# HW 3

Genavieve Middaugh

2025-02-07

## Problem 1 - Gas Station Prices

In Problem 1 we are assessing multiple theory given the prices of gas stations with multiple outside factors. The goal is to try to understand the variation in gas station prices and what causes that variation. The data given by GasPrices.csv a dataset made by students in spring of 2016 on a project.

The important variables are below:

- ID: Order in which gas stations were visited
- Name: Name of gas station
- Price: Price of regular unleaded gasoline, gathered on Sunday, April 3rd, 2016
- Highway: Is the gas station accessible from either a highway or a highway access road?
- Stoplight: Is there a stoplight in front of the gas station?
- Competitors: Are there any other gas stations in sight?
- Zipcode: Zip code in which gas station is located
- Income: Median Household Income of the ZIP code where the gas station is located based on 2014 data from the U.S. Census Bureau
- Brand: ExxonMobil, ChevronTexaco, Shell, or Other.

The theorys we are trying to either prove or disprove are below:

- A) Gas stations charge more if they lack direct competition in sight.
- B) The richer the area, the higher the gas prices.
- C) Gas stations at stoplights charge more.
- D) Gas stations with direct highway access charge more.
- E) Shell charges more than all other non-Shell brands.

**1. Gas stations charge more if they lack direct competition in sight.**

Bootstrap sampling distribution for difference in mean gas prices based on
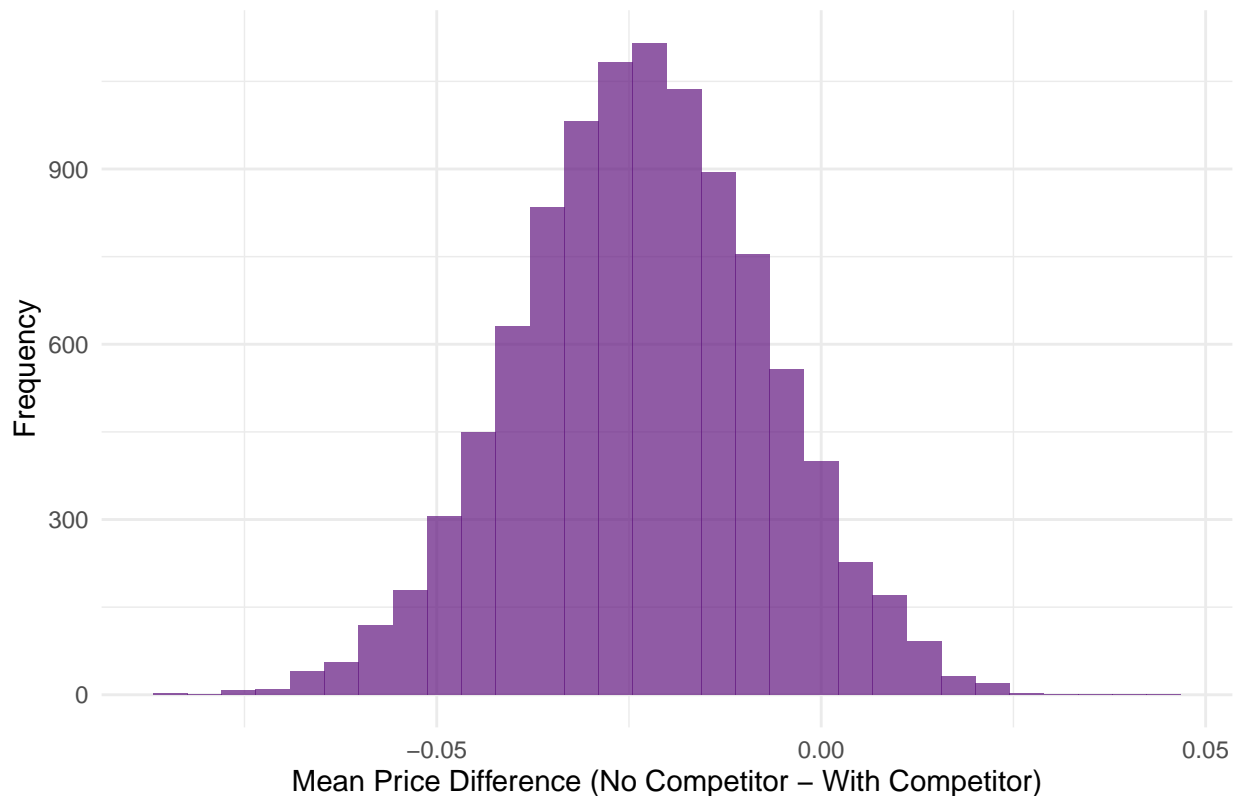


