Homework 3

Ryann Liu (rl35976)

2025-01-28

Link to my Github repo: https://github.com/ryannrliu/Homework-3-Scott

Problem 1: Gas Station Theories

Part A

1) Claim: Gas stations charge more if they lack direct competition in 2) sight.
2) Evidence: The difference in mean prices between both groups (No competition vs. Yes) lies between -0.06 and 0.01, with 95% confidence. The graphs also show relatively similar prices between each group.
3) Conclusion: This theory is unsupported by the data.


Do Competitors Affect Gas Price?
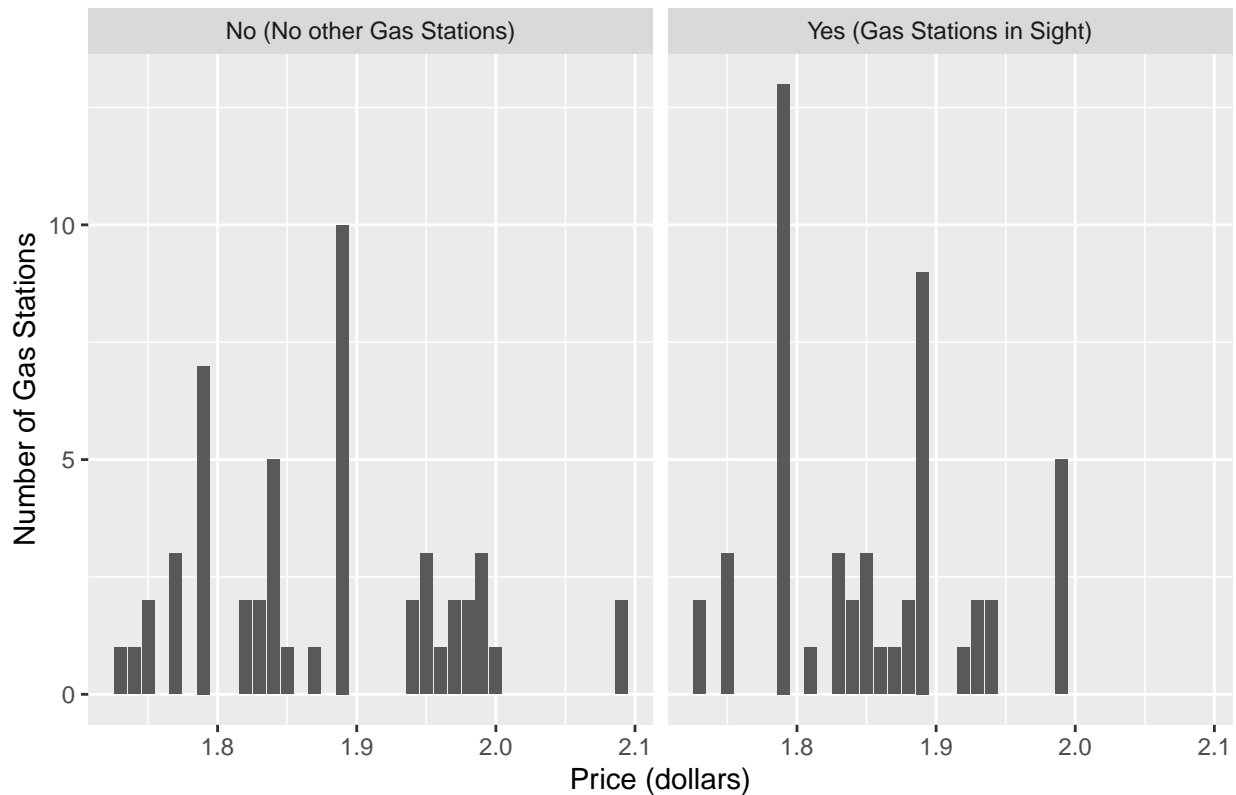
```
##       name       lower      upper level     method    estimate
## 1 diffmean -0.05394621 0.008305822  0.95 percentile -0.02348235
```

Part B

1) Claim: The richer the area, the higher the gas prices.
2) Evidence: From the graph, there is a positive correlation between income and gas price. So, as income increases, gas price also increases. The correlation coefficient between the two variables is 0.4, so there is a moderately strong positive correlation.
3) Conclusion: The theory is supported by the data.

