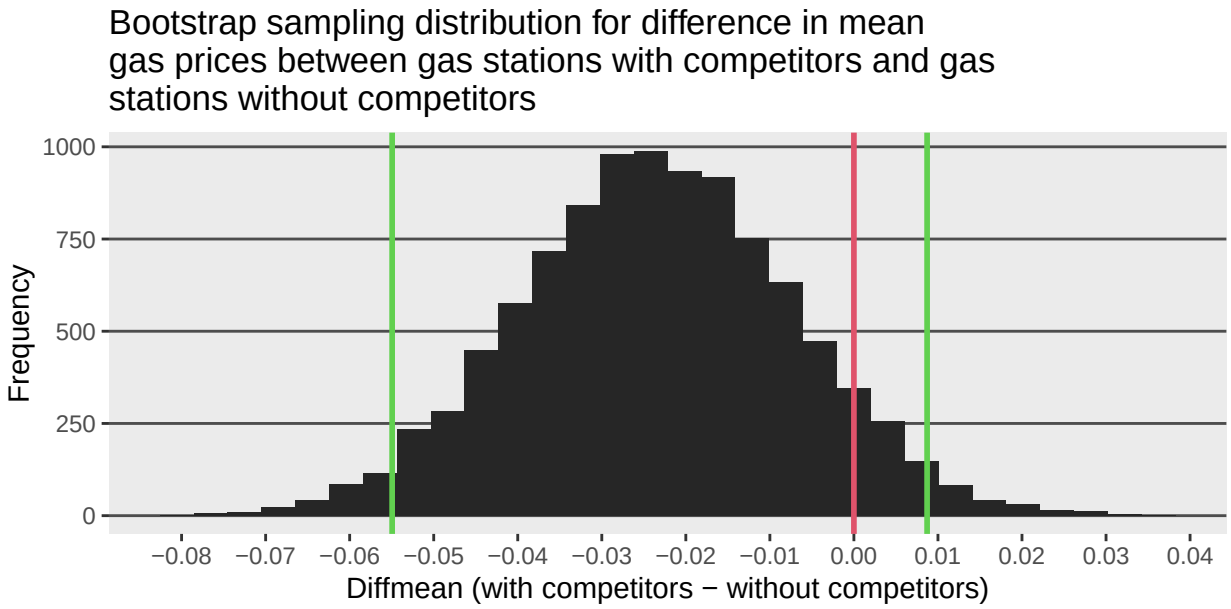
Hw 3

Jaxon Lara

2/10/2025

(Github_R_Code)

1(A)

