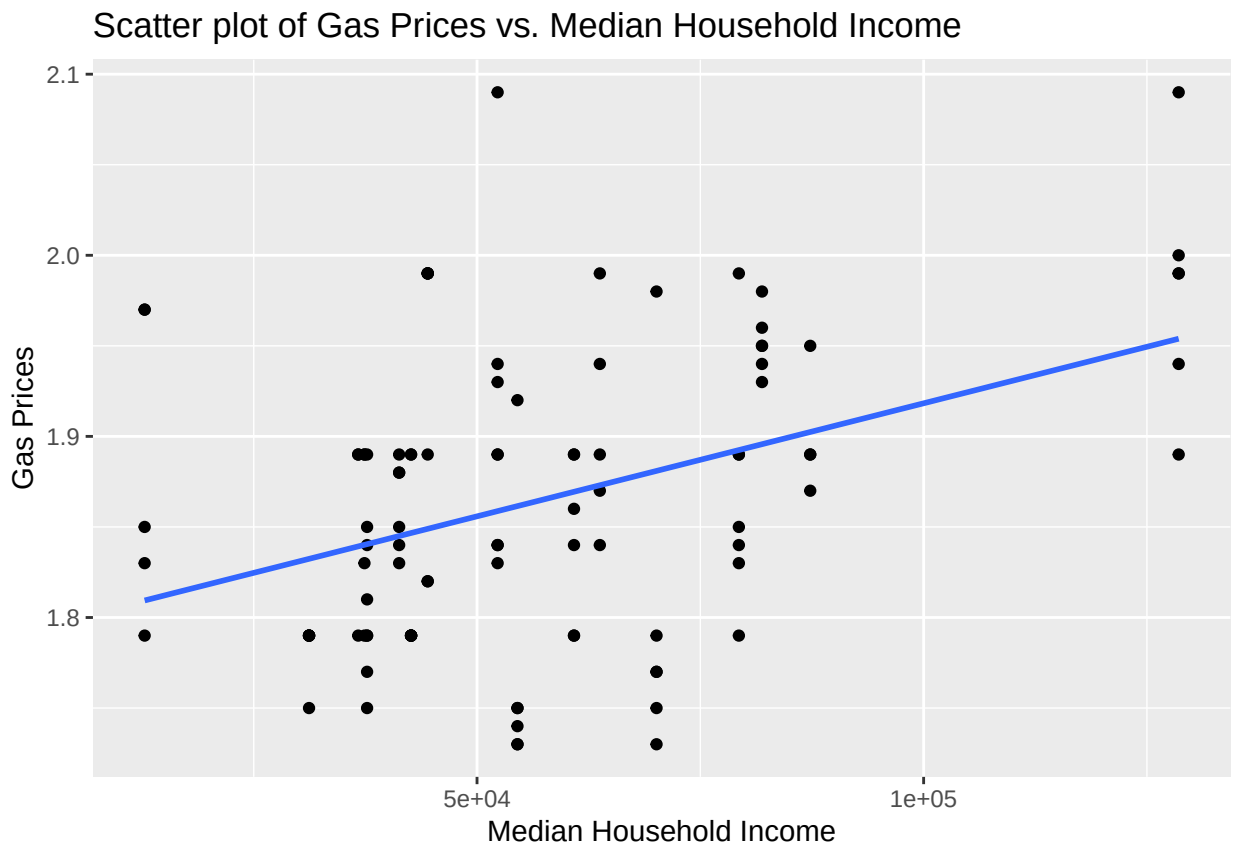
The difference between gas prices in gas stations with direct competition in sight vs. without direct competition in sight is somewhere in between -0.055 and 0.008 with 95% confidence.

Conclusion: According to the results calculated, this theory is not supported by the data. In fact, it is shown that gas stations with competition charge ~1.5 cents less than gas stations without competition on average.

Theory B: The richer the area, the higher the gas prices.

