Homework 4

Isaiah Benny EID: ieb357

2024-02-20

Github

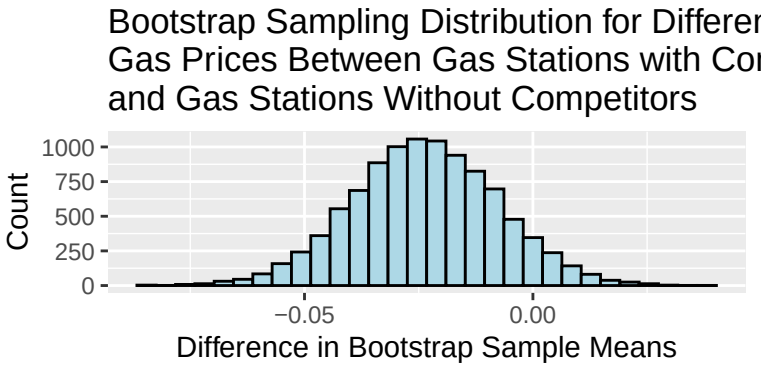
Problem 1

Theory A

Claim

This theory states that gas stations charge more for gas if they lack direct competition in sight.

Evidence



Given the data in our sample, our best guess at the difference in mean gas prices between gas stations with and without competition is -0.023 dollars. Additionally, the bootstrap sampling distribution above shows us that we can be 95% confident that the difference in mean prices between gas stations with and without competition is between -0.055 and 0.008 dollars.

Conclusion

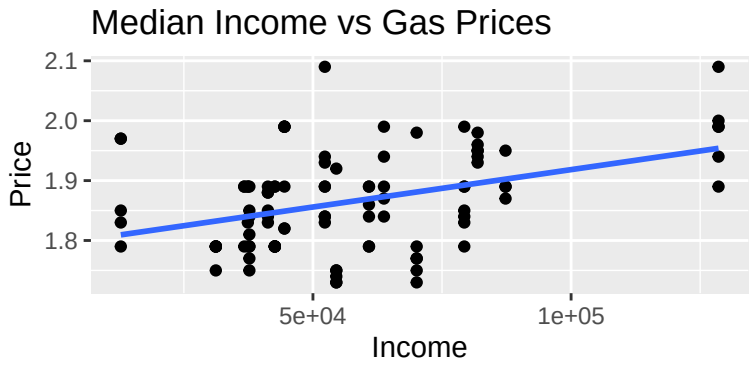
Since the confidence interval contains the value of 0, it is entirely plausible that there is no difference in mean gas prices when comparing gas stations with and without competition in sight. Thus, this theory is unsupported by this collected data.

Theory B

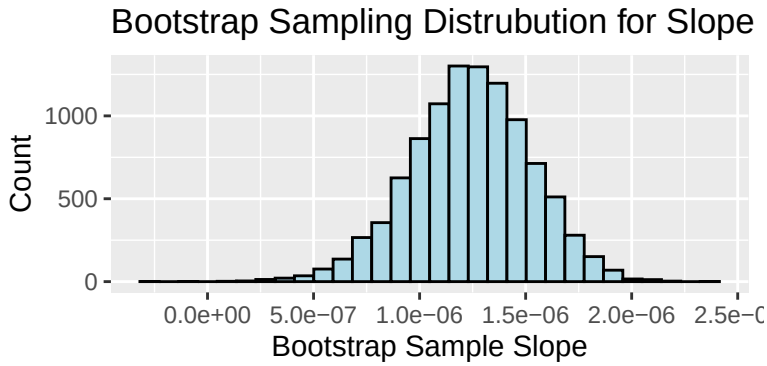
Claim

This theory claims that gas prices increase as the income of the area in which they are located increases.

Evidence



The scatterplot above shows the relationship between median income and gas prices. The line represents the linear model fitted to this data using least squares regression. The slope of this line is 1.2483413×10^{-6}, which represents a predicted average 0.012 dollar change in gas prices for every 10,000 dollar increase in median income.



Using this bootstrap sampling distribution, we are 95% confident that the population slope of the line used to predict gas prices given income is somewhere between 6.5864792×10^{-7} and 1.7776799×10^{-6}.