If you have ever driven around, anywhere in america im sure you've seen your fair share of gas stations, the fluctuation of prices varies from place to place, but one may wonder if having competitors near by lead to gas stations having lower prices. The Competitors and Price variables from gasprices data was used to understand if there were competitors around or not and what the prices were for each gas station.

The above graph shows a bootstrap sampling distribution of the difference in the means of these 2. The graph skews to the left of 0 only slightly, so we can say with 95% confidence that the average difference in gas price between gas stations with no competitors and with competitors is between -0.055 and 0.008 cents. Since the graph is so close to 0, this means it is a common occurrence for there to be no difference, meaning the average prices were the same or at least almost the same. It seems that there is very little correlation especially considering the mean is -0.023, a very low number and when you think about the difference in gas prices this looks like 2.34 and 2.32 difference is quite small.

To conclude, the data does not present any real evidence of the idea that gas stations prices are correlated to whether there are competitors nearby. But we must remember that this data was only done over the course of 1 day, in the spring of 2016 in Austin. This data does not cover enough time for us to draw conclusions on if over time this is the same conclusion.

**2. The richer the area, the higher the gas prices.**

Because gas stations have various different pricing depending on where you go in America, its reasonable to quesiton if the general income level of the community the gas station is in has any affect on the prices of gas. Below is a graph that is a distribution of samples looking for correlation between gas price and income of the community it is located in.

Bootstrap Sampling Distribution for Correlation Between Gas Price and In
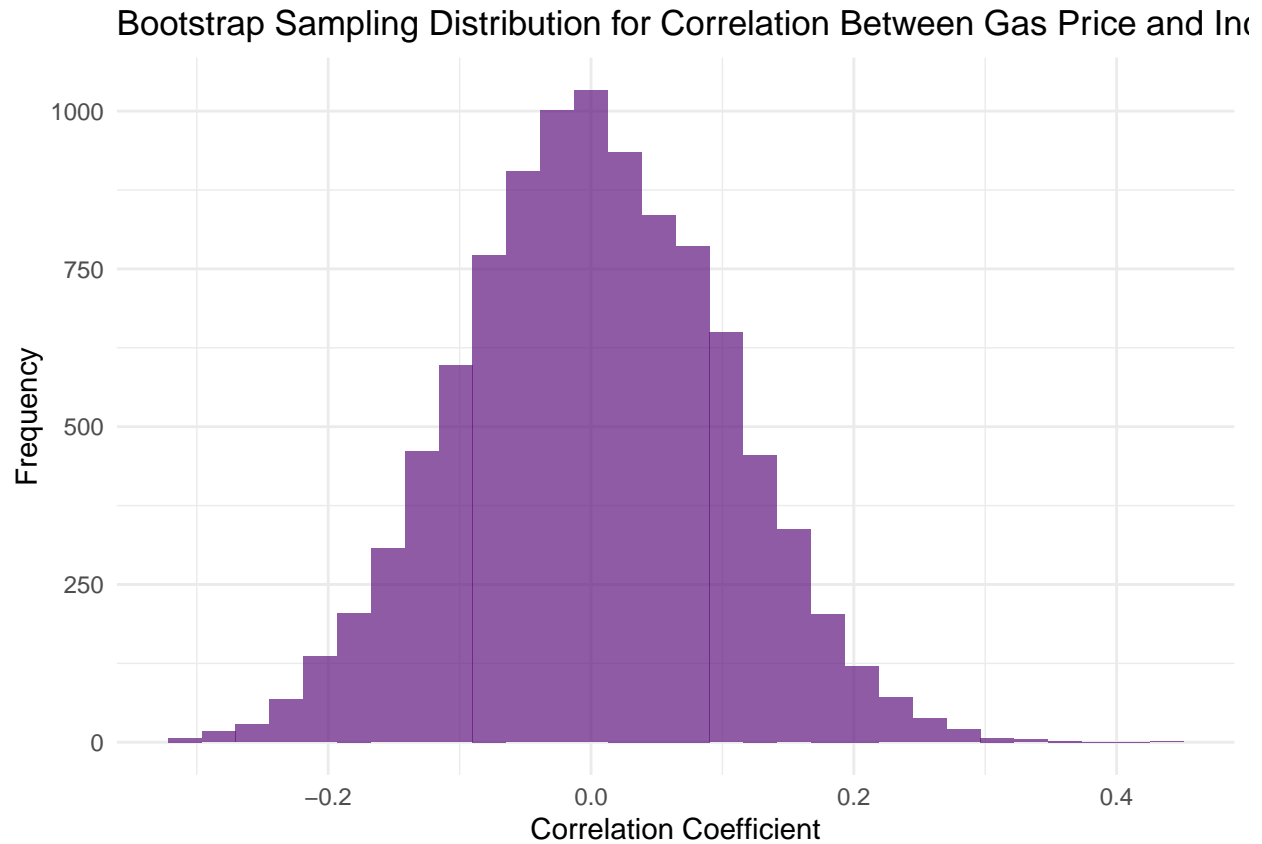