Does Income Affect Gas Price?
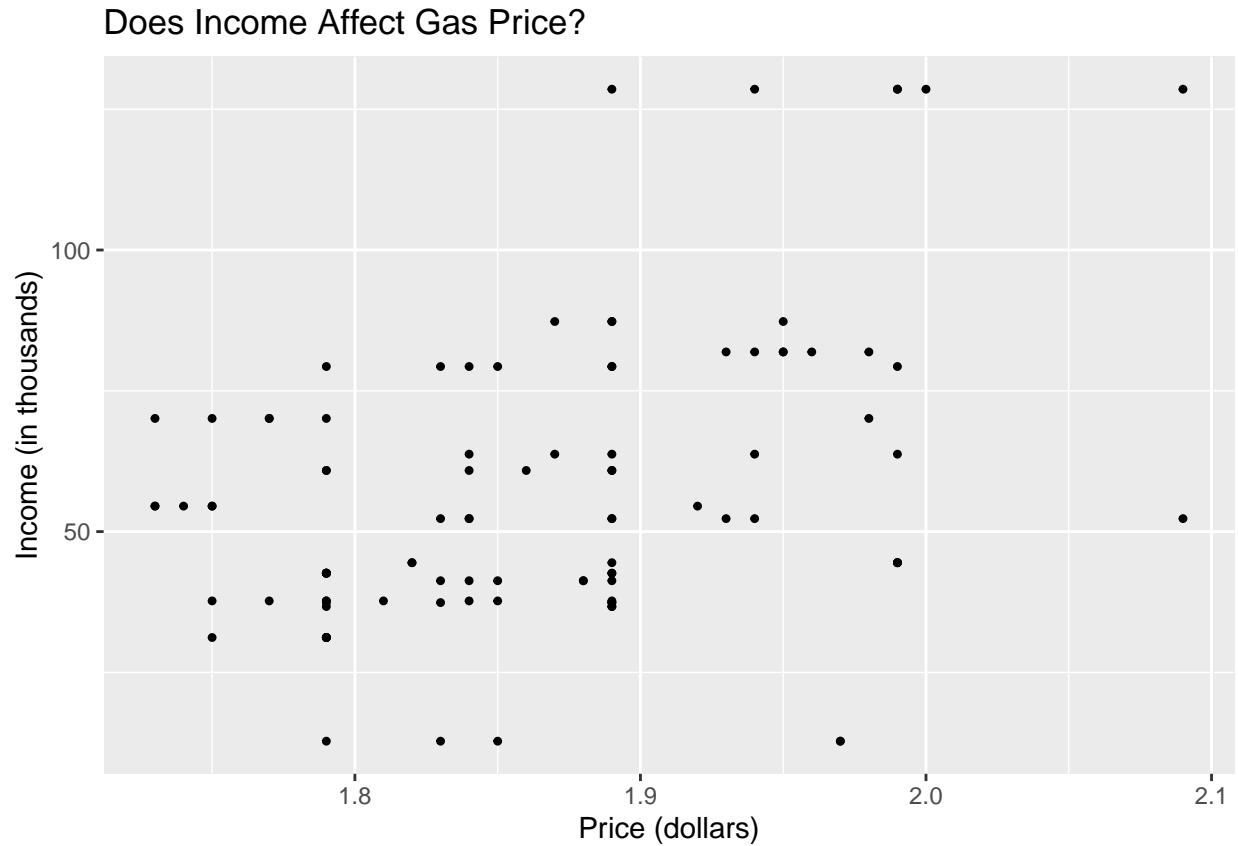
Part C

1) Claim: Gas stations at stoplights charge more.
2) Evidence: The difference in mean gas price between each group(By a stoplight vs. not by a light) lies between -0.04 and 0.03, with 95% confidence. The graphs also show relatively similar prices between each group.
3) Conclusion: This theory is unsupported by the data.

Do Gas Stations at Stoplights Charge More?
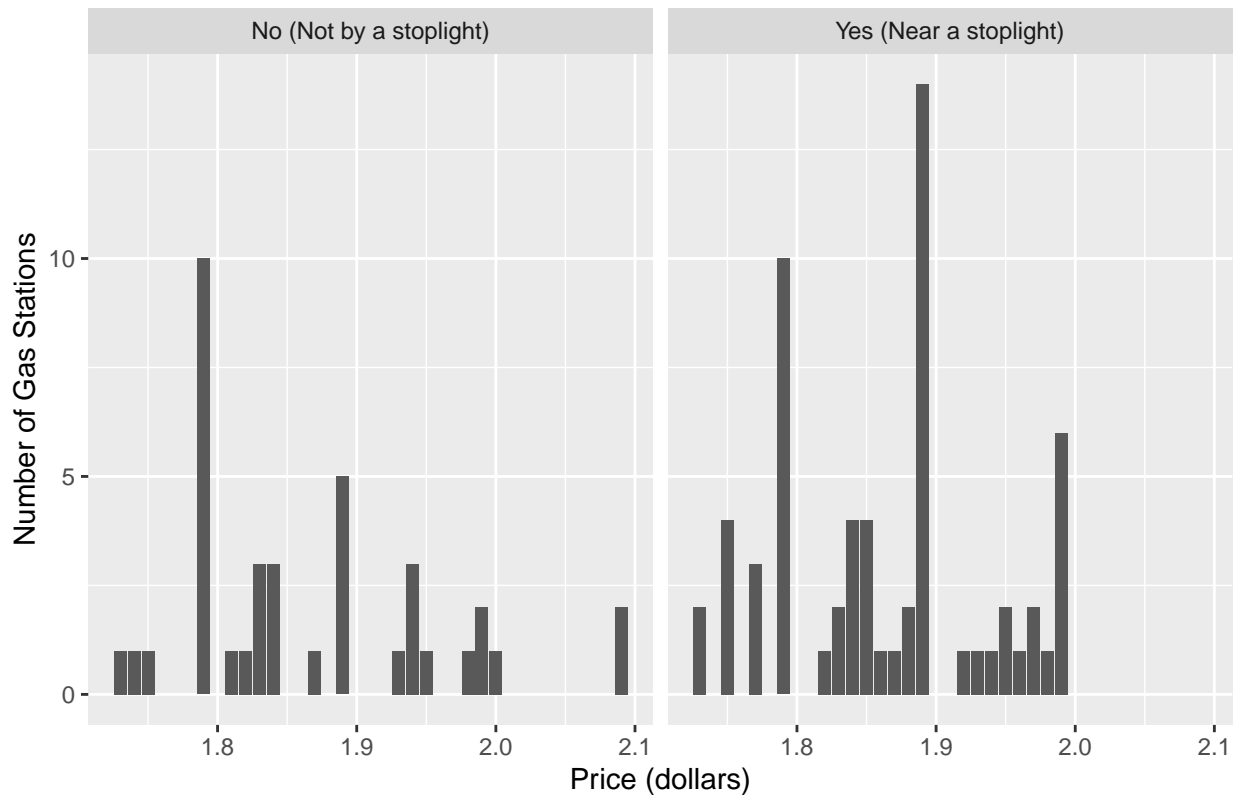

```
##        name       lower      upper level     method      estimate
## 1 diffmean -0.03761887 0.03109085  0.95 percentile -0.003299916
```

