Homework4

Anish Ravindran

2024-02-19

Name: Anish Ravindran
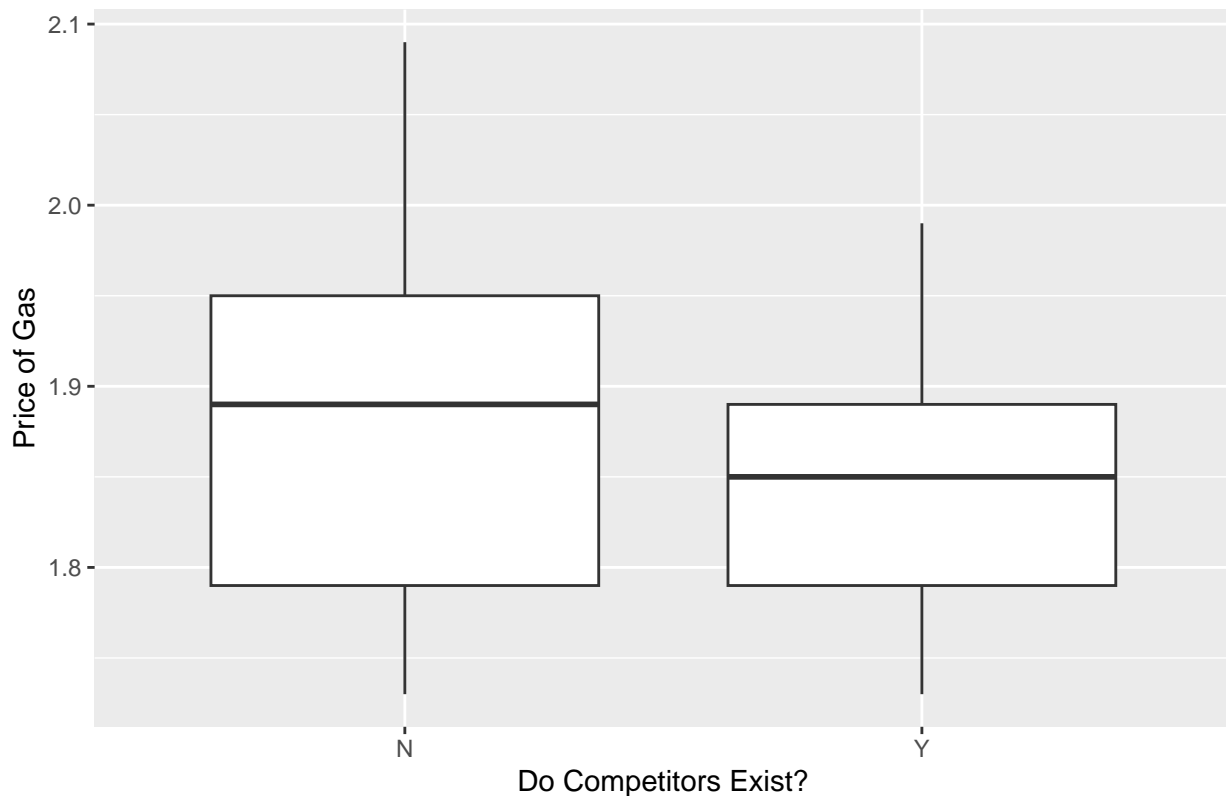uteid: ar76454
github link: https://github.com/AnishRavi5/Homework4

Question 1

Part A

Price of Gas With and Without Direct Competition


```
##   name     lower    upper level      method estimate
## 1    N 1.851607 1.900204  0.95 percentile 1.884884
## 2    Y 1.832692 1.872392  0.95 percentile 1.854310
```

Claim - Gas stations charge more if they lack direct competition in sight. Evidence - When looking at the 95 percent confidence interval we see that when there is no competition in sight gas stations charge between 1.85 to 1.90 with 95 percent confidence while when there is competition they charge between 1.83 to 1.87 with 95 percent confidence. This difference is significant enough as it comes out to a difference of 2 to 3 cents. Conclusion - The theory is supported by the data as the confidence intervals tell us that there is a difference

of 2 to 3 cents when there is competition compared to when there is no competition. This difference may seem small but when it comes to gas prices the difference multiplies quickly as when filling up a full tank the difference can come out to around 50 cents to a dollar. Due to this we know that when there is no direct competition gas stations tend to charge more.