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

Evidence:



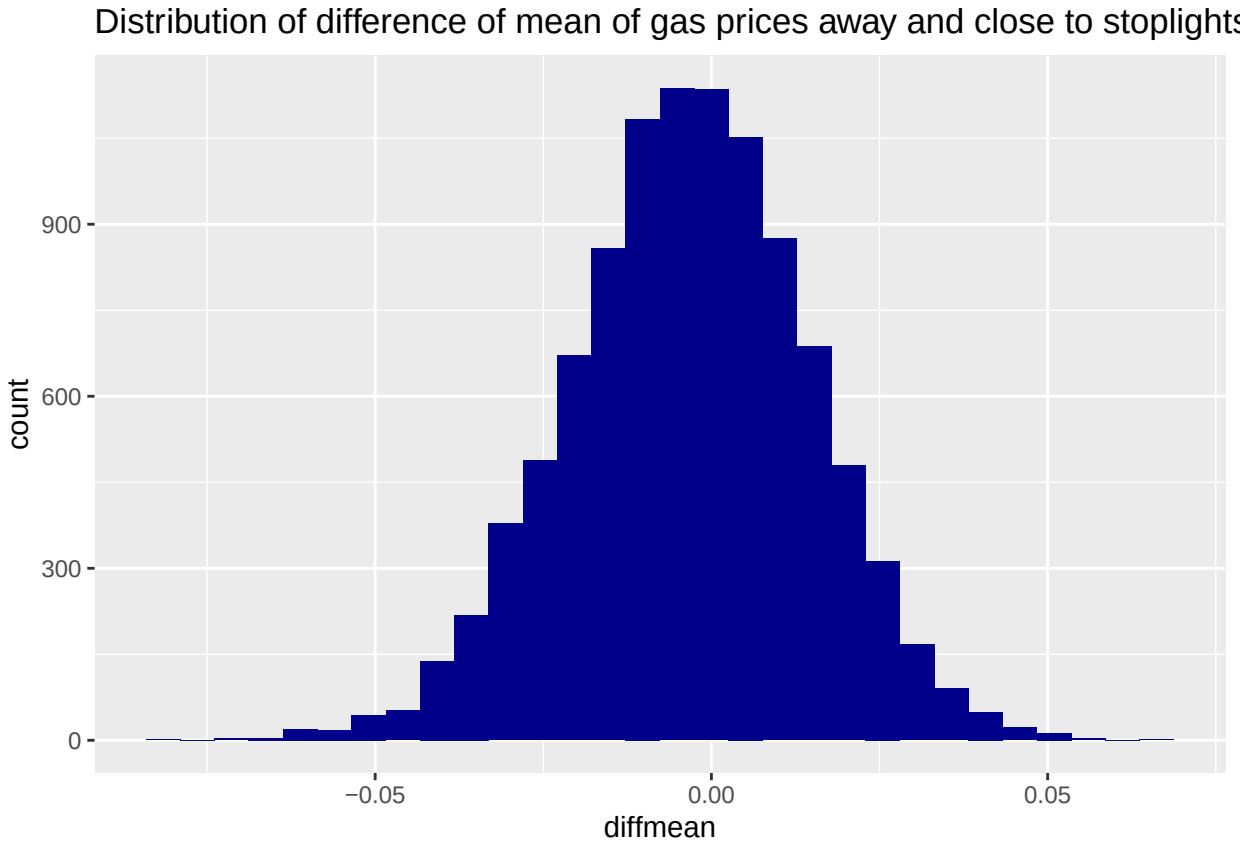
The R square value of gas prices vs. the surrounding median household income is somewhere in between 0.04 and 0.325 with 95% confidence.

Conclusion: This claim is supported by the data as the linear regression line has a positive slope (gas prices increase as house income increases)

Theory C: Gas stations at stoplights charge more.

Claim: Gas stations at stoplights charge more.