Part D

1) Claim: Gas stations with direct highway access charge more.
2) Evidence: The difference in mean gas price between the two groups (highway access vs. no direct highway access) lies between 0.01 and 0.08, with 95% confidence. The graphs show that gas stations without direct highway access charge more.
3) Conclusion: The theory is unsupported by the data, as it appears that gas stations without direct highway access charge more.

Highway Access vs. Gas Price
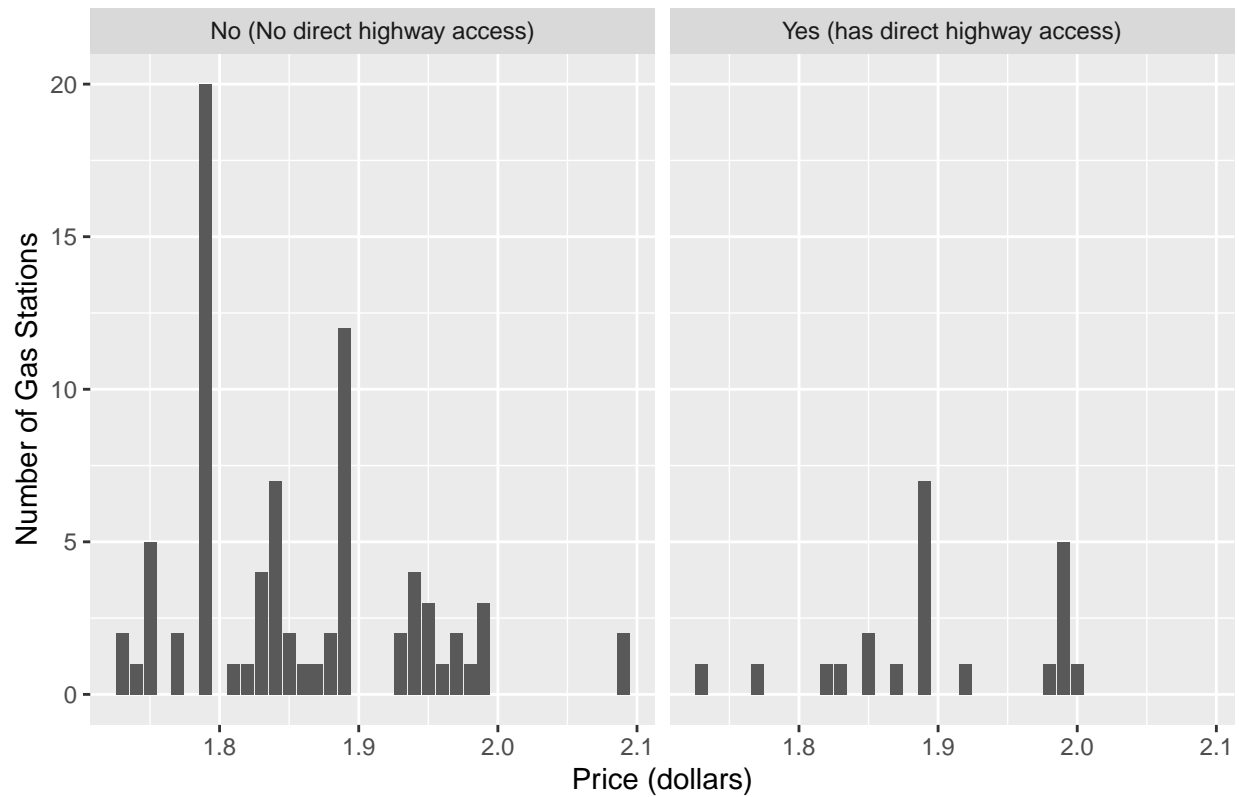

```
##      name       lower      upper level    method   estimate
## 1 diffmean 0.008984123 0.08150036  0.95 percentile 0.0456962
```

Part E

1) Claim: Shell charges more than all other non-Shell brands.
2) Evidence: The difference in mean gas price between the two groups (Shell vs. Non-Shell) lies between -0.01 and 0.07, with 95% confidence. The graphs also show no significant difference in price between Shell and non-Shell brands.
3) Conclusion: The theory is unsupported by the data.

