SDS 315 HW 3

https://github.com/sophiayang5/SDS-315-HW-3

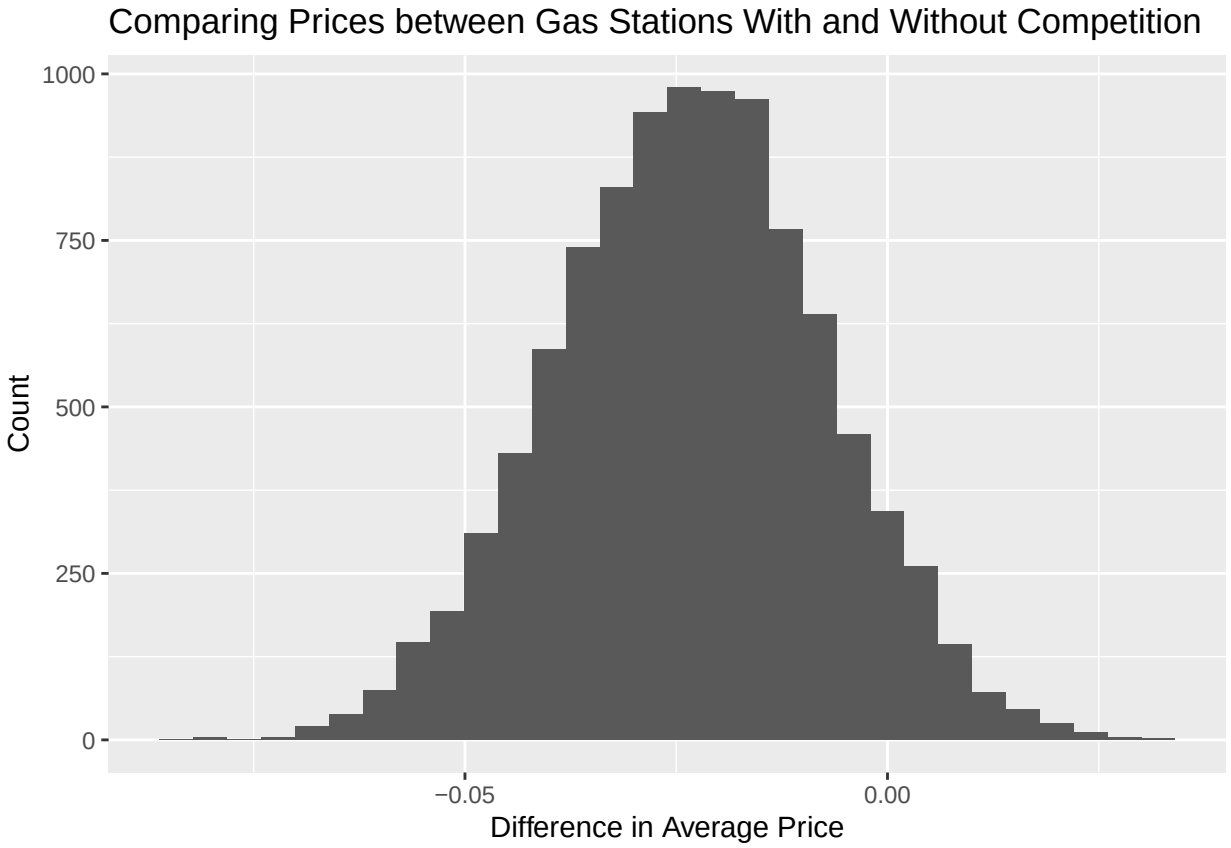
Sophia Yang (sy23928)

