

HW03

Shashwat Mishra, EID: sdm4329

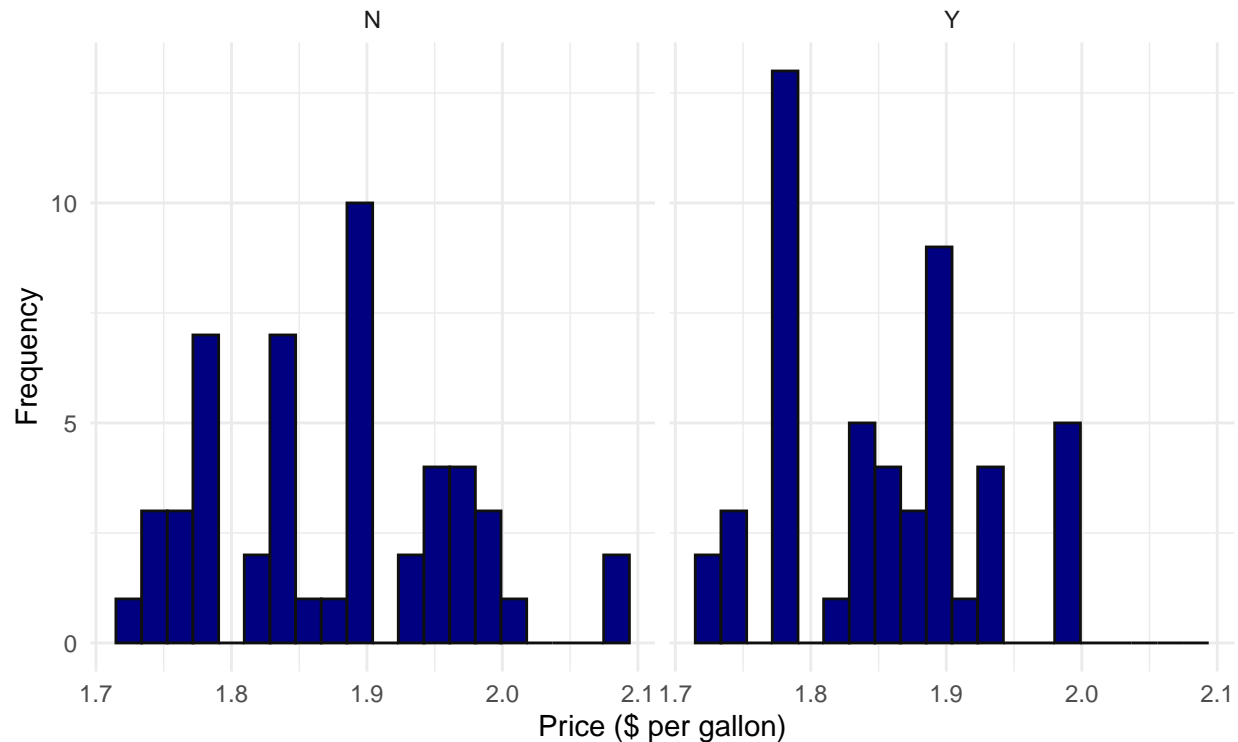
<https://github.com/shashwat-m/SDS315>

Question 1

Claim 1: Gas stations charge more if they lack direct competition

Gas Price Distribution by Competitor Presence

Comparing gas prices for stations with vs. without nearby competitors



```
##      name      lower      upper level      method      estimate
## 1 diffmean -0.05477641 0.007715707 0.95 percentile -0.01284929
```