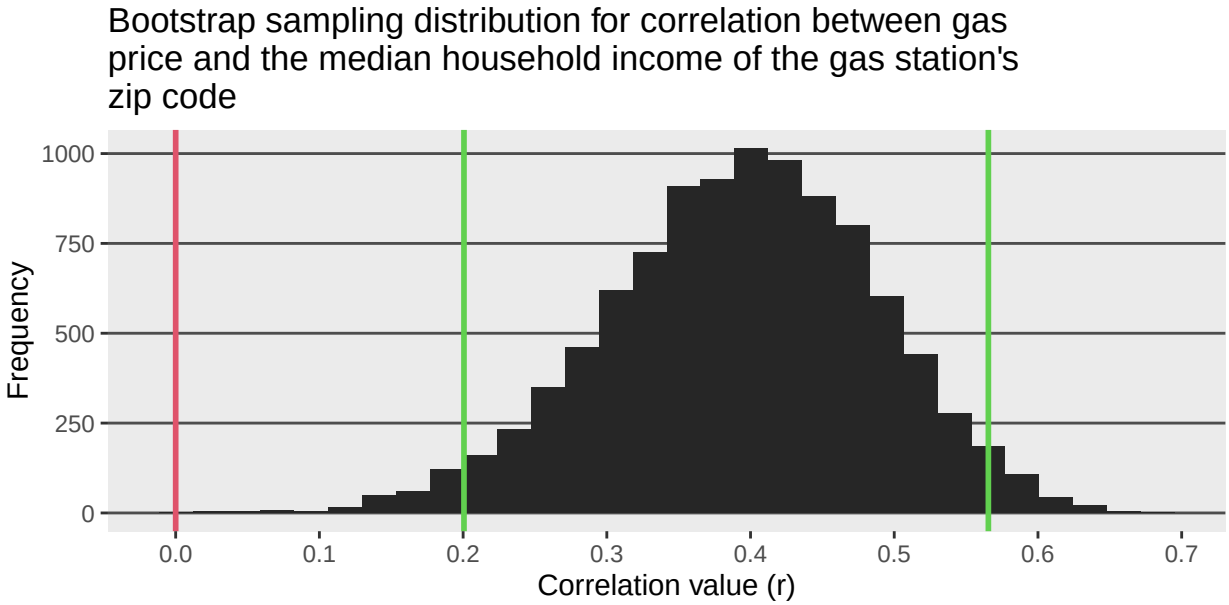


Claim: gas stations charge more if they lack direct competition in sight.

Evidence: Based on the bootstrap sampling distribution I am 95% confident that the difference in mean gas prices between gas stations with competitors and gas stations without competitors is between −0.055 and 0.0087. The theory is unsupported by the evidence because 0 is within the confidence interval which describes that the confidence interval is not statistically significant.

Conclusion: The theory is unsupported, not enough evidence to suggest gas stations charge more if they lack direct competition in sight.

1(B)

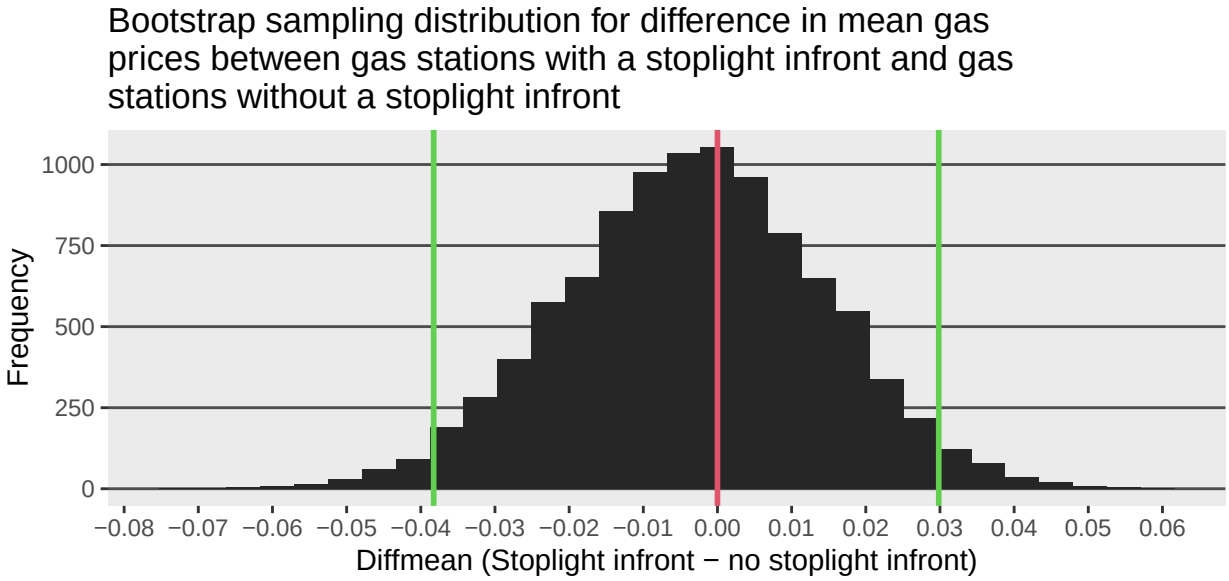


Claim: The richer the area, the higher the gas prices.