Part B

```
## `geom_smooth()` using formula = 'y ~ x'
```


Relationship between Median Income and Price of Gas

```
##   name      lower     upper level      method   estimate
## 1  cor 0.1977935 0.5626592  0.95 percentile 0.3471397
```

Claim - The richer the area, the higher the gas prices Evidence - When looking at the 95 percent confidence interval between the correlation of income and gas prices we see a range between 0.1999 and 0.5713 with 95 percent confidence. This is a confidence interval with a very big range. Conclusion - The theory is not supported as the confidence interval spans from around .2 to .57 which is a very big range showing us that the data has no real correlation. This means that the theory of there being higher gas prices in richer areas is false.

2

Part C

Price of Gas at Stoplights vs. No Stoplights
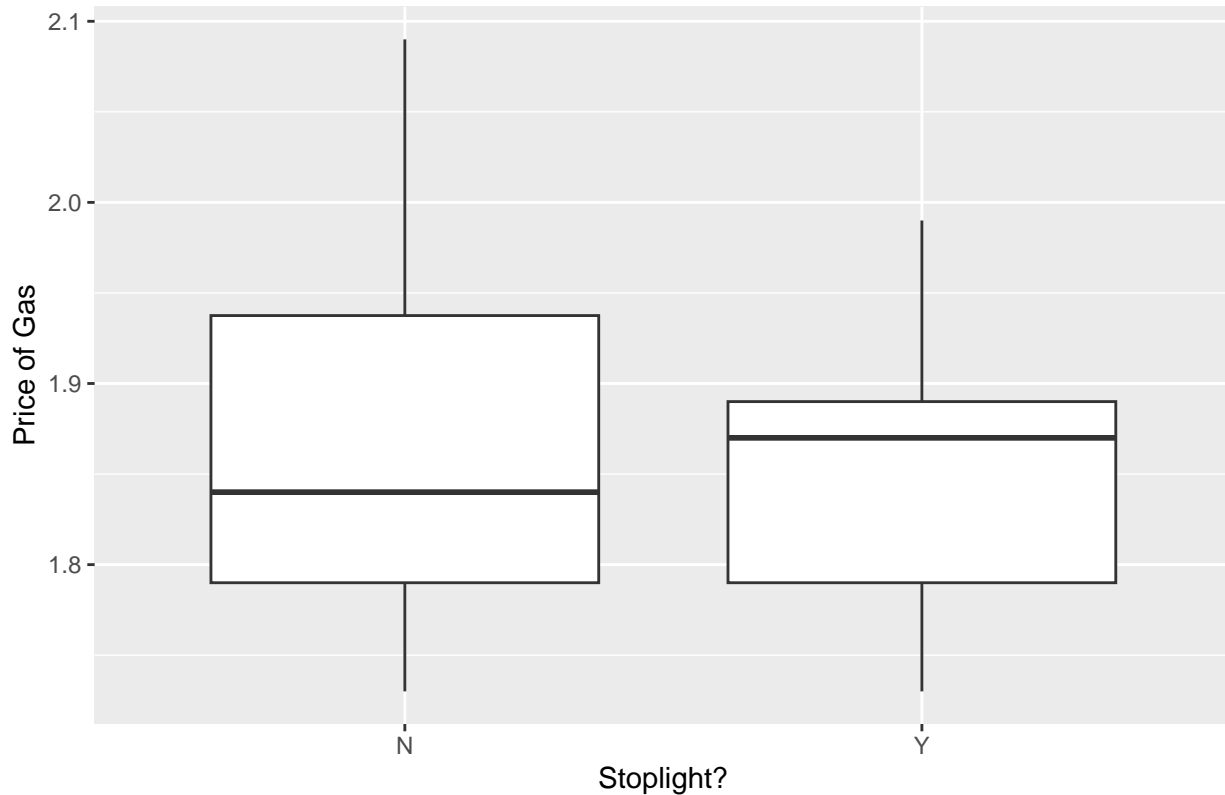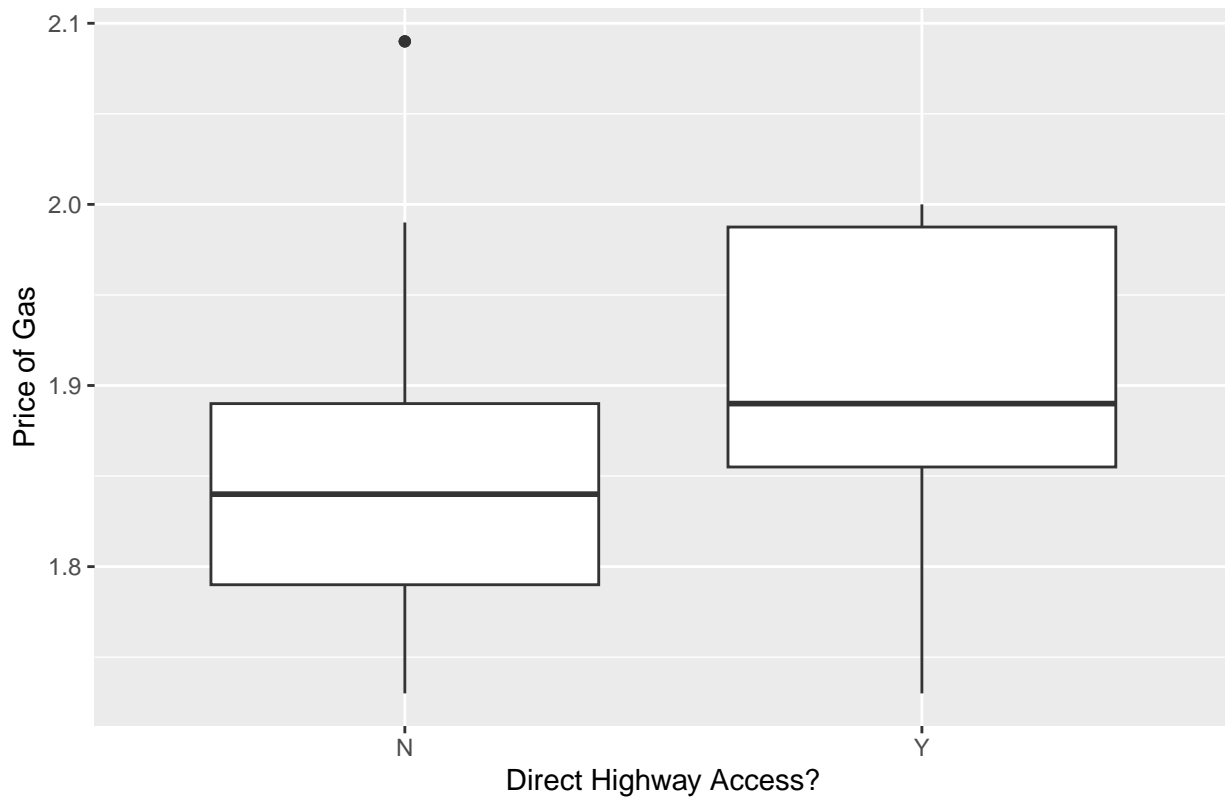


```
##   name    lower    upper level     method estimate
## 1    N 1.838235 1.896776  0.95 percentile 1.854737
## 2    Y 1.844782 1.881755  0.95 percentile 1.849206
```