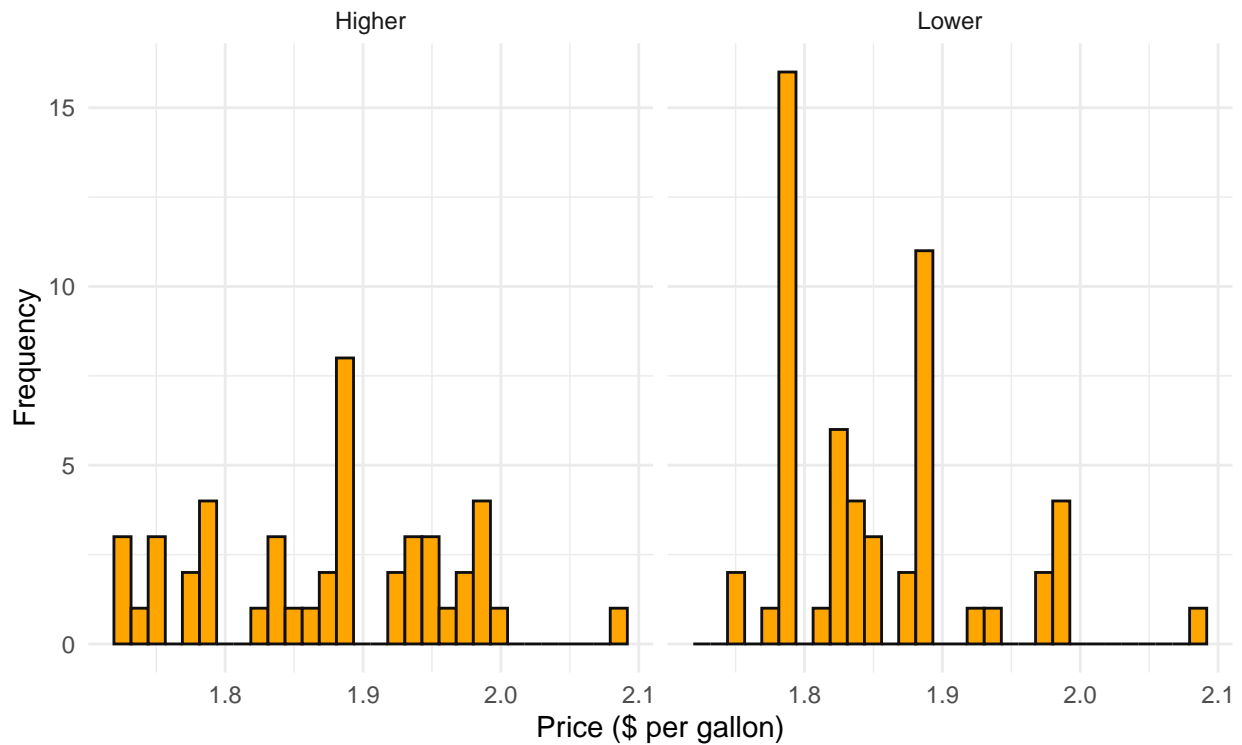
Evidence: When looking at the graph, it's shown that whether or not a gas station has competitors nearby, the distributions of the gas prices are relatively the same. The difference in price between gas stations with and without competition in sight is somewhere between -0.0555 and 0.0076, with 95% confidence.

Reasoning: Because the difference in price includes 0, there's not conclusive evidence to show that competitors in the view of a certain gas station affect gas prices.

Claim 2: The richer the area, the higher the gas prices.

Gas Price Distribution by variation in Median Income

Comparing gas prices for in areas that earn higher vs. lower than the median income (~52