Evidence:



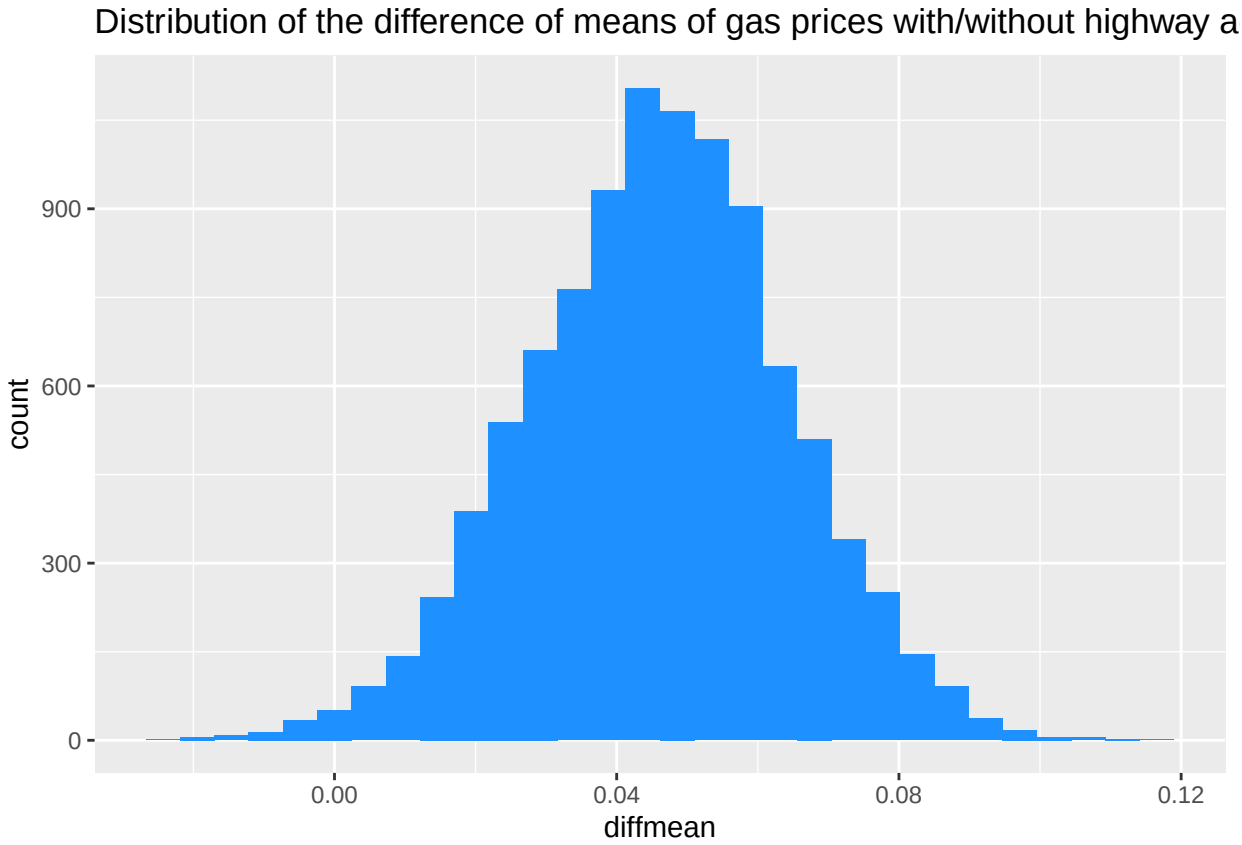
The difference in means of gas prices at gas stations near stoplights and gas stations away from stoplights is somewhere in between -0.039 and 0.031 with 95% confidence.

Conclusion: This claim is not supported by the data as gas prices do not to seem to change at gas stations close to stoplights and away from stop lights.

Theory D: Gas stations with direct highway access charge more.

Claim: Gas stations with direct highway access charge more.

Evidence:

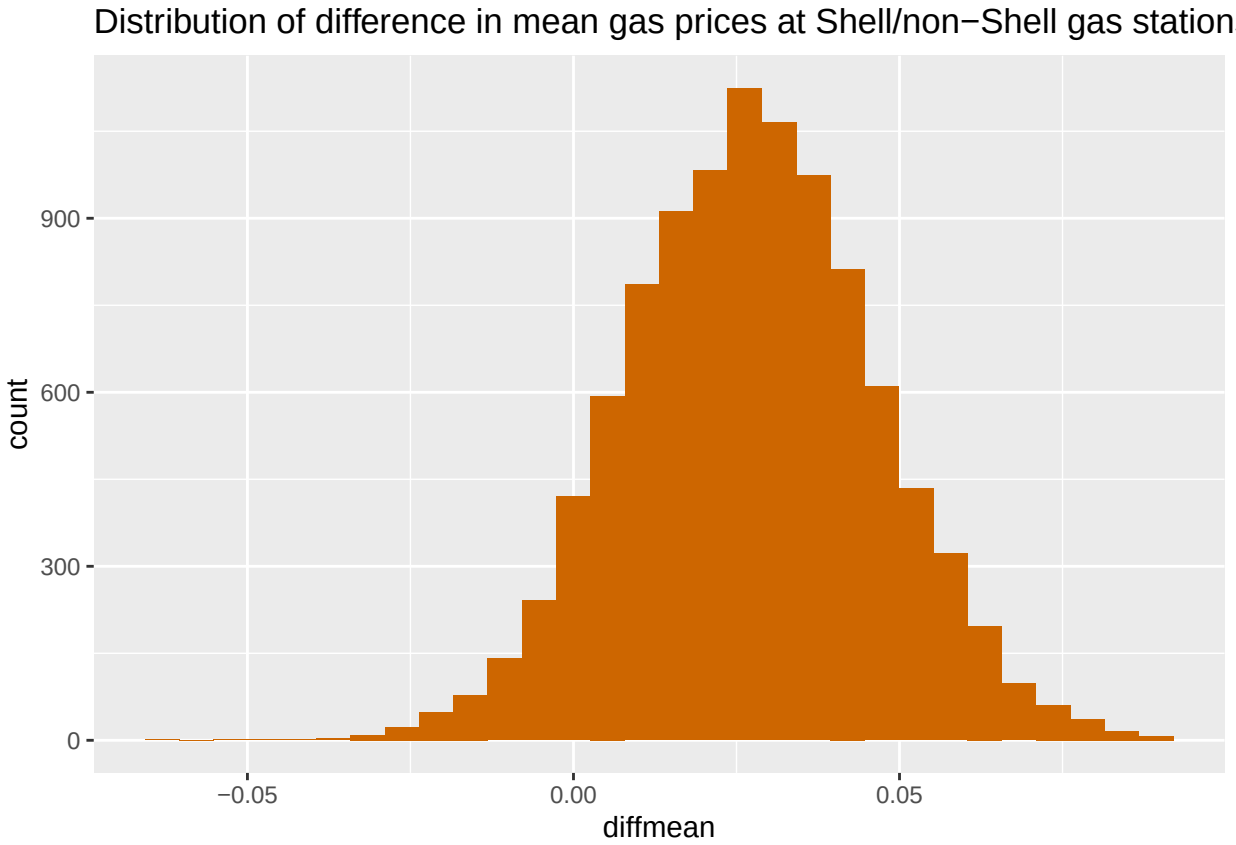


The difference in means of gas prices between gas stations with direct highway access and without direct highway access is somewhere in between 0.009 and 0.082 with 95% confidence.

Conclusion: This claim is supported by the data as gas prices at gas stations with highway access are about 5 cents higher than gas stations without on average.

Theory E: Shell charges more than all other non-Shell brands.

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



Evidence:

The difference in means of gas prices at Shell gas stations and non-Shell gas stations is somewhere in between -0.01 and 0.065 with 95% confidence.

Conclusion: The claim is supported by the data as Shell gas stations charge almost 3 cents more than non-Shell gas stations on average.

Problem 2

Part A

The average mileage for 2011 S-Class 63 AMGs that were hitting the used-car market when this data was collected is somewhere in between 2.6271138×10^4 and 3.1807724×10^4 with 95% confidence.

Part B

The proportion of all 2014 S-Class 550s that were painted black when this data was collected is somewhere in between 0.417 and 0.452 with 95% confidence.

Problem 3

Part A

Question: Who makes people happier: Ed or Earl?

Approach: For this question we filter the data to only contain the 2 shows of interest and then use the diffmean() and confint() functions to calculate the confidence intervals.