Claim - Gas stations at stoplight charge more. Evidence - When looking at the 95 percent confidence interval we see that when there is no stoplight gas stations charge between 1.84 to 1.9 with 95 percent confidence while when there is a stoplight they charge between 1.84 to 1.88 with 95 percent confidence. The confidence interval for both cover a very similar range as the prices overlap. Conclusion - The theory is not supported as the confidence intervals tell us that the price is usually very similar at gas stations regardless of whether or not there is a stoplight present in the area. The range is slightly bigger for the gas stations located where there is no stoplight but the range still encompasses the range of the confidence interval for gas stations located with stoplights. Due to this we know that stoplights no do affect the price at gas stations.

Part D

Price of Gas With and Without Highway Access


```
##   name    lower    upper level      method estimate
## 1    N 1.836667 1.872105  0.95 percentile 1.853780
## 2    Y 1.868419 1.931111  0.95 percentile 1.866316
```

Claim - Gas stations with direct highway access charge more. Evidence - When looking at the 95 percent confidence interval we see that when there is no highway access gas stations charge between 1.84 to 1.87 with 95 percent confidence while when there is highway access they charge between 1.87 to 1.93 with 95 percent confidence. This difference is significant enough as it comes out to a difference of 3 to 6 cents. Conclusion - The theory is supported by the data as the confidence intervals tell us that there is a difference of 3 to 6 cents when there is highway access compared to when there is no highway access. This difference is significant as the confidence intervals barely even overlap and a difference of 3 to 6 cents can come out to a difference of a few dollars when filling up a full tank of gas. Due to this we know that when there is direct highway access gas stations tend to charge more.