Evidence: Based on the bootstrap sampling distribution I am 95% confident that the correlation value between gas price and the median household income of the gas station's zip code is between 0.2 and 0.57. The claim is supported by the evidence because 0 is not in the confidence interval which describes that the confidence interval is statistically significant.

Conclusion: The theory is supported, there is enough evidence to suggest the richer the area the higher the gas prices.

1(C)

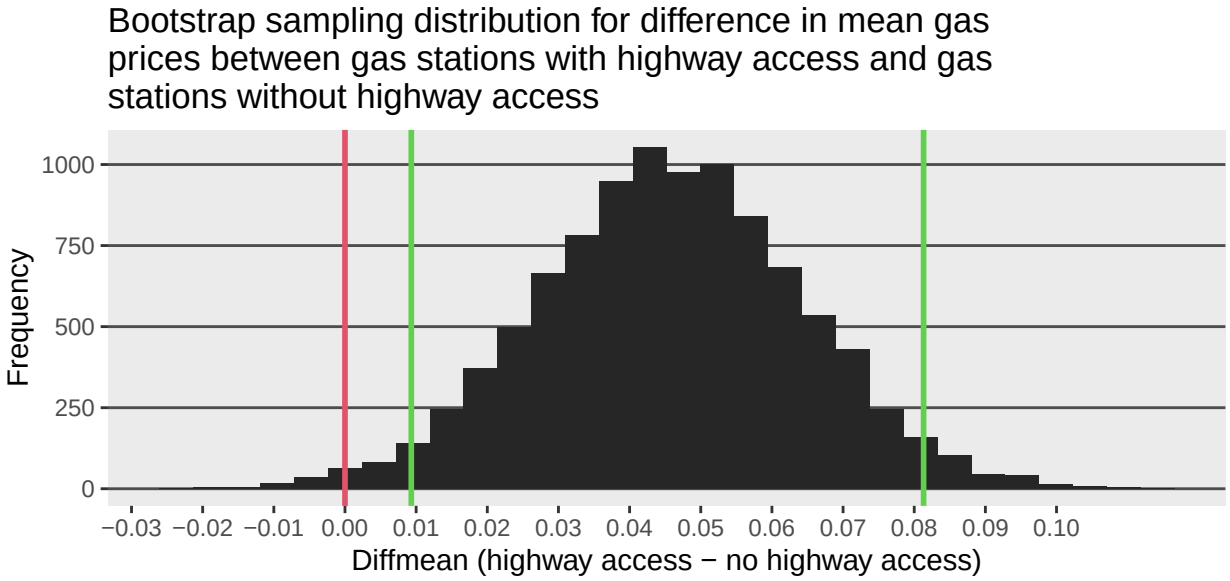


Claim: Gas stations at stoplights charge more.

Evidence: Based on the bootstrap sampling distribution I am 95% confident that the difference in mean gas prices between gas stations with a stoplight infront and gas stations without a stoplight infront is between −0.038 and 0.030. The claim is not supported by the evidence because 0 is within the confidence interval which describes that the confidence interval is not statistically significant.

Conclusion: The theory is not supported, there is not enough evidence to suggest gas stations at stoplights charge more.

1(D)