2025-02-08

Contents

Problem 1

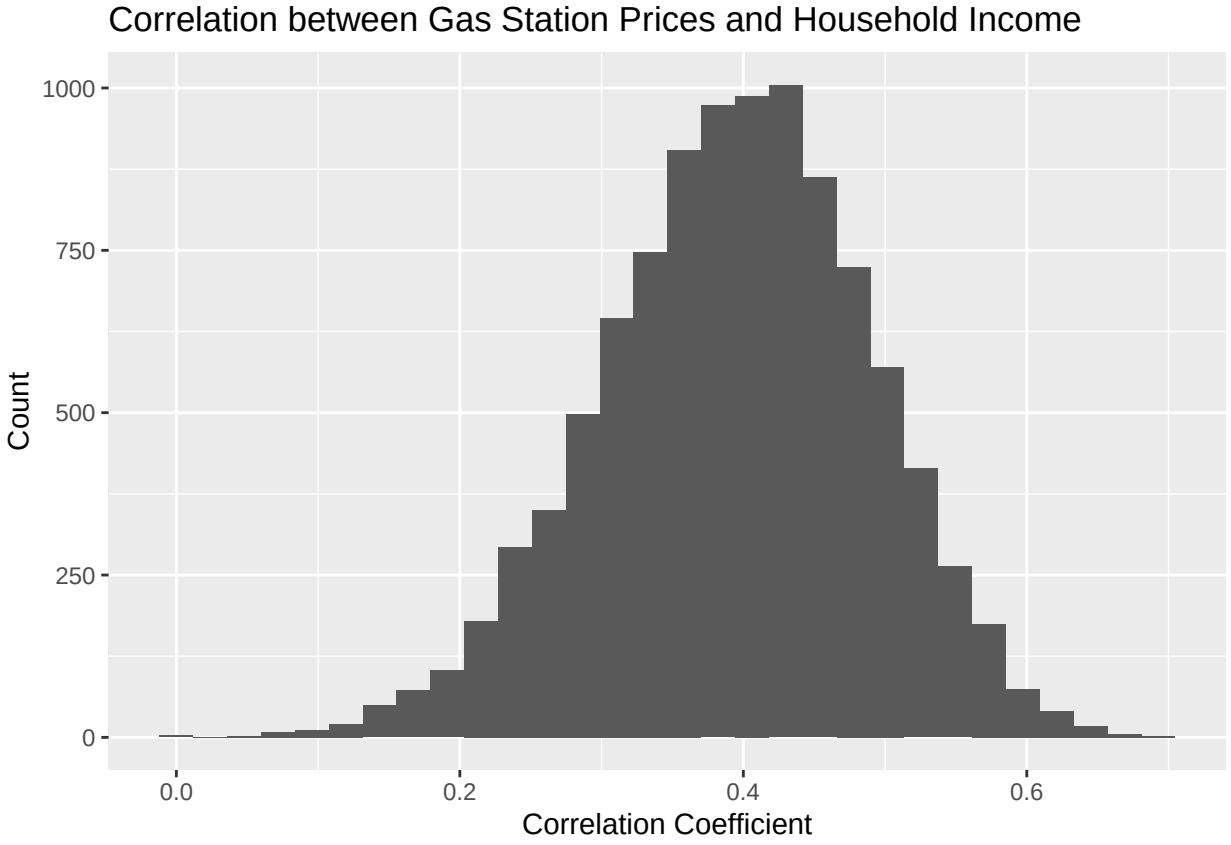
Part A.




```
##       name       lower       upper level     method    estimate
## 1 diffmean -0.05506388 0.007167577  0.95 percentile -0.02348235
```

The claim was that gas stations charge more if they lack direct competition in sight. However, the bootstrapped histogram of the difference in means between gas stations with and without direct competition, as well as the 95% confidence interval, show that no difference is likely. More specifically, the difference in price between gas stations with and without direct competition is somewhere between -0.055 and 0.008 (which contains 0), with 95% confidence. Thus, the claim is not supported.