Results:

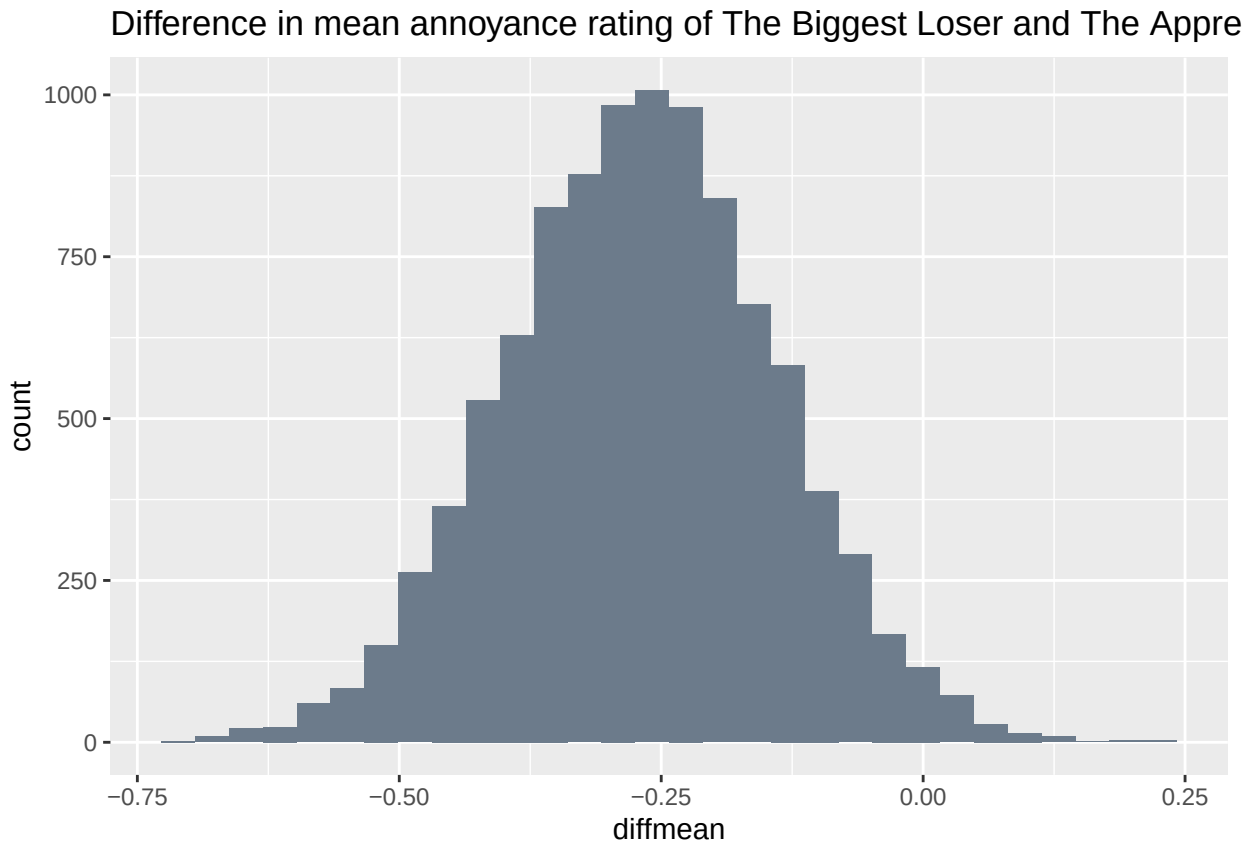
The difference in mean happiness rating between "Living with Ed" and "My Name is Earl" is somewhere in between -0.394 and 0.105 with 95% confidence.

Conclusion: Based on the data analyzed, the show "Living with Ed" makes people happier than the show "My Name is Earl." We can see that "Living with Ed" got almost 0.2 points higher happiness rating by viewers on average.

Part B

Question: Which reality/contest show made people feel more annoyed?

Approach: For this question we filter the data to only contain the 2 shows of interest and then use the diffmean() and confint() functions to calculate the confidence intervals.



Results:

The difference in mean annoyance rating between "The Biggest Loser" and "The Apprentice: Los Angeles" is somewhere in between -0.52 and -0.017 with 95% confidence.