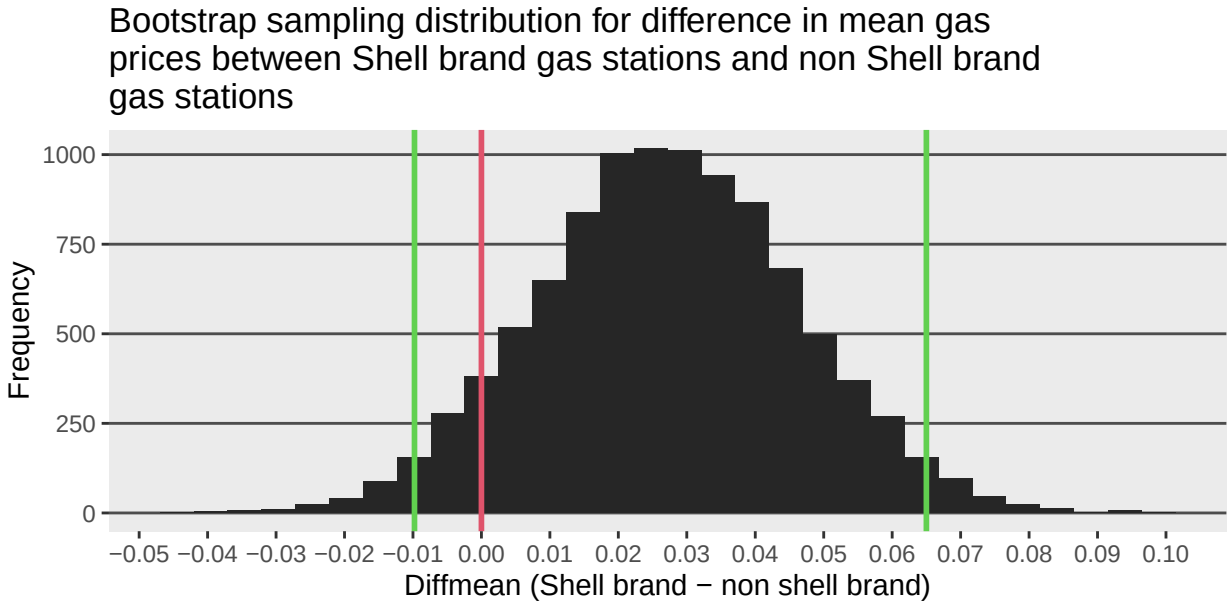


Claim: Gas stations with direct highway access charge more.

Evidence: Based on the bootstrap sampling distribution I am 95% confident that the difference in mean gas prices between gas stations with highway access and gas stations without highway access is between 0.0093 and 0.081. The claim is supported by the evidence because 0 is not within the confidence interval which describes that the confidence interval is statistically significant.

Conclusion: The theory is supported, there is enough evidence to suggest gas stations with direct highway access charge more.

1(E)