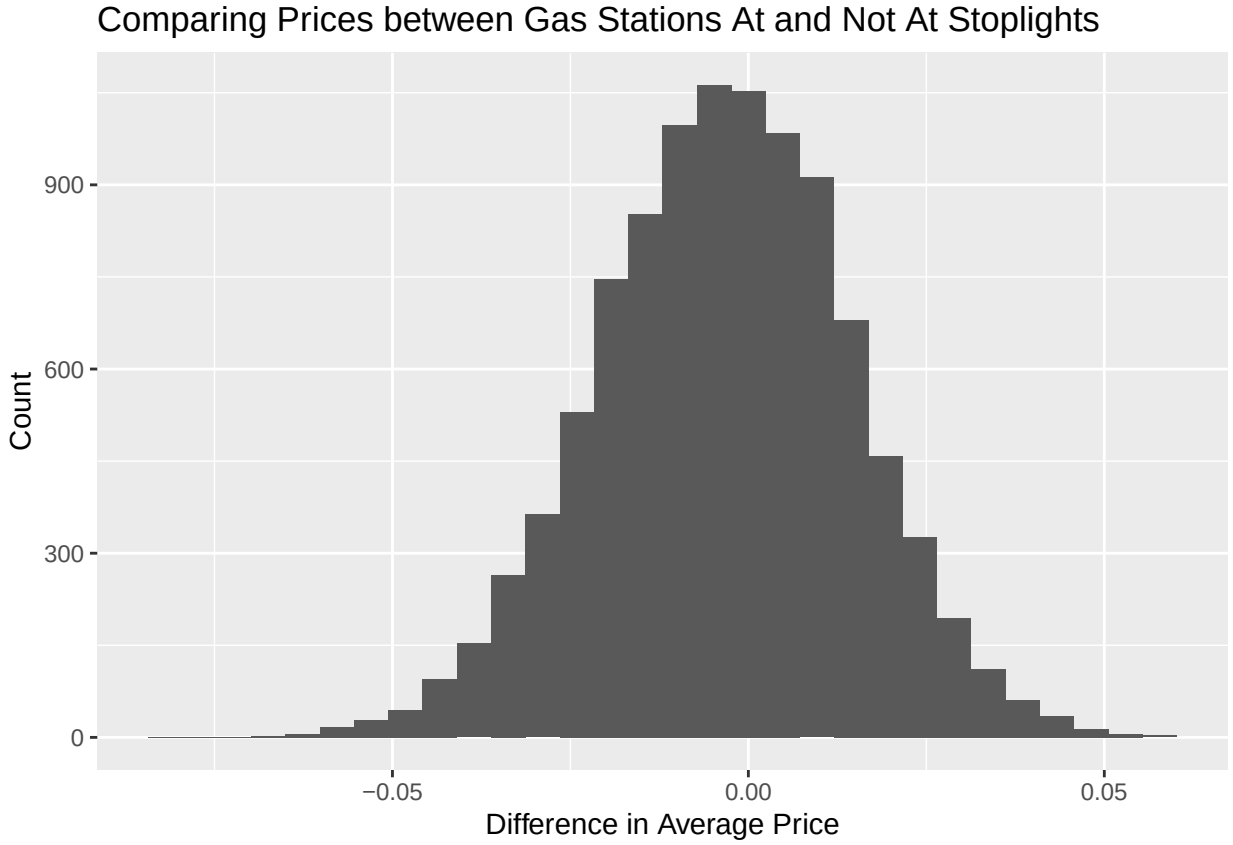
Part B.




```
##   name    lower     upper level     method  estimate
## 1  cor 0.200088 0.5695352  0.95 percentile 0.3961546
```

The claim was that gas stations have higher gas prices when the surrounding median household income is higher. The bootstrapped histogram of the correlation coefficients between gas stations' prices and the surrounding median household income, as well as the 95% confidence interval, show that there is likely not much of a correlation between higher gas prices and household income. More specifically, the correlation coefficient in gas stations' prices and household income is somewhere between 0.120 and 0.569, with 95% confidence. Thus, the claim is not strongly supported.