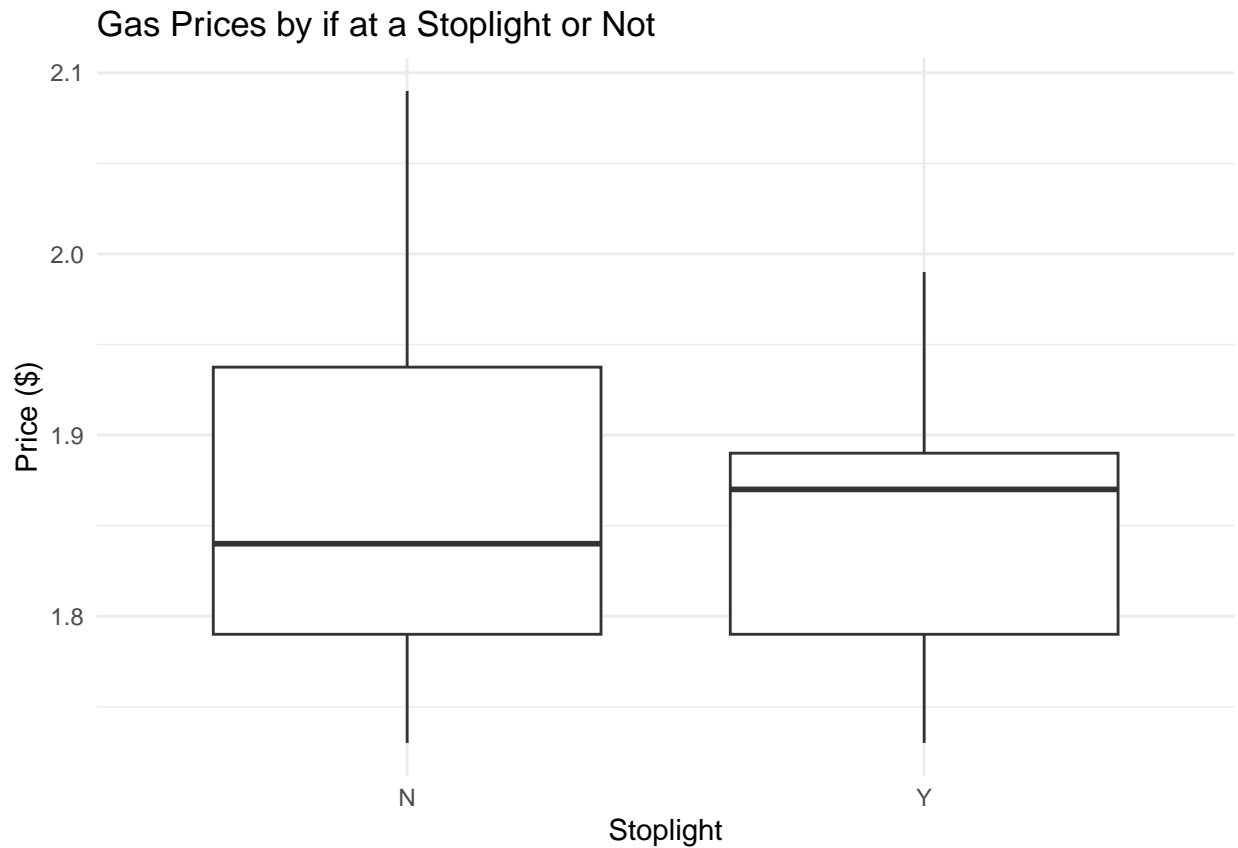


```
##      name      lower  upper level  method  estimate
## 1 diffmean -0.05394901 0.0119574  0.95 percentile -0.02132411
```

Evidence: Although it's seen that there is some noticable difference in the gas prices in areas with higher median income, the difference in price between gas stations in lower and higher income areas is somewhere between -0.0116 and 0.0527, with 95% confidence.

Reasoning: Because the range of the difference includes 0, the variance in median income does not show significant evidence that the median income of an area will increase or decrease gas prices.

Claim 3: Gas stations at stoplights charge more.