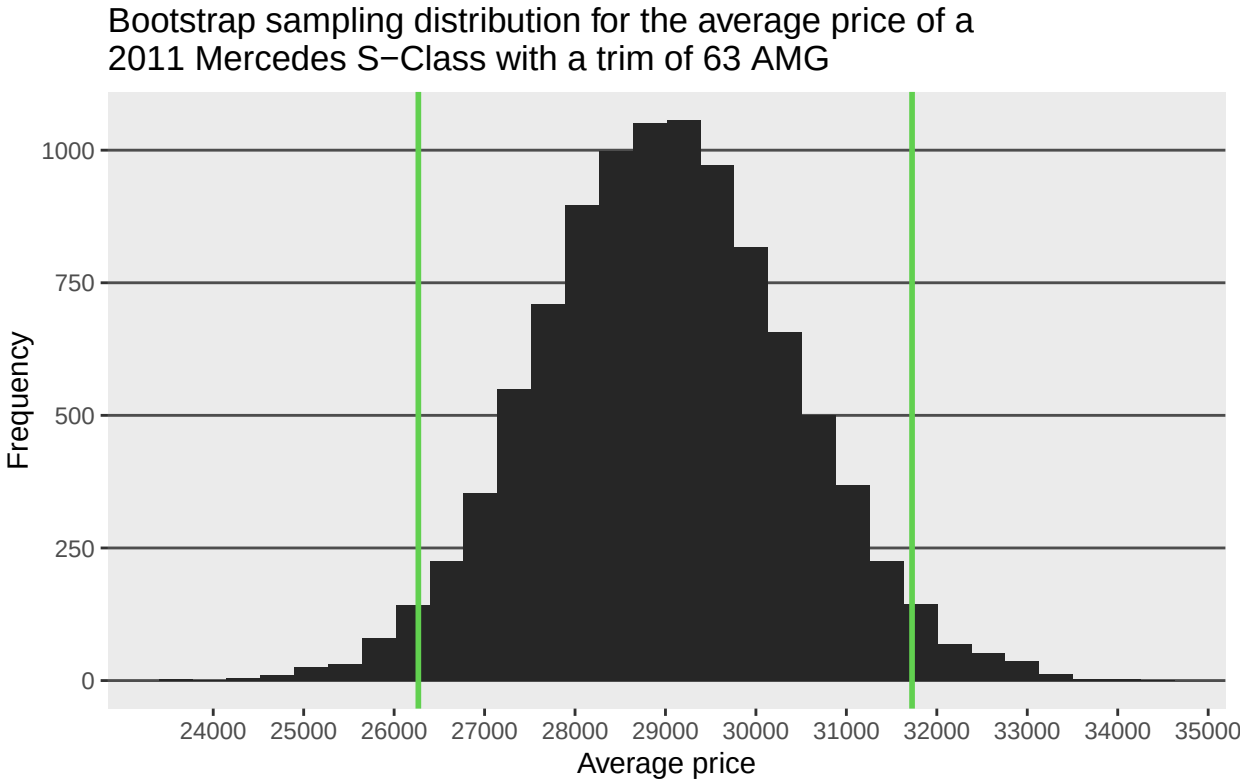


Claim: Shell charges more than all other non−shell brands.

Evidence: Based on the bootstrap sampling distribution I am 95% confident that the difference in mean gas prices between shell brand gas stations and non shell brand gas stations is between −0.0098 and 0.065. The claim is not supported by the evidence because 0 is within the confidence interval which describes that the confidence interval is not statistically significant.

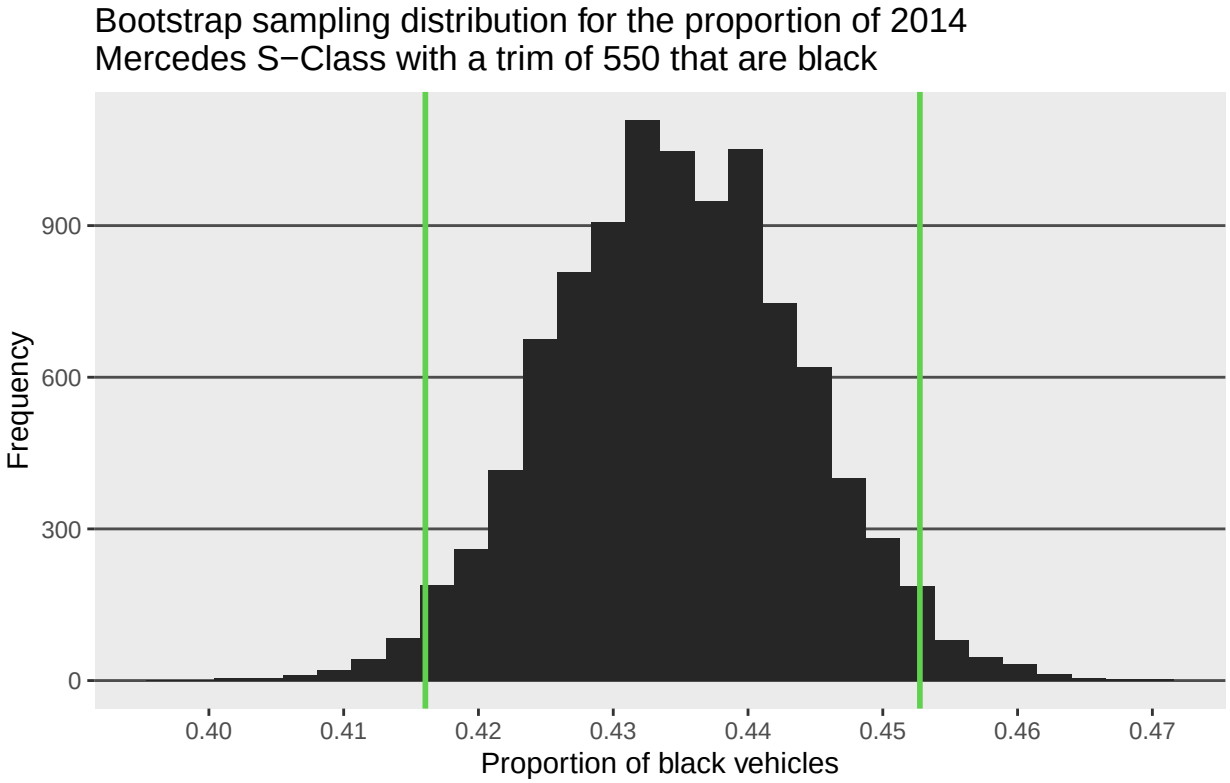
Conclusion: the theory is not supported, there is not enough evidence to suggest the Shell brand charges more than all other non−Shell brands.