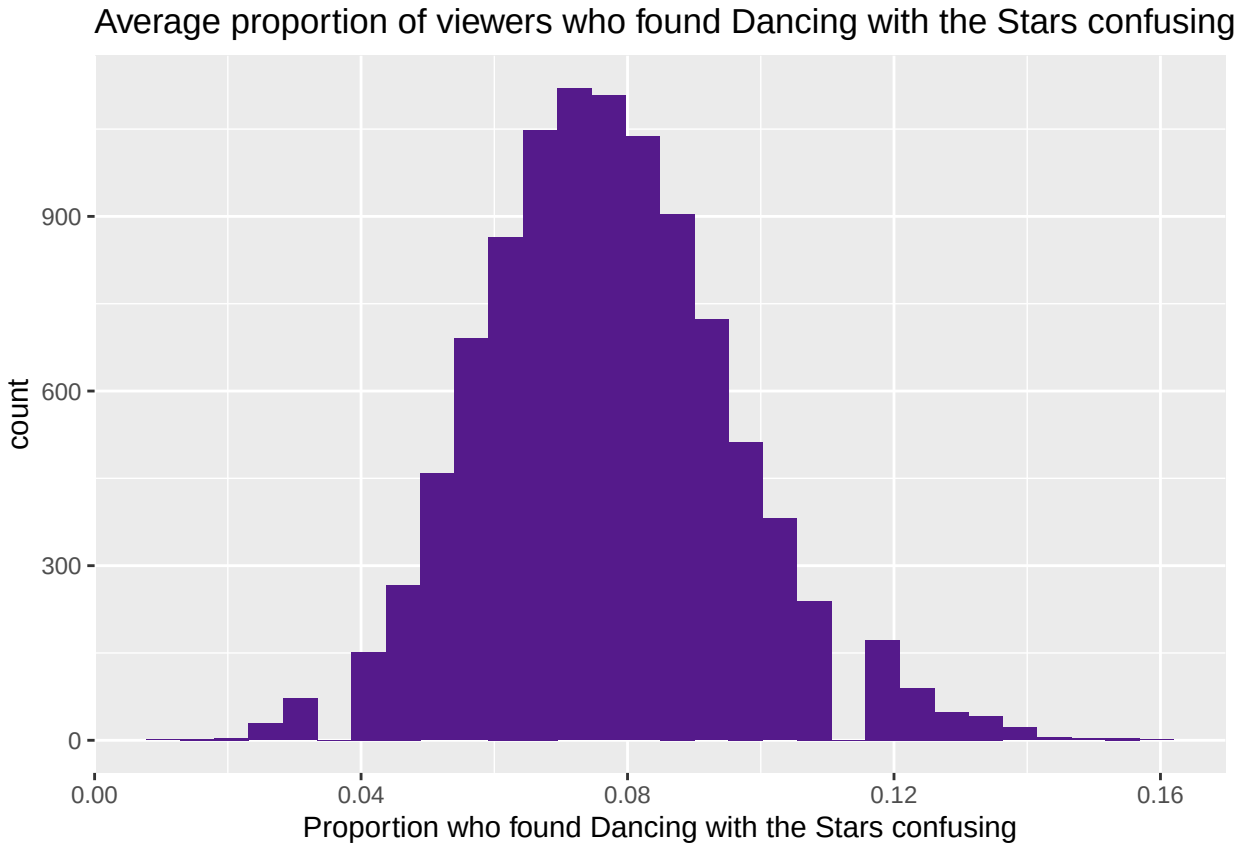
Conclusion: According to the data, "The Apprentice: Los Angeles" made people more annoyed. "The Apprentice: Los Angeles" recieved a 0.3 point higher annoyance rating from viewers on average.

Part C

Question: Is the Show "Dancing With the Stars" confusing to viewers?

Approach: For this question we filter the data to the show of interest and add another boolean variable (whether the viewer thought the show was confusing or not). Then we use the prop() and confint() functions to calculate the confidence intervals for the proportion of people who found the show confusing.