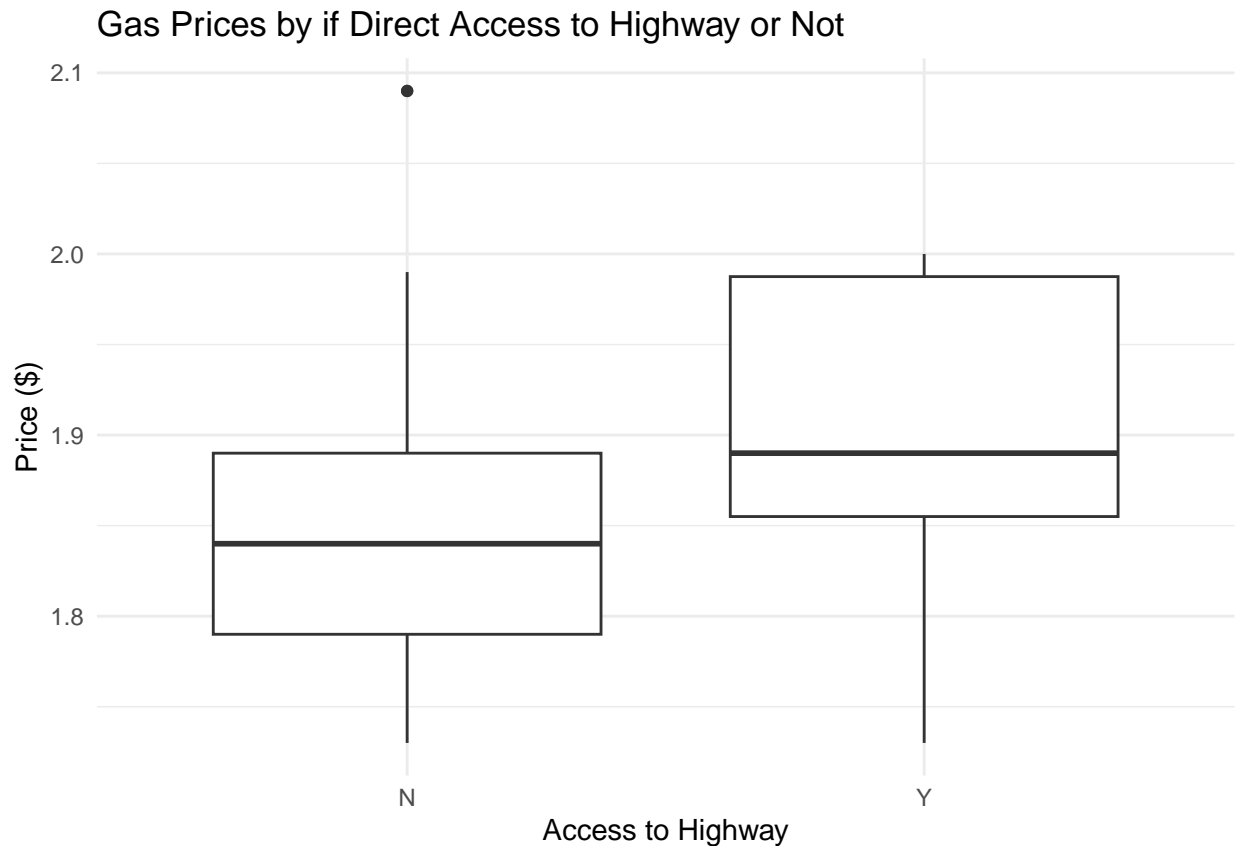


```
##      name      lower      upper level      method      estimate
## 1 diffmean -0.03798271 0.03025081 0.95 percentile -0.003299916
```

Evidence: It's shown by the graph that the median of the gas prices when a stoplight is present is higher than when no stoplight is present. With the prices being \$1.87 and \$1.84, respectively. The difference in price between gas stations in near a stoplight and not near a stoplight is somewhere between -0.0382 and 0.0300, with 95% confidence.

Reasoning: Similar to the other 2 claims, the confidence interval's range includes 0, which means a stoplight's presence is insignificant to the price of gas at the station.