Shell vs. Non–Shell Gas Prices
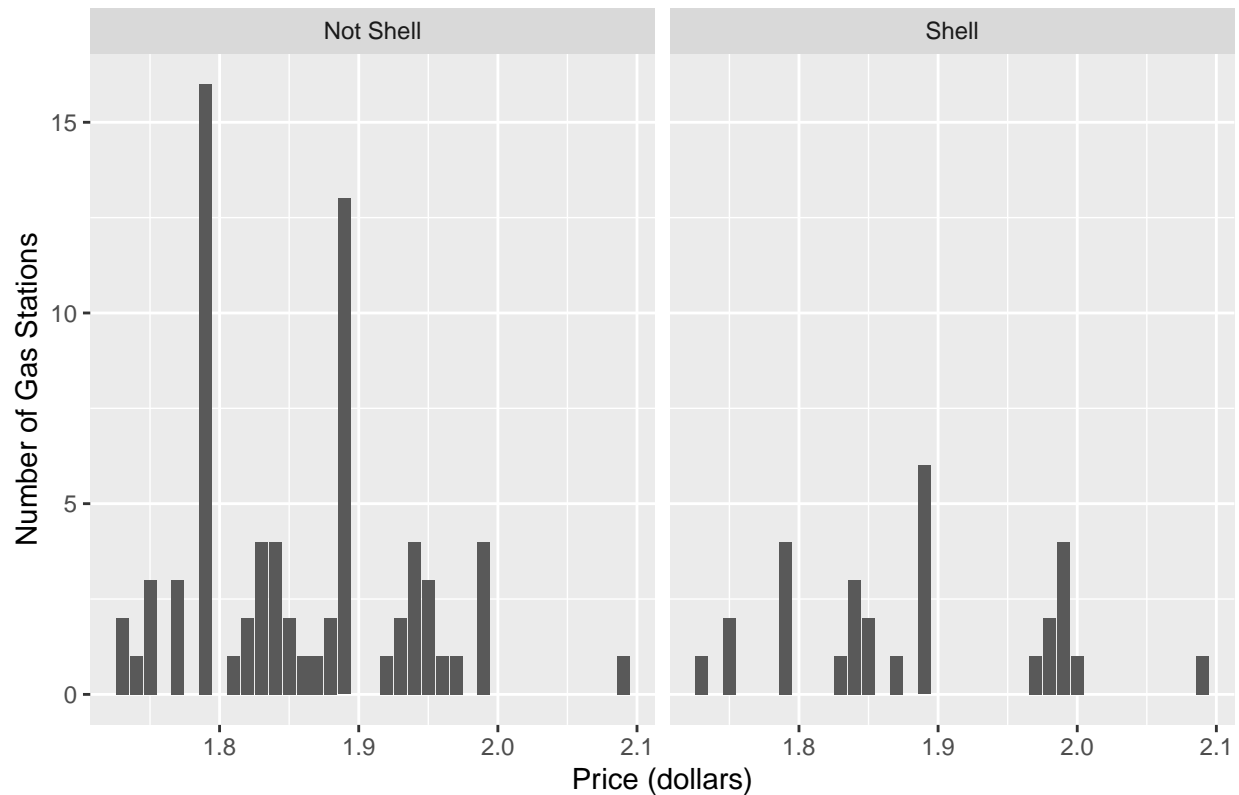

```
##       name       lower      upper level    method   estimate
## 1 diffmean -0.009344376 0.06461563  0.95 percentile 0.02740421
```

Problem 2: Mercedes S-Class Data

For each of these questions, I created a sample of each bootstrap, a confidence interval, and a histogram showing the distribution of the calculated bootstrap.