2(A)



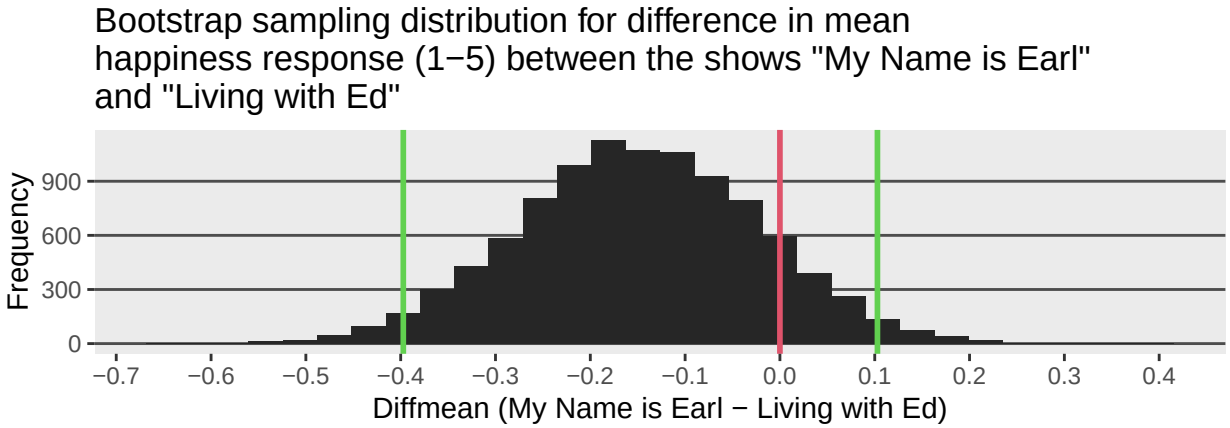
I am 95% confident that the average price of a 2011 Mercedes S−Class with a trim of 63 AMG is between $26,267.88 and $31,725.97.

2(B)



I am 95% confident that the proportion of Mercedes S−Class with a trim of 550 that are black is between 0.42 and 0.45.

3(A)