Given the Distribution we can say with 95% confidence that the correlation between gas prices and the average income of the area around it is between -0.194 and 0.196. This confidence interval includes 0 meaning that there is not much evidence of these variables causing one another. The income of the area does not seem to suggest the price of gas in the same area. There is a wide range of correlation values between incomr and gas prices, and 0 is right in between the bell shaped distribution, which suggest there is no correlation between the two.

This graph shows the correlation between the two variables, which is 0.396. The correlation is moderate or at least above 0, which suggest some relationship between the two variables.

After looking at these two facts, the initial correlation between the two variables shows there is a moderate correlation between the two, but the sampling distribution shows that it varies and that there is almost no correlation. I think with this information we can conclude that the data is two small, and done over such a short amount of time, we cannot conclude anything about gas stations and their prices based on their average income level using this data.

**3. Gas stations at stoplights charge more.**

The next theory is that some Gas stations charge more for gas if it is near a stop light. I think this is a reasonble idea, but lets see if it follows through.

**Bootstrap Sampling distribution of the differences in mean between gas prices and stoplight status**
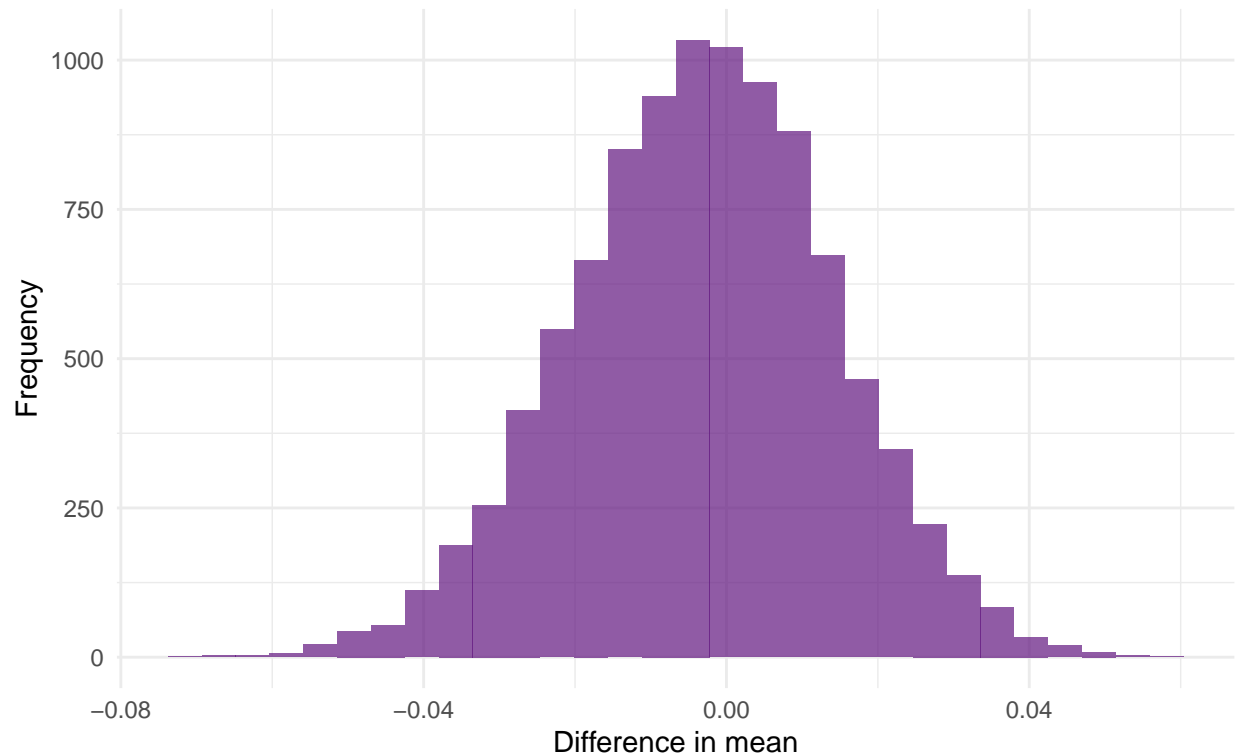


After bootstrap sampling the distribution of the Price and Stoplight variables we find that the mean price for gas stations if there is no light is 1.866 and for gas stations near a light were 1.863. This graph is clearly bell shaped with a center ranging near zero, meaning that the difference in means often ends up being zero or close, this means there is little difference between the means.I can say with 95% confidence that the difference in mean between the stores with stop lights and without gas prices was between -0.038 and 0.03 cents. Zero is included in the interval once again meaning there is too many outcome and a lot of variance.

The conclusion here is that we once again cant say for sure if there is any evidence of gas stations charging more for gas if they are near a gas station.

**4. Gas stations with direct highway access charge more.**

Related to the last theory this one asks if gas stations with direct highway access charge more for gas.