Claim 4: Gas stations with direct highway access charge more



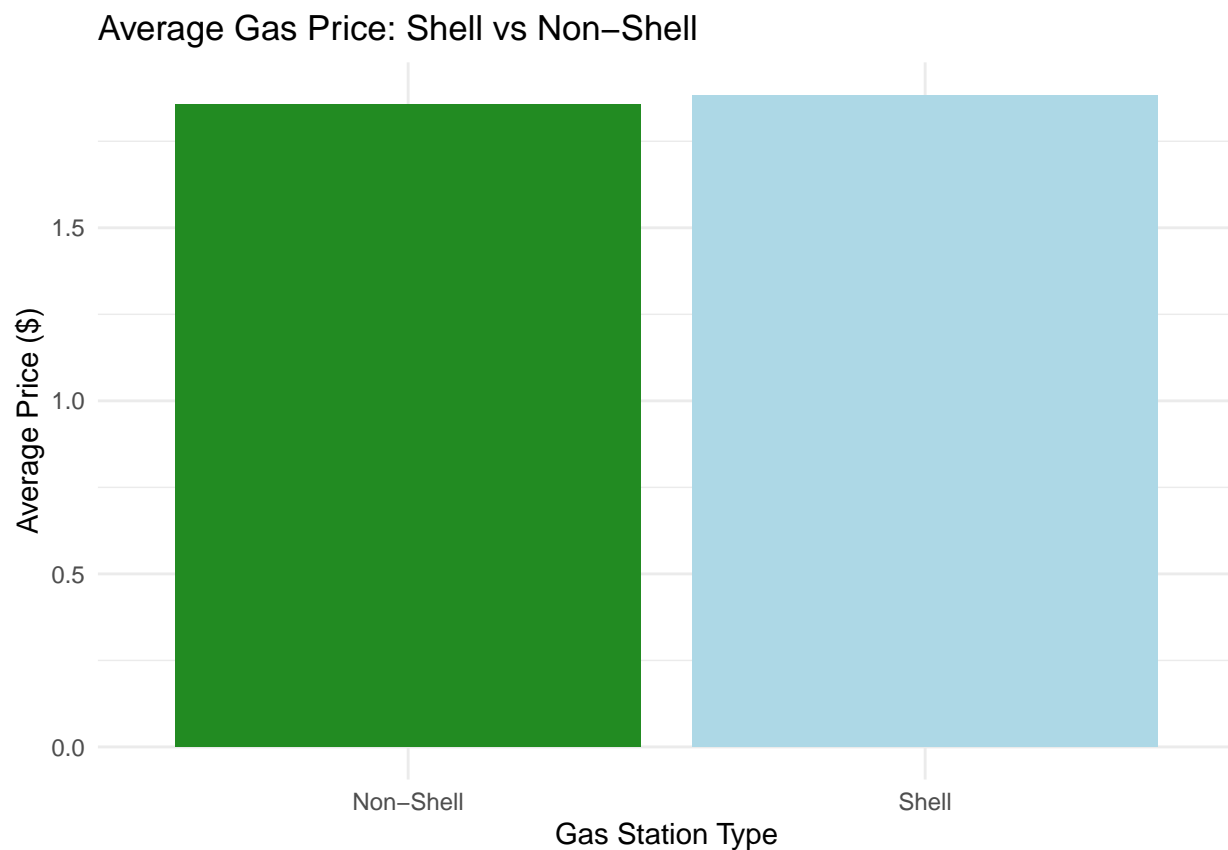
```
## # A tibble: 2 x 2
##   Highway Median_Price
##   <chr>         <dbl>
## 1 N             1.84
## 2 Y             1.89
```

```
##      name      lower      upper level    method  estimate
## 1 diffmean 0.008916383 0.08122078 0.95 percentile 0.0456962
```

Evidence: The box plots show that there is a higher median price at stations that have direct highway access, being \$1.89 - as opposed to no highway access, where the median price is \$1.84. The confidence interval shows that there is slight variance in gas prices whether or not there is direct highway access - somewhere between 0.0087 and 0.0804, with 95% confidence.

Reasoning: Because the confidence interval is above 1, and does not include 0 in it, we can say with 95% confidence that there is a variance of approximately \$0.05 per gallon dependent on whether or not there is direct highway access or not.

Claim 5: Shell charges more than all other non-Shell brands.



```
##      name      lower      upper level      method      estimate
## 1 diffmean -0.009859759 0.06498003 0.95 percentile 0.02740421
```

Evidence: It's shown that the prices of gas dependent on whether the station is a Shell or not are very similar. The difference in price is in the range of -0.009 to 0.064, with 95% confidence.

Reasoning: Because the confidence interval range includes 0, we cannot say that Shell's ownership changes gas prices with certainty.