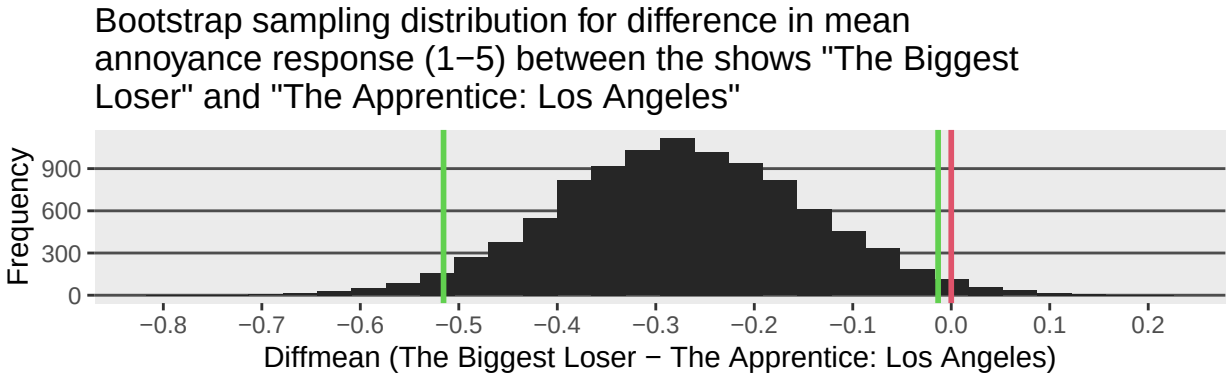
Question: Is there evidence that one show consistently produces a higher mean Q1_Happy response among viewers?

Approach: I first filtered the nbc data set to include only Earl and Ed. I then bootstrapped a sampling distribution of the difference in mean happiness responses between the Earl and Ed shows. After that I constructed a histogram and confidence interval to visualize and quantify the results.

Results: Based on my confidence interval I am 95% confident that the difference in mean happiness response between the shows "My Name is Earl" and "Living with Ed" is between −0.40 and 0.10.

Conclusion: There is not enough evidence to suggest that either Earl or Ed consistently produced a higher mean Q1_Happy response. This is because 0 is within the confidence interval which describes that the confidence interval is not statistically significant and suggests there is not enough evidence that either Earl or Ed makes people happier.

3(B)



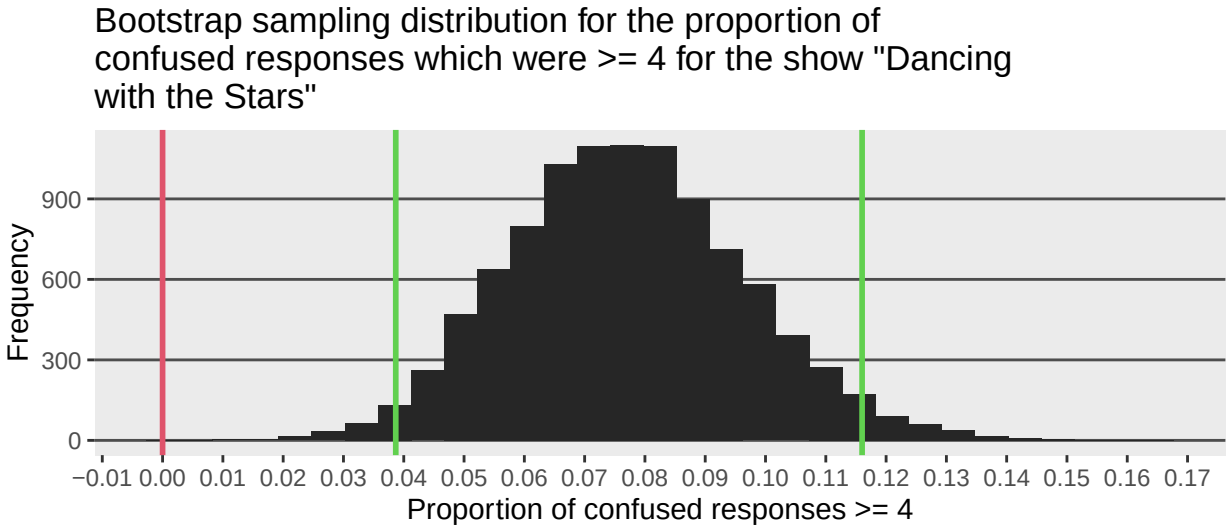
Question: Is there evidence that one show consistently produces a higher mean Q1_Annoyed response among viewers?