Conclusion

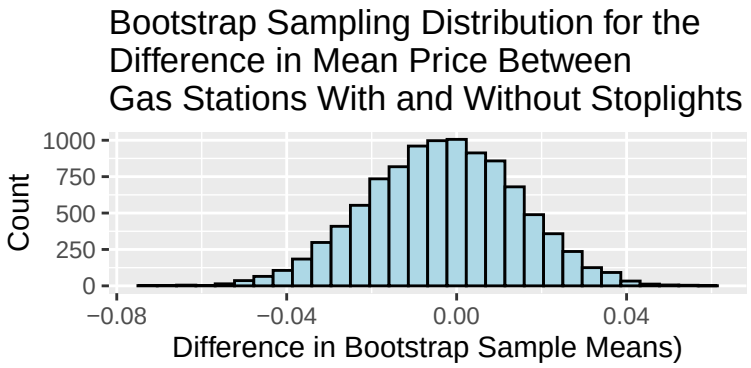
As a result of the confidence interval only containing positive slope values, it seems as if the data supports the idea that there is a positive linear relationship between the income of an area and that area's gas prices. To get an idea of how strong this relationship is, the upper and lower bounds of the confidence interval tells us that a 100,000 dollar increase in median income would lead to a predicted increase in gas prices somewhere in between \$0.07 and \$0.18, which could make a difference when comparing two areas of disparate income over time. However, when the difference in incomes are less than \$100,000, the difference is gas prices are not as pronounced.

Theory C

Claim

This theory states that gas stations at stoplights charge more

Evidence



The bootstrap sampling distribution above tells us that the difference in price between gas stations with and without a stoplight is somewhere between -0.038 dollars and 0.03 dollars, with 95% confidence. The sample estimate for this difference is -0.003 dollars.

Conclusion

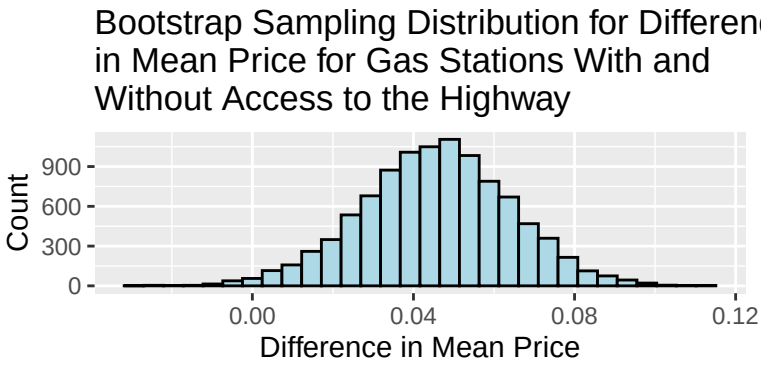
Since 0 is a value that is contained by the 95% confidence interval, it is certainly plausible that there is no difference between gas prices between gas stations with stoplights and gas stations without stoplights. Thus, Theory C is not supported by this data.

Theory D

Claim

Theory D holds the claim that gas stations with direct access to the highway charge more.

Evidence



Using this bootstrap sampling distribution, we are 95% confident that the true difference in mean gas prices between gas stations with and without direct access to the highway is somewhere between 0.008 dollars and 0.081 dollars. The sample estimate for this difference is 0.046 dollars.

Conclusion

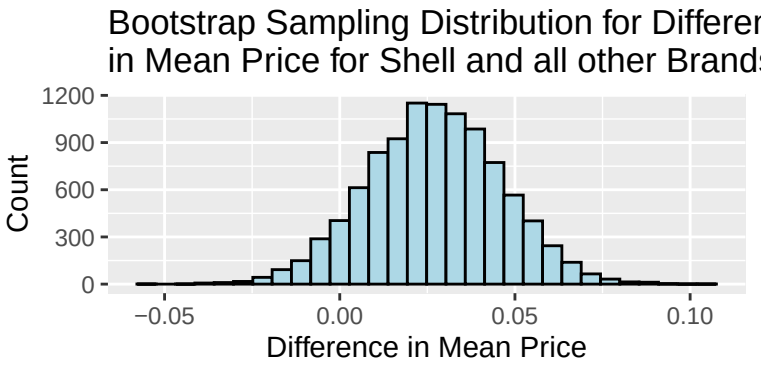
Because there are only positive values contained in the 95% confidence intervals, we can conclude that it is likely that gas stations with access to the highway charge more than stations without access to the highway. Our sample estimate shows us that this difference is somewhere around 4.6 cents. Thus, the sample data does support Theory D, but the data does not support the idea that access to the highway would be associated with a very large difference in gas prices.