Results:



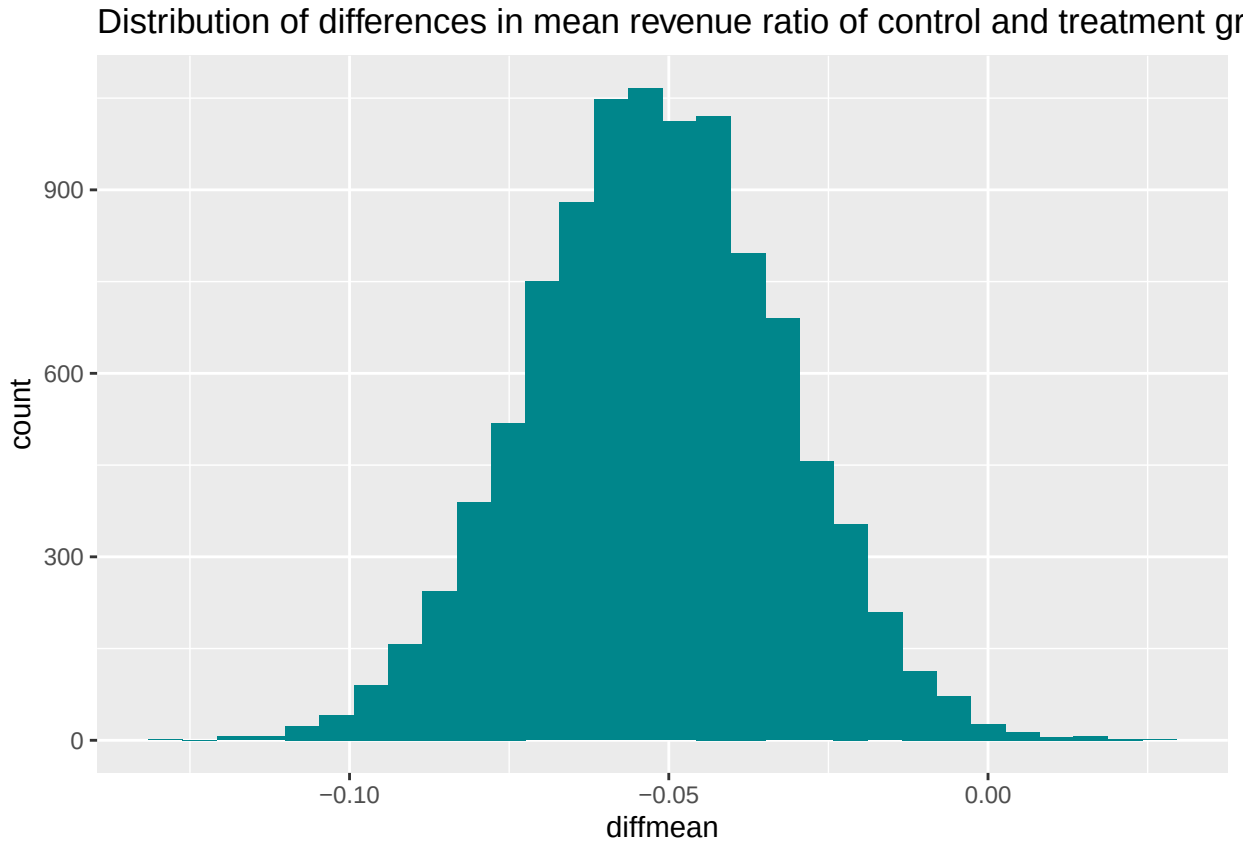
The proportion of viewers who found Dancing with the Stars confusing is somewhere in between 0.039 and 0.116 with 95% confidence.

Conclusion: According to the data, the show "Dancing with the stars" is not very confusing to viewers. The data shows that only around 7% of users found the show confusing on average.

Problem 4

Question: Does Ebay's paid advertising on Google's search platform improve EBay's revenue?

Approach: For this question we create a new variable tracking the ratio of revenue before and after the start of the expirement. We then use the diffmean() and confint() function to calculate confidence intervals for the revenue ratios.



Results:

The difference in mean revenue ratio between the control and treatment groups is somewhere in between -0.091 and -0.014 with 95% confidence.

Conclusion: According to the data, Ebay's paid advertising on Google's search platform does improve Ebay's revenue. While both the control and treatment groups saw a decrease in revenue on average, the treatment groups fall was more drastic was a difference of about 5% on average.