Part A

The bootstrap suggests that the true population parameter (average mileage of 2011 S-Class 63 AMGs that were used) lies within the interval 26,907 to 34,802, with 95% confidence.

```
##        CPO     Used
## 1 30599.77 31854.64
## 2 26626.93 29707.34
## 3 29097.95 31309.16
## 4 25678.76 32265.47
## 5 26457.51 30200.02
## 6 24413.96 35270.87
```

```
##   name    lower    upper level     method estimate
## 1  CPO 22838.00 30190.62  0.95 percentile 26495.19
## 2 Used 26886.86 34708.81  0.95 percentile 30763.56
```
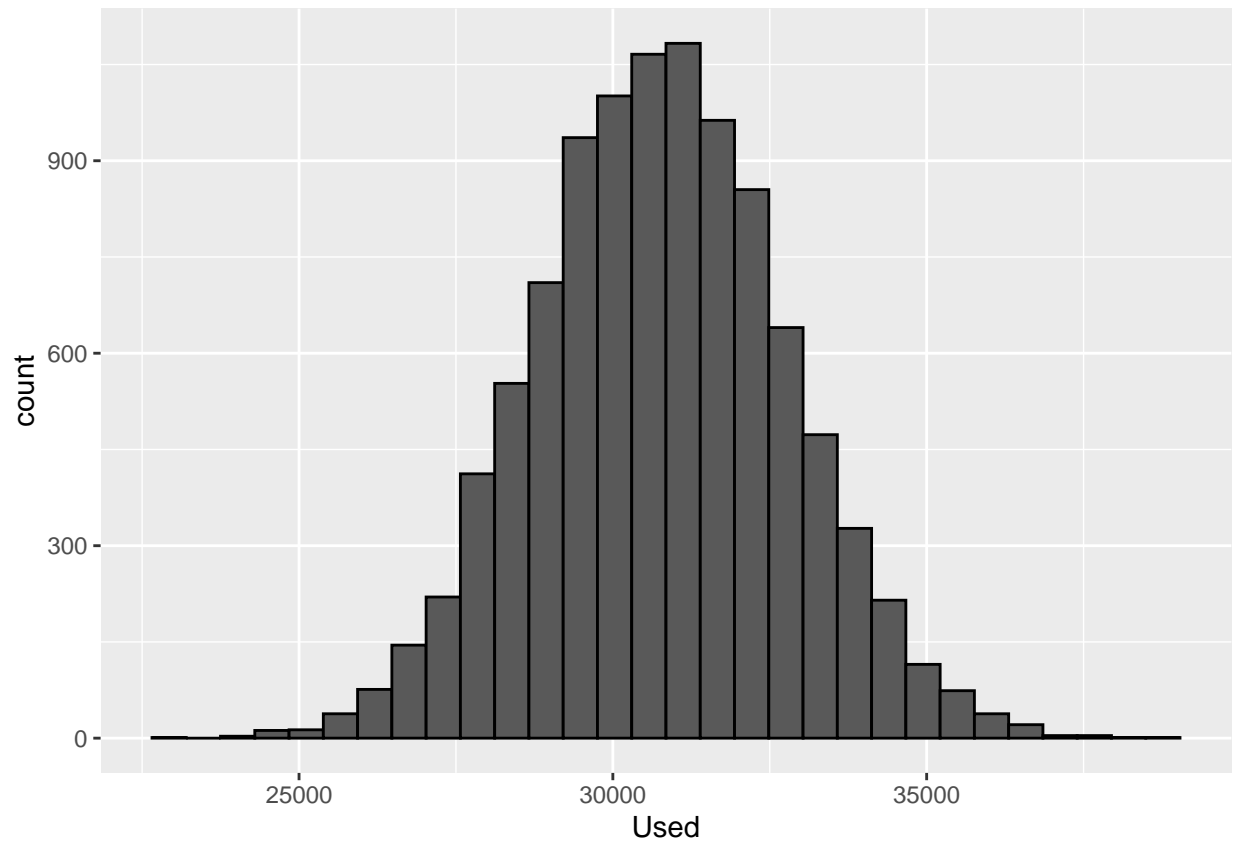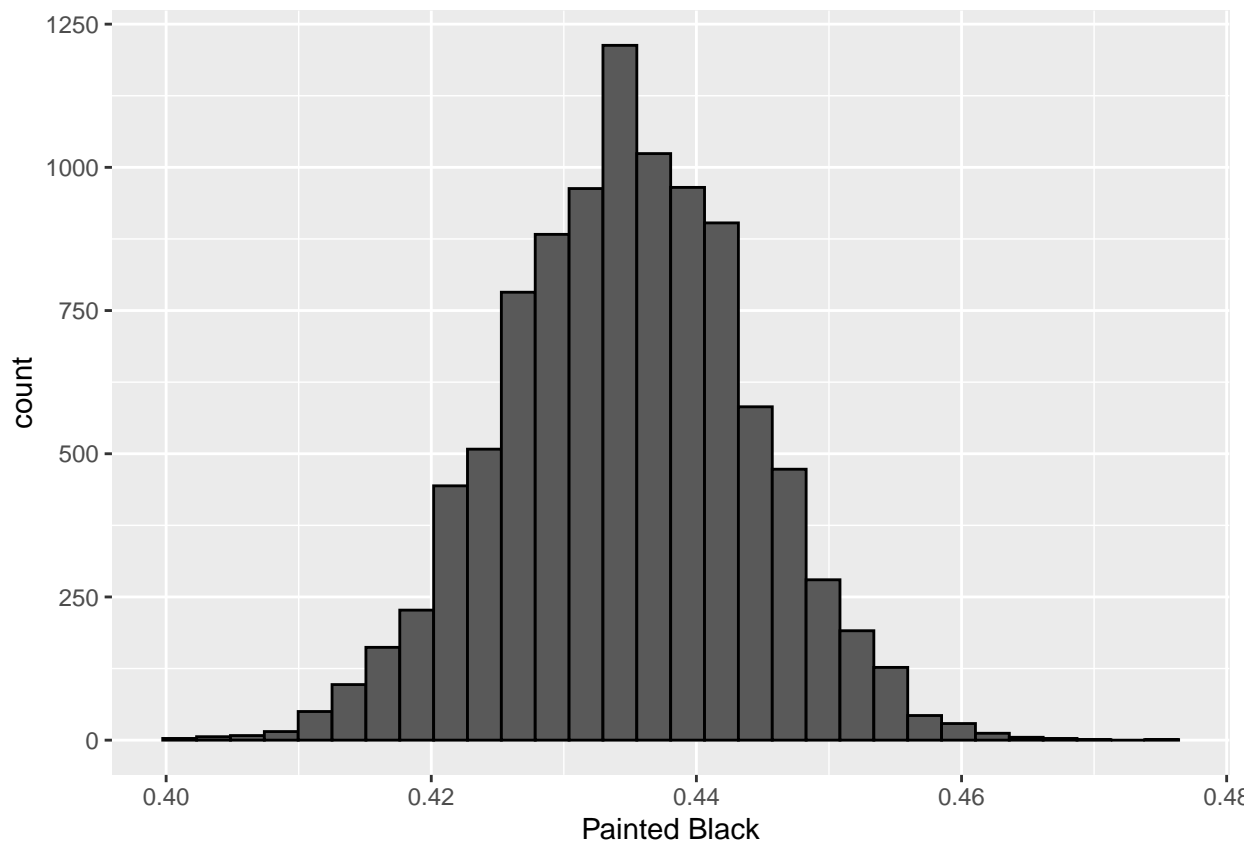
Part B