Part E

Price of Gas by Brand
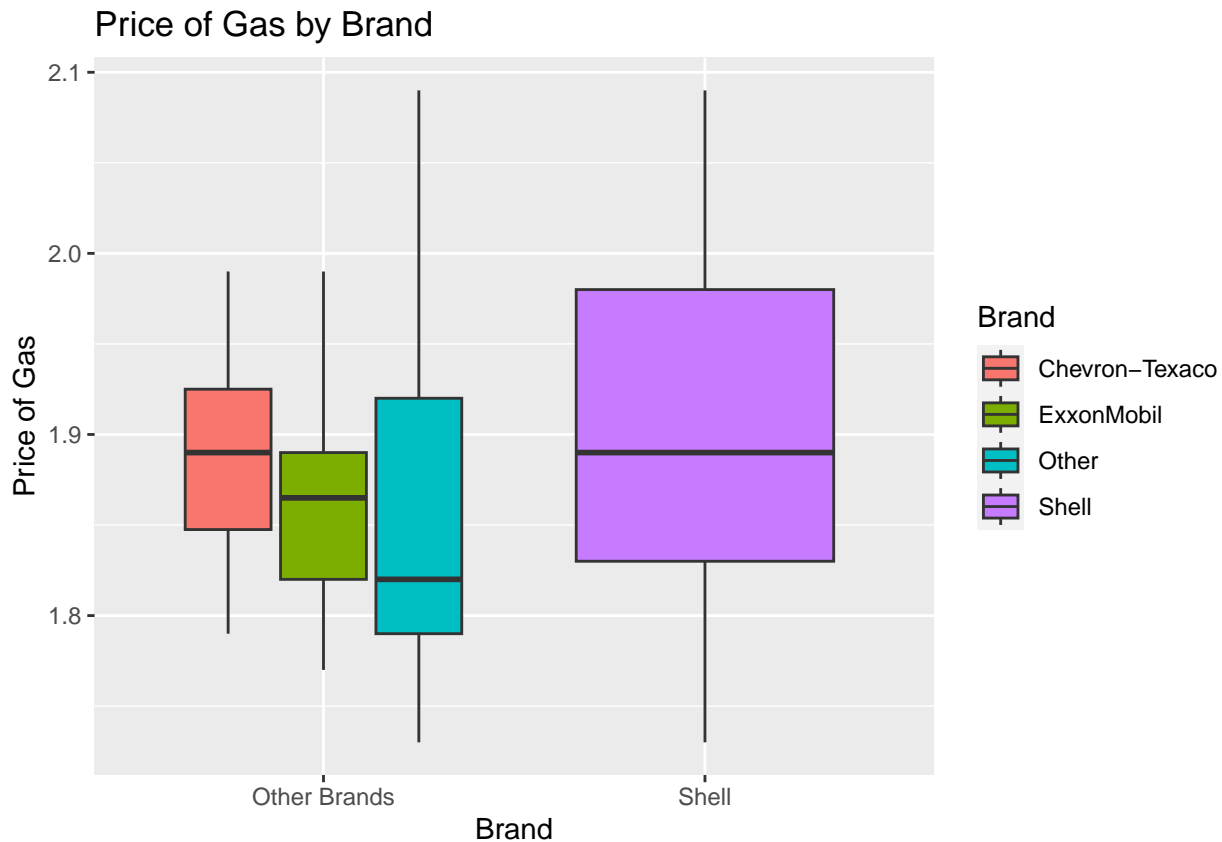

```
##       name       lower      upper level     method    estimate
## 1 diffmean -0.009795418 0.06456113  0.95 percentile 0.008389474
```

Claim - Shell charges more than all other non-Shell brands. Evidence - When looking at the 95 percent confidence interval of the difference between the mean price at Shell compared to competitors shows us a range between -0.0096 and 0.0656 with 95 percent confidence. This translates to around 0 to 6 cents more that Shell charges compared to competitors. Conclusion - The theory is supported by the data as the confidence interval tells us that Shell charges more than competitors from anywhere between -0.0096 to 0.0656 cents. This means that Shell charges anywhere from 0 to 6 cents more than any other gas station on average. This proves the theory that Shell charges more than all other non-Shell brands.

Question 2

Part A

The dataset was trimmed for all the sclass cars from the year 2011 and with the 63 AMG trim. When looking at the average mileage for these cars we use a bootstrapped 95 percent confidence interval to get an average mileage between 26300 to 31700 miles with 95 percent confidence. This means that for all the 2011 63 AMG cars hitting the used car market the average mileage is between 26300 to 31700 miles with 95 percent confidence.

Part B

The dataset was trimmed for all the sclass cars from the year 2014 and with the 550 trim. After this a new variable was created called isBlack to store whether a car's color was black or not. After this a bootstrap 95 percent confidence interval was calculated and we calculated a proportion of sclass cars painted black were between 0.4168 to 0.4528 percent with 95 percent confidence. This means that for all the sclasses from the

year 2014 and with the trim 550 the proportion of cars painted black was between 0.4168 to 0.4528 with 95 percent confidence.