Approach: I first filtered the nbc data set to include only "the Biggest Loser" and "The Apprentice: Los Angeles". I then bootstrapped a sampling distribution of the difference in mean annoyance response between the shows Loser and Apprentice. After that I constructed a histogram and confidence interval to visualize and quantify the results.

Results: Based on my confidence interval I am 95% confident that the difference in mean annoyance response between the shows Loser and Apprentice is between −0.52 and 0.013.

Conclusion: There is not enough evidence to suggest that either the Show "The Biggest Loser" or "The Aprentice: Los Angeles" consistently produces a higher mean Q1_Annoyed response among viewers. This is because 0 is within the confidence interval which describes that the confidence interval is not statistically significant and suggests there is not enough evidence that either Loser or Apprentice makes people more annoyed.

3(C)



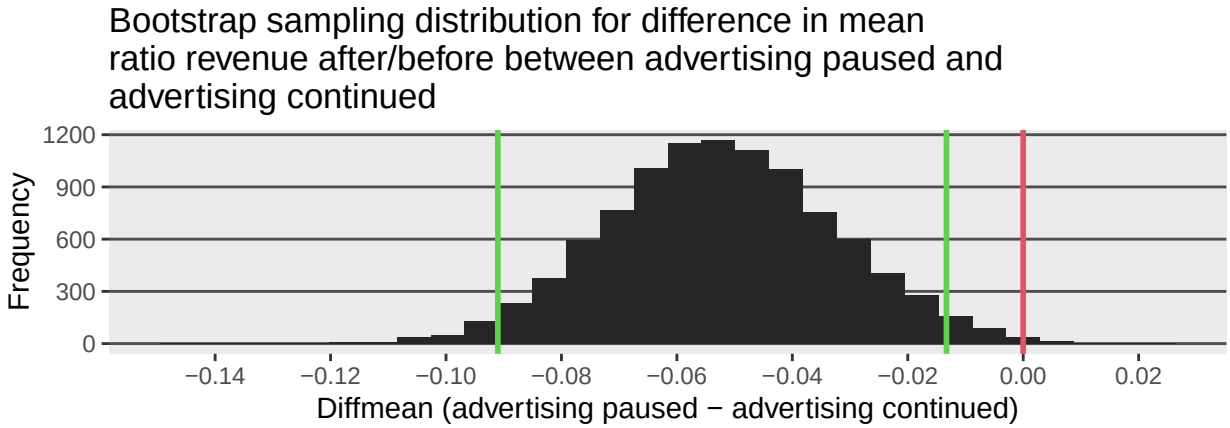
Question: What proportion of American TV watchers would we expect to give a response of 4 or greater to the "Q2_Confusing" question?

Approach: I first filtered the nbc data set to include only "Dancing with the stars". I then bootstrapped a sampling distribution of the proportion of Q2_Confusing responses that were >=4. After that I constructed a histogram and confidence interval to visualize and quantify the results.

Results: Based on my confidence interval I am 95% confident that the proportion of Q2_Confusing responses that were >=4 is between 0.039 and 0.12.

Conclusion: From this it is safe to say the vast majority did not think the show was confusing to the level of a 4 or 5 response.

4(A)



Question: Is there evidence that the revenue ratio is the same in the treatment and control groups? (Treatment group − paused search advertising) (control group − continued search advertising)

Approach: I first created a new variable called "revenueRatio" which was calculated by dividing revenue after by revenue before. I then created a bootstrapped sampling distribution for the difference of means between the two control groups. After that I constructed a histogram and confidence interval to visualize and quantifty the results

Results: Based on my confidence interval I am 95% confident that the difference in mean revenue ratio between the two control groups is between −0.091 and −0.013.

Conclusion: There is not enough evidence to suggest that the revenue ratio is the same in the treatment and control group. This is because 0 is not within the confidence interval which describes the confidence interval as statisticaly significant. More specifically the confidence interval contained all negative values which in this case means there is enough evidence to suggest the advertisng paused group had a smaller revenue ratio compared to the advertising continued group.