The bootstrap suggests that the true population parameter (2014 S-Class 550s that were painted black) lies within the interval 0.416 to 0.452, with 95% confidence.

```
##   prop_TRUE
## 1 0.4254067
## 2 0.4395985
## 3 0.4326757
## 4 0.4427137
## 5 0.4569055
## 6 0.4337141
```

```
##        name     lower     upper level     method  estimate
## 1 prop_TRUE 0.4164071 0.4527518  0.95 percentile 0.4347525
```

Problem 3: NBC Shows

Part A

1) Question: Does either the show "Living with Ed" or "My Name is Earl" generally produce happier viewers?
2) Approach: I used the bootstrap method to calculate the difference in means, and constructed a 95% confidence interval by re-sampling 10,000 times.

3) Results: The difference of mean viewer response to the "Happy" question between the two shows lies somewhere between -0.39 and 0.10, with 95% confidence.
4) Conclusion: Because the estimated mean difference is relatively small (< 0.5), both shows consistently produce the same average response to the "Happy" question. So, both shows make people equally happy.

```
##      diffmean
## 1 -0.24082091
## 2 -0.08013504
## 3 -0.09234015
## 4 -0.16325212
## 5 -0.08193847
## 6  0.02376120
```

```
##        name      lower      upper level     method    estimate
```

```
## 1 diffmean -0.4004568 0.1012461  0.95 percentile -0.1490515
```
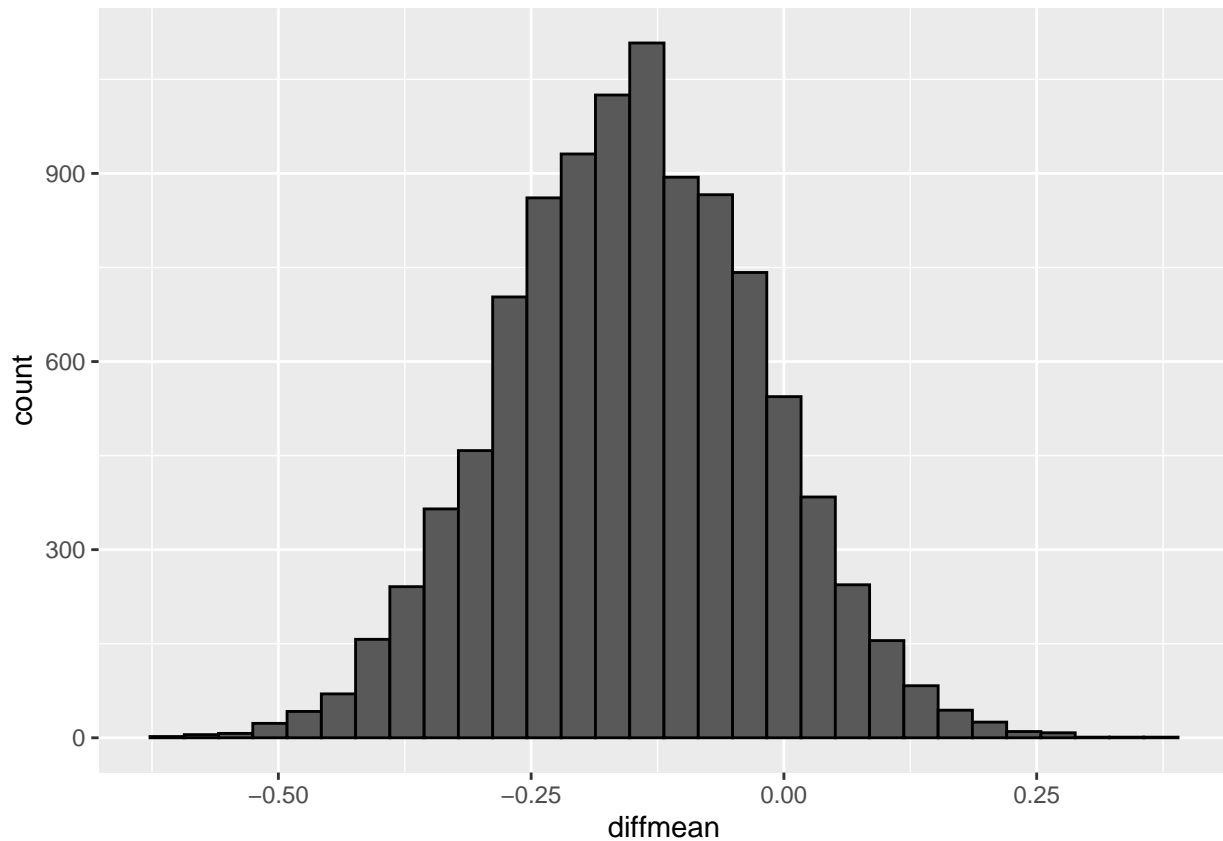