## Bootstrap Sampling distribution of the differences in mean between gas prices and nearby highway status
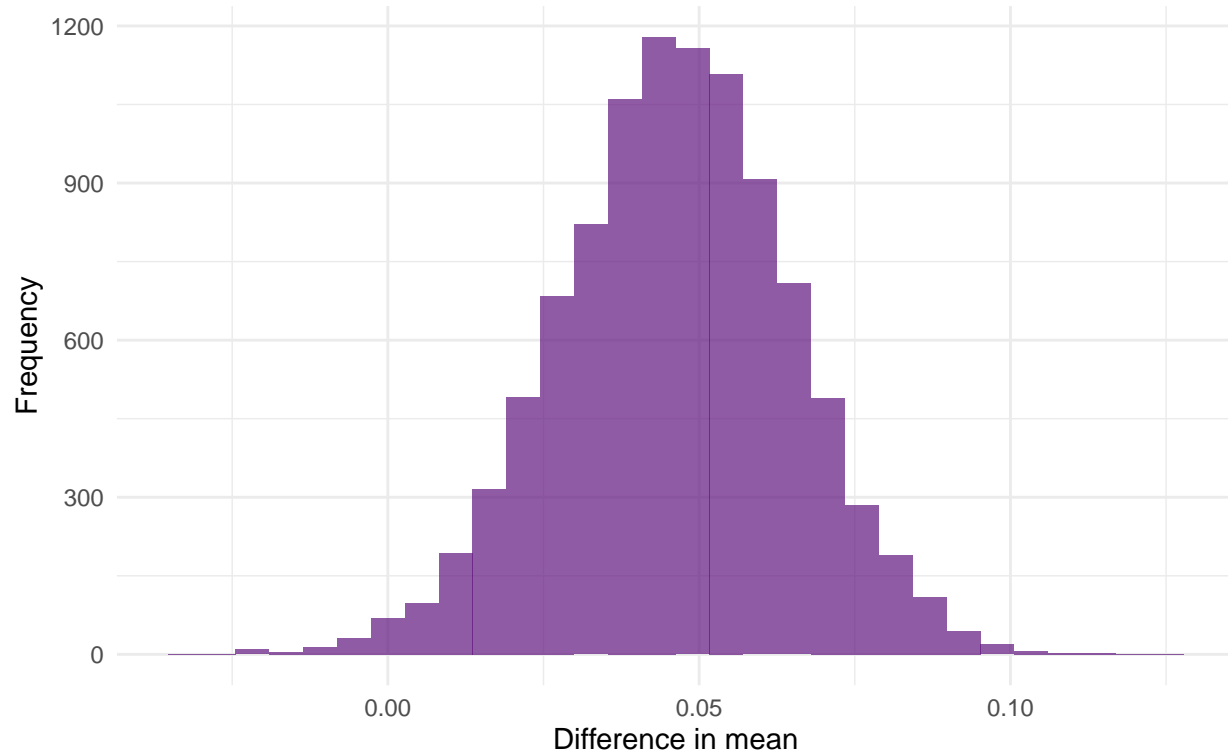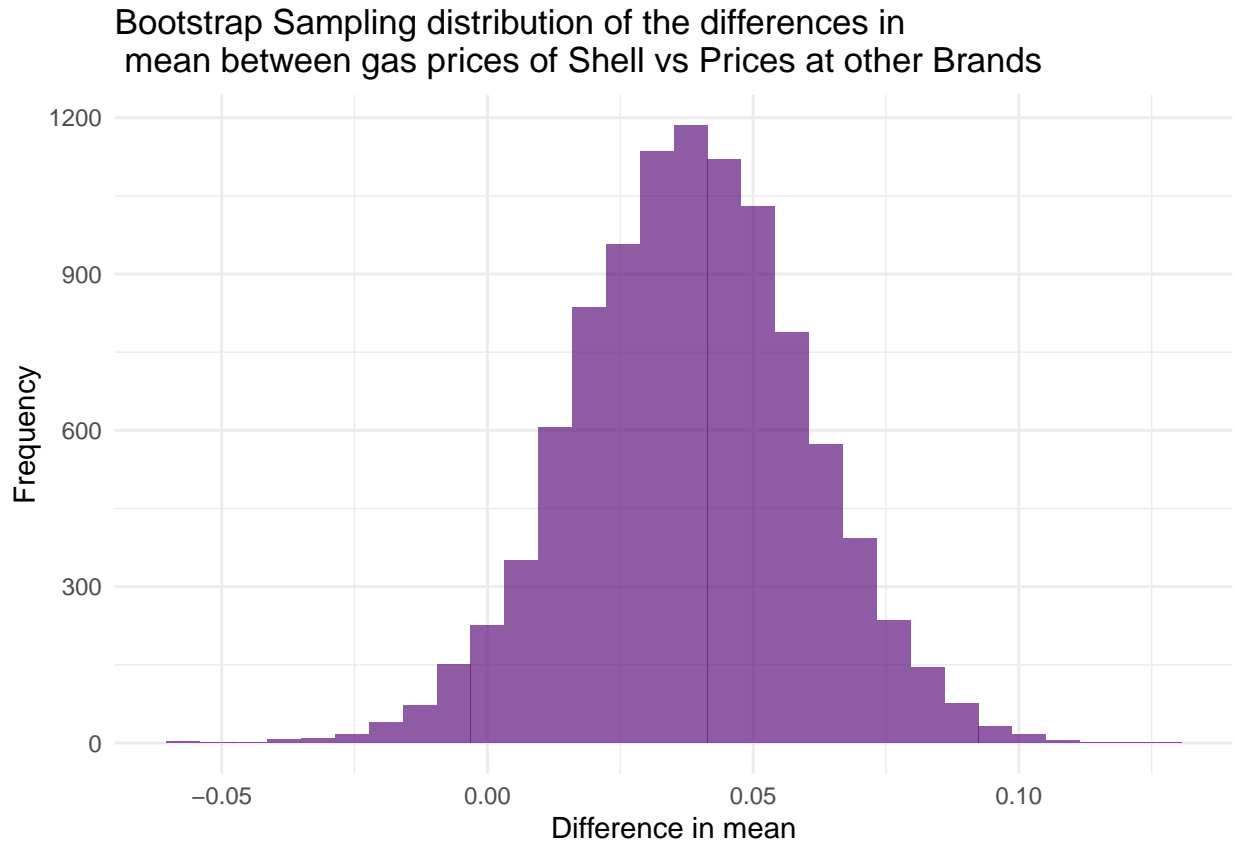