Theory E

Claim

This theory asserts that Shell charges more for gas than all other brands.

Evidence



We are 95% confident that the population difference in mean prices for Shell gas stations and all other gas stations lies between -0.011 dollars and 0.064 dollars.

Conclusion

Since the 95% confidence interval contains 0, it is possible that there is no difference between mean prices between Shell and all other non-Shell brands. Therefore, this theory is not supported by the data in this sample.

Problem 2

Part A

Using the bootstrap sampling distribution, we are 95% confident that the average mileage of 2011 S-Class 63 AMGs that were hitting the used-car market when this data was collected is between 26266.4 and 31751.52 miles.

Part B

Using the bootstrap sampling distribution, we are 95% confident that the the proportion of all 2014 S-Class 550s that were painted black is somewhere between 0.417 and 0.453.

Problem 3

Part A

Question

Does the show "My Name is Earl" make people feel happier than the show "Living with Ed"?

Approach

To answer this question, I will be using bootstrap samples to approximate the sampling distribution of the difference in mean viewer response to the statement "this show made me feel happy" for the two shows. This will involve the use of the resample() and diffmean() functions from the mosaic library. Then, I will construct a confidence interval using the confint() function, which is also from the mosaic library.

Results

We are 95% confident that the difference in mean viewer response to the statement "this show made me feel happy" for the two shows "My Name is Earl" and "Living with Ed" is between -0.402 and 0.101.

Conclusion

Since the value of zero is contained in the 95% confidence interval, it is possible that the show "My Name is Earl" makes viewers feel happy to the same degree that "Living with Ed" does. In other words, there isn't evidence that one show consistently produces a higher mean Q1_Happy response among viewers.

Part B

Question

Between the shows "The Biggest Loser" and "The Apprentice: Los Angeles," which show makes people feel more annoyed?

Approach

in order to answer this question, I will be using bootstrap samples to approximate the sampling distribution of the difference in mean viewer response to the statement "this show made me feel annoyed" for the two shows. This will involve the use of the same functions, resample(), diffmean(), and confint(), to create the bootstrap sampling distribution and the confidence interval.

Results

The difference in mean viewer response to the statement between the show "The Biggest Loser" and "The Apprentice: Los Angeles" is somewhere between -0.521 and -0.017, with 95% confidence.

Conclusion

As a result of the 95% confidence interval only containing negative values, we can conclude that the show "The Apprentice: Los Angeles" consistently produces a higher mean Q1_Annoyed response among viewers when compared to the show "The Biggest Loser."

Part C

Question

What proportion of American TV watchers would we expect to give a response to 4 or greater to the statement "I found this show confusing" when the show is "Dancing with the Stars"?

Approach

To answer this question, I will use a similar approach to the previous questions and make a bootstrap sampling distribution. In this case, the parameter of interest is the proportion of Americans who report themselves to be confused by the show. This will call for the use of the functions resample(), prop(), and confint(), which are all from the mosaic library.

Results

We are 95% confident that the proportion of American TV watchers would we expect to give a response of 4 or greater to the "Q2_Confusing" question lies in the interval of 0.039 and 0.116.

Conclusion

The 95% confidence interval tells us that we would expect anywhere from 3.9% to 11.6% of Americans to be confused by the show "Dancing with the Stars."

Problem 4

Question

What is the difference in Ebay's revenue ratio when Ebay pauses their advertising on Google AdWords versus when they don't?

Approach

In order to answer this question, I will use bootstrap samples to create an approximation for the sampling distribution of the difference in mean revenue ratios for DMAs where advertising on Google AdWords is paused and is not paused.

Results

We are 95% confident that the difference in revenue ratios between DMAs where advertising on Google AdWords is paused and DMAs where the advertising is not paused is in the interval from -0.09 to -0.013.

Conclusion

Since the confidence interval only contains negative values, we can conclude that pausing advertising on Google AdWords is associated with a decrease in Ebay's revenue ratio.