Part B

1) Question: Does either the show "The Biggest Loser" or "The Apprentice: Los Angeles" produce more annoyed viewers than the other?
2) Approach: I used the bootstrap method to calculate the difference in means, and constructed a 95% confidence interval by re-sampling 10,000 times.
3) Results: The difference of mean viewer response to the "Annoyed" question between the two shows lies somewhere between -0.52 and -0.02, with 95% confidence.
4) Conclusion: Because the estimated mean difference is relatively small (< 0.5), both shows consistently produce the same average response to the "Annoyed" question. So, both shows make people equally annoyed.

```
##      diffmean
## 1 -0.2314622
## 2 -0.3968875
## 3 -0.3067005
## 4 -0.3510421
## 5 -0.1649826
## 6 -0.3467515


##      name      lower      upper level     method  estimate
## 1 diffmean -0.521353 -0.01363181  0.95 percentile -0.270997
```
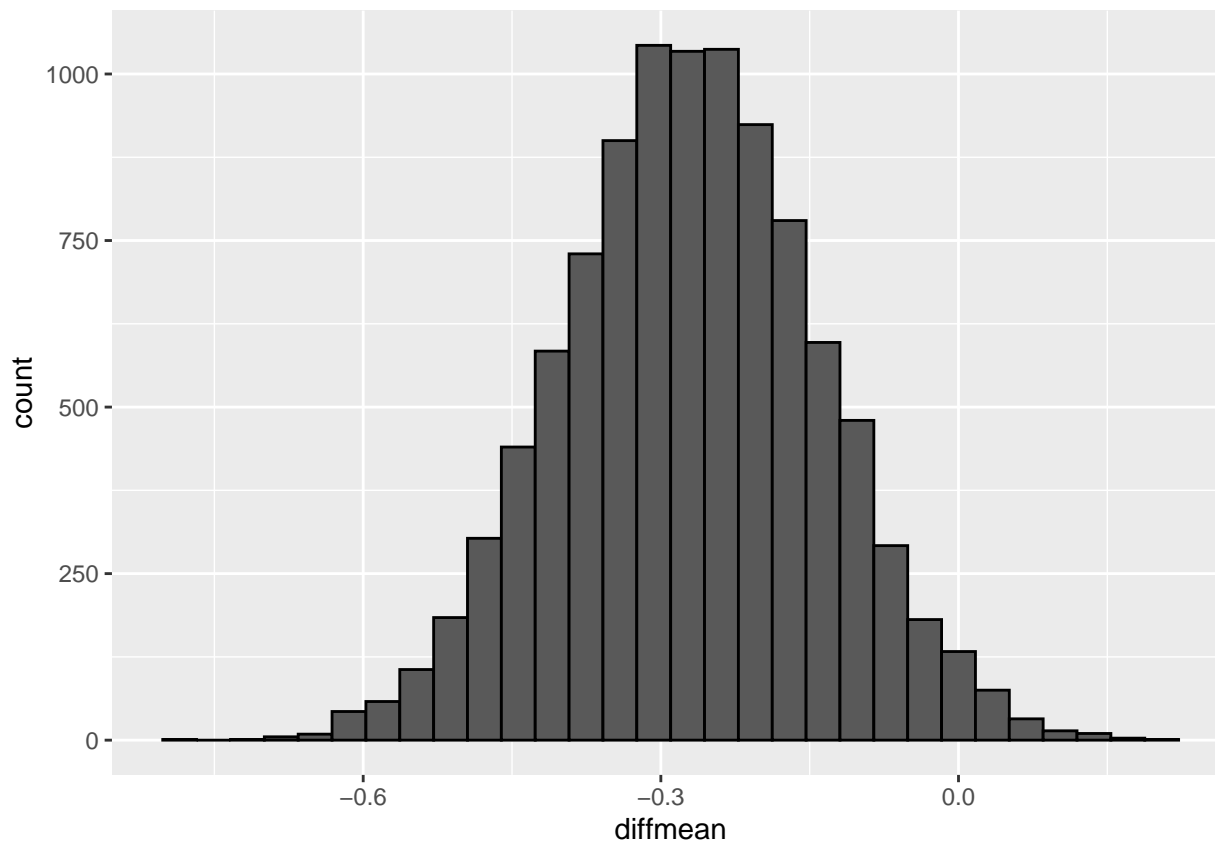
Part C

1) Question: What proportion of the population (American TV watchers) should be expected to think "Dancing with the Stars" is confusing (have a response greater than or equal to 4 to the "Confusing" question)?

2) Approach: I used the bootstrap method to calculate the population proportion, and constructed a 95% confidence interval by re-sampling 10,000 times.