First we can see that the general mean from the data for stores not near a highway was 1.854 and for stores near a highway 1.9. They are quite close in range and don't seem to differ, although we can acknowledge that the stores not near highways do have a lower average in price. In the Graph above it shows the sampling distribution of the difference in means from these two variables. We see that it is bell shaped and consistent, and we also see that these are higher values, meaning that there were differences in the mean, suggesting that stores near highways charge more. So I am 95% confident that the difference of means between gas prices and their highway status is between 0.009 and 0.082 cents.

In conclusion , given that the confidence interval is above 0, The graph does show that there is some evidence, stating that stores near highways charge more for gas,but its not a strong difference in means so its only a possibility. So we can say that on the day the data was gathered and the area (2016, and in Austin), that on this day there was slight evidence that gas stations near highways charged more for gas.

**5. Shell charges more than all other non-Shell brands.**

This last theory asks if Shell charges more for gas than any other brand. I have always personally avoided Shell, becasue they seem to always be the most expensive gas around me, so I am curious what the information will show.

## Bootstrap Sampling distribution of the differences in mean between gas prices of Shell vs Prices at other Brands



The graph above is consistent and slightly above zero, this could suggest that there is a slight difference in price between Shell and other brands gas prices. When I get the confidence interval I am 95% confident that the difference of mean prices between Shell brand and other brands gas prices is between NA and NA cents. This interval does not include zero, but it is pretty low meaning there is a slight significance to this evidence. Shells average price given this data is 1.88 which the others Chevron is 1.88 , exxon is 1.86 and other is 1.85 approximatly.

To conclude, this data shows that there is a slight significance to the evidence of Shell being more expensive than other brands, but not very signifigant. The difference in mean of price is low, and doesn't have a high range.

## Problem 2 - Mercedes S-Class

The dataset has data on 30,000 used Mercedes S-Class vehicles sold on cars.com. Mercedes is a popular luxury car brand and here we will be looking at

- price
- mileage
- trim
- color
- year

### Part A

In part A, we need to look at cars from 2011 with the trim being 63 AMG, and we are going to simulate a boostrap sampling distribution to find the average mileage of the 2011 S-Class 63 AMG at this time when the data was gathered.

## [1] 26261

## [1] 31814

After making those calculations, I am 95% confident that the average mileage of 2011 S-Class 63 AMG cars hitting the used car market falls between $2.6261 \times 10^4$ and $3.1814 \times 10^4$.

### Part B

In part B, we are looking for the proportion of all 2014 S-Class 550s that were Black. To do this we will filter the dat, mutate the data, and find a 95% confidence interval.

## [1] 41.68

## [1] 45.34

I filtered the data to only contain 2014 years and trim 550 and also added a variable that indicated if the car was black or not using true and false. After bootstrapping to find the proportion, so I am 95% confident that the proportion of all 2014 S-Class 550's that are black is between 41.68% and 45.34%.

## Problem 3 -