Question 3

Part 1

1) Question - Does one show produce a higher mean Q1_Happy response among viewers?
2) Approach - For this question I first filtered the data so I could only view one show at a time for the two shows that I was interested in. After this I calculated 95 percent bootstrap confidence interval using bootstrapping, mean, mosaic resample, and, confint for Q1_Happy for both shows using the data from the file.
3) Results - The results show us that for the show, Living with Ed, we can say that the viewers were happy between the numbers 3.732 to 4.123, out of 5, with a 95 percent confidence. For the other show, My Name is Earl, we can say that the viewers were happy between the numbers 3.627 to 3.922, out of 5, with 95 percent confidence.
4) Conclusion - The results show us that most viewers of both shows answered the statement, this show made me feel happy, with an agree not a disagree or a strongly agree. Both shows had a similar audience happiness rating but if we had to pick a show with a higher mean Q1_Happy rating we would say the first show, Living with Ed. However when taking a look at both shows from the viewpoint of the TV Network we would group both these shows into the same category when it comes to the viewers happiness when watching. This means that neither show is greater than the other in this category.

Part 2

1) Question - Does one show produce a higher mean Q1_Annoyed response among viewers?
2) Approach - For this question I first filtered the data so I could only view one show at a time for the two shows that I was interested in. After this I calculated 95 percent bootstrap confidence interval using bootstrapping, mean, mosaic resample, and, confint for Q1_Annoyed for both shows using the data from the file.
3) Results - The results show us that for the show, The Biggest Loser, we can say that the viewers were annoyed between the numbers 1.862319 to 2.217, out of 5, with a 95 percent confidence. For the other show, The Apprentice: Los Angeles, we can say that the viewers were annoyed between the numbers 2.127 to 2.488, out of 5, with 95 percent confidence.
4) Conclusion - The results show us that the viewers of the first show, The Biggest Loser, answered the statement, this show maded me feel annoyed, with a strongly disagree to a disagree. The second show, The Apprentice: Los Angeles, answered the same statement with disagree. This tells us that viewers of the second show were more likely to be annoyed when watching compared to viewers of the first show answering our question telling us that The Apprentice: Los Angeles produced a higher mean Q1_Annoyed response. This can be important for the TV Network as they might want to know which show is more likely to rub viewers the wrong way and they can try to adjust that show for the future.

Part 3

1) Question - What proportion of Dancing with the Stars viewers can we expect to answer Q2_Confusing with a 4 or higher.
2) Approach - For this question I first filtered the data so I would only have data from the show Dancing with the stars. After this I created a new variable called Q2_Confusing4 which stored whether or not the response to Q2_Confusing was greater than or less than 4. After this I calculated a 95 percent bootstrap confidence interval using bootstrapping, prop, mosaic resample, and confint for the new variable Q2_Confusing4.
3) Results - The results show us that for the show Dancing with the Stars the proportion of people who answered Q2_Confusing with a score of 4 or more was somewhere between 0.0442 to 0.1160 with 95 percent confidence.

4) Conclusion - The results show us that the viewers of the Show Dancing with the Stars were not very confused when watching the show as the proportion of the viewers who were confused when watching the show was between 0.0442 and 0.1160 with 95 percent confidence. This can be important for the TV Network as they know that they have a show that doesn't confuse a significant majority of the viewers so they can use this show as a model for other shows were the audience answered with scores of 4 or higher for the question Q2_Confusing.

Question 4

1) Question - Is there a difference in the revenue ratio when there is no advertising used in the DMA compared to when there is advertising used?
2) Approach - For this question I first created two different data sets, one to represent the treatment, no advertising, and one to represent the control, paid advertising. This was done by using filter and looking at what adwords_pause was set to. After this I created to different bootstrap confidence intervals, one for the treatment and one for the control. This was done by using bootstrapping, mean, mosaic resample, and confint.
3) Results - The results show that for the treatment group where there was no advertising the revenue ratio was between 0.8647 to 0.9285 with 95 percent confidence. For the control group with the advertising the revenue ratio was between 0.9270 and 0.9713.
4) Conclusion - The results show that there is a significant difference in the confidence intervals when it comes the treatment with no advertising and the control with advertising. The difference is around 5 to 6 percent which is very significant. This means that the DMAs where there was no money spent on advertising during this experiments time period lost revenue at a greater rate than the DMAs where there was money spent on advertising. This tells the company EBay that even though they are a really big company they can't rely on going organic as they will lose revenue and they also have to pay for advertising on google.