3) Results: The proportion of the population that is estimated to respond that "Dancing with the Stars" is confusing lies between 0.03 and 0.11, with 95% confidence.

4) Conclusion: The estimated percentage of people to respond that "Dancing with the Stars" is confusing is very low, showing that not many people are expected to find the show confusing.

```
##    prop_TRUE
## 1 0.05524862
## 2 0.07182320
## 3 0.12707182
## 4 0.06629834
## 5 0.09392265
## 6 0.09392265
```

```
##       name      lower     upper level     method   estimate
## 1 prop_TRUE 0.03867403 0.1160221  0.95 percentile 0.07734807
```
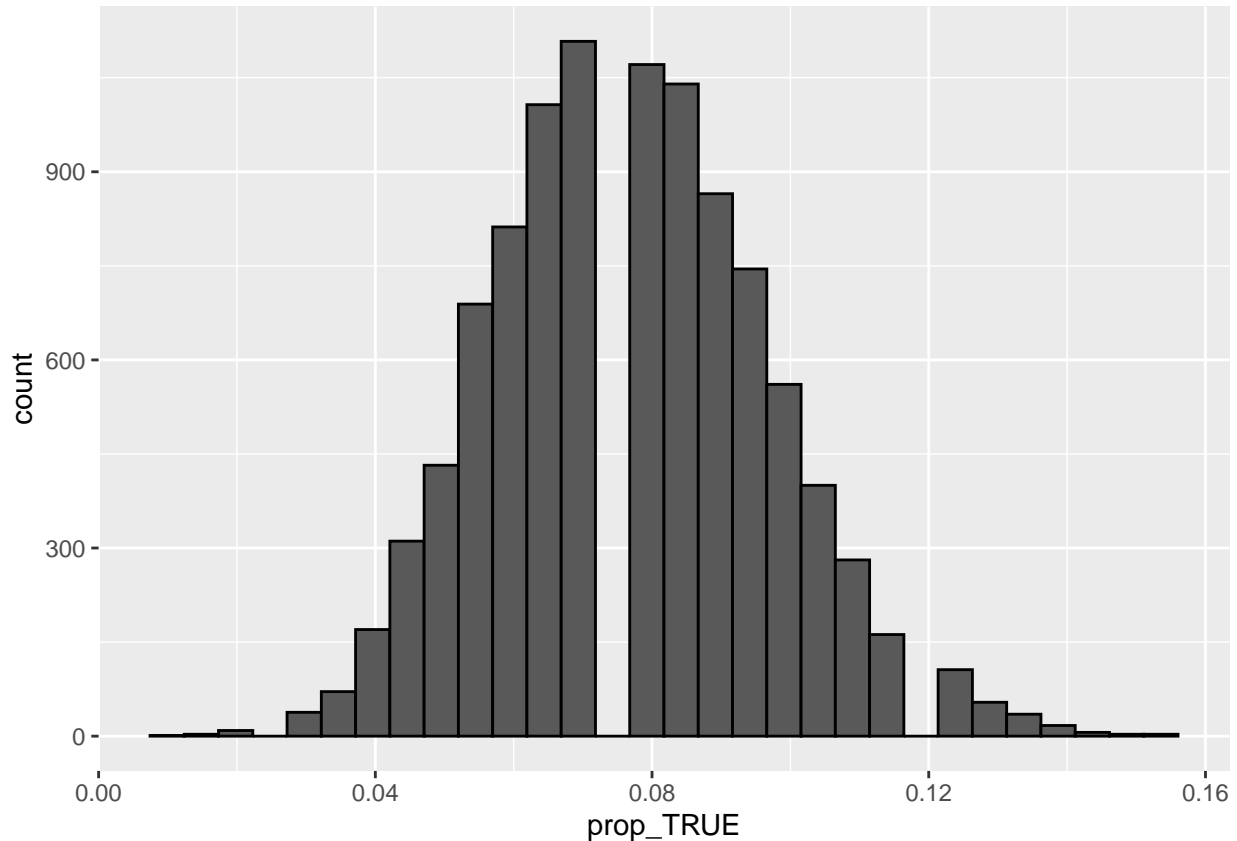
Problem 4: Ebay

1) Question: What is the difference in revenue proportion between the treatment and control DMAs, and does the difference support the idea that paid search advertising on Google creates extra revenue for EBay?
2) Approach: I used the bootstrap method to calculate the difference in mean revenue ratio, and constructed a 95% confidence interval by re-sampling 10,000 times.
3) Results: The difference in mean ratios between the control and treatment group lies somewhere between -0.091 and -0.013, with 95% confidence. The estimated difference in average revenue ratio between the groups is -0.05 (or a 5% difference).
4) Conclusion: The average revenue ratio consistently decreases more for the first group over the other. From this, we can conclude that certain paid search advertising on Google could create extra revenue for EBay.

```
##      diffmean
## 1 -0.06049145
## 2 -0.08495854
## 3 -0.06368247
## 4 -0.05498534
## 5 -0.03893333
## 6 -0.04089150
```

```
##       name       lower       upper level     method    estimate
## 1 diffmean -0.09055485 -0.01325074  0.95 percentile -0.05217532
```