Question 2

Part A

```
##      lower      upper
## 1 26842.42 34693.61
```

Based on the confidence interval's range being between 26926.91 and 34716.97, we are 95% confident that the true average mileage of used 2011 S-Class 63 AMG cars in the population falls between 26,926.91 and 34,716.97 miles.

Part B

```
##      lower      upper
## 1 0.4168 0.4531
```

This means we are 95% confident that the true proportion of all 2014 S-Class 550s that were painted black falls between 41.68% and 45.28%. If we could analyze all 2014 S-Class 550s, the proportion of black cars would likely be within this range. Since the confidence interval is around 42%-45%, this suggests that nearly half of all 2014 S-Class 550s were black.

Question 3

Part A

```
##      lower      upper
## 1 -0.394 0.108
```

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

Approach:

- 1) Filter the dataset for these two shows.
- 2) Bootstrap a 95% confidence interval for the difference in mean Q1_Happy scores.
- 3) Check if the CI includes zero to determine if there's a significant difference.

Results: It was shown that the difference in mean happiness between the shows "Living with Ed" and "My Name is Earl" is somewhere between -0.4007 and 0.1036, with 95% confidence.

Interpretation: Since the confidence interval includes 0 (-0.397, 0.101), there's not strong evidence that one show produces significantly higher happiness ratings than the other. This means viewers of "Living with Ed" and "My Name is Earl" had similar happiness responses on average.

Part B

Question: Do viewers of “The Biggest Loser” feel more annoyed (Q1_Annoyed) than viewers of “The Apprentice: Los Angeles”?

Approach:

- 1) Filter the dataset for these two shows.
- 2) Bootstrap a 95% confidence interval for the difference in mean Q1_Annoyed scores.
- 3) Check if the CI includes zero to determine if there’s a significant difference.

```
##      lower  upper  
## 1 -0.529 -0.023
```

Results: The confidence interval does not include 0, therefore we can say, with 95% confidence, that this *The Apprentice: Los Angeles* is “more annoying” than *The Biggest Loser*, as `diffmean(Q1_Annoyed ~ Show)`, shows the difference as Mean Annoyance of “Biggest Loser” – Mean Annoyance of “Apprentice: LA”, which is always negative.

Conclusion: This means that The Apprentice was consistently more annoying to viewers.

Part C

Question: What proportion of viewers agreed (4) or strongly agreed (5) that “Dancing with the Stars” was confusing?

Approach:

- 1) Filter dataset for responses where Show == “Dancing with the Stars”.
- 2) Create a binary variable (isConfused): TRUE if Q2_Confusing ≥ 4, FALSE otherwise.
- 3) Bootstrap the proportion of confused viewers and compute a 95% confidence interval.

Results:

```
##      lower upper  
## 1 0.0387 0.116
```

Conclusion: We estimate that between ~3.87% and ~11.6% of American TV viewers found “Dancing with the Stars” confusing with 95% confidence

This means a very small population finds the show confusing (A little embarrassing)

Question 4

Question: Does turning off paid search advertising significantly reduce eBay's revenue?

Approach:

- 1) Load and filter the dataset.
- 2) Create a new variable for the revenue ratio ($\text{rev_ratio} = \text{rev_after} / \text{rev_before}$).
- 3) Bootstrap a 95% confidence interval for the difference in mean revenue ratios between the treatment and control groups.
- 4) Assess whether the confidence interval includes 0 (i.e., no significant difference).

```
##      lower  upper
## 1 -0.0909 -0.0129
```

Results: The results show that the entire confidence interval is negative, ranging from (-0.0918, -0.0134). Because this interval does not cross 0, we can say that with 95% confidence, the revenue ratio was lower in the treatment group (where ads were paused) compared to the control group (where ads continued) because it stays negative.

The estimated drop in revenue ratio falls between 1.34% and 9.18%.

Conclusions:

The treatment group (ads paused) saw a statistically significant decrease in revenue ratio compared to the control group.

Since the entire confidence interval is negative, there is strong evidence that paid ads were driving additional revenue.

As for what this implies for the business, turning off paid search ads led to a real revenue loss, suggesting that eBay should continue investing in Google AdWords.