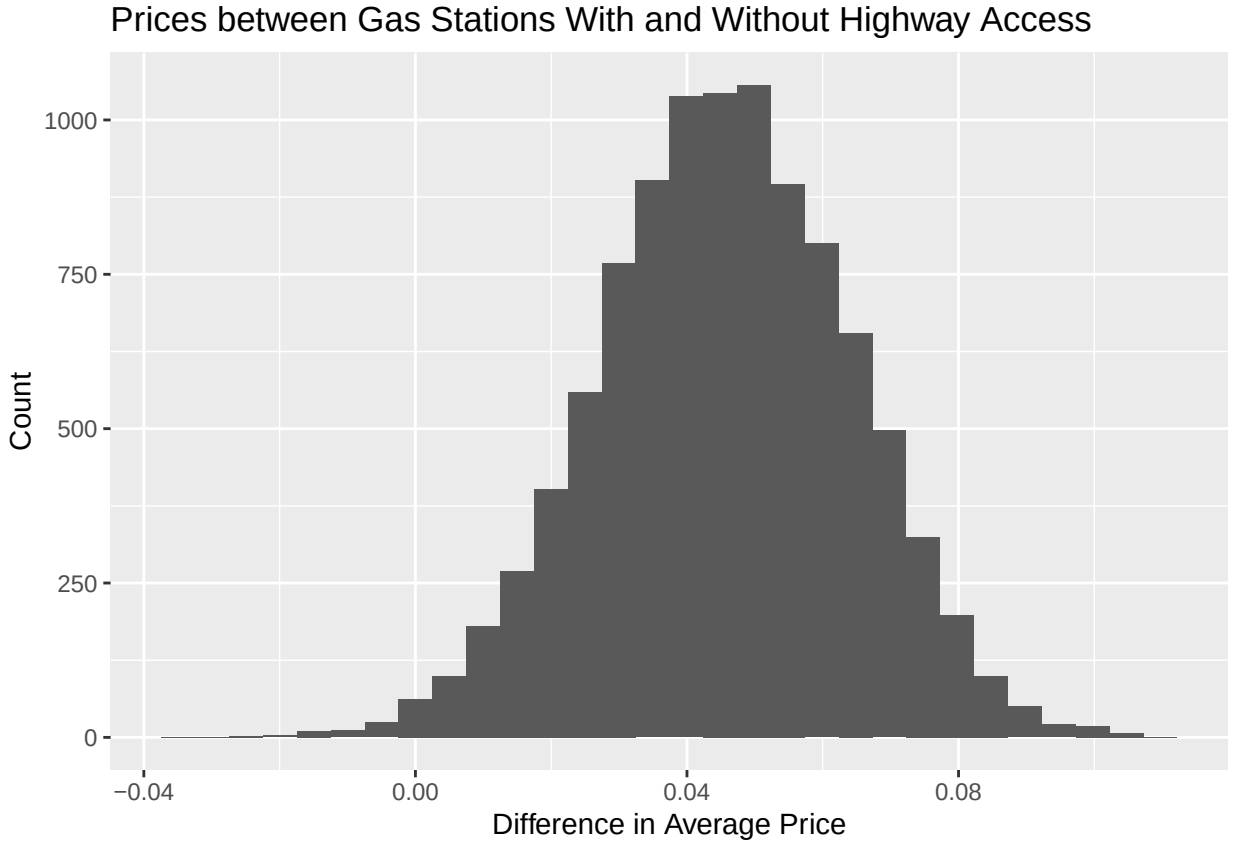
Part C.




```
##       name       lower      upper level     method     estimate
## 1 diffmean -0.03892662 0.03062433  0.95 percentile -0.003299916
```

The claim was that gas stations at stoplights charge more than gas stations not at stoplights. However, the bootstrapped histogram of the difference in means between gas stations at and not at stoplights, as well as the 95% confidence interval, show that no difference is likely. More specifically, the difference in price between gas stations at and not at stoplights is somewhere between -0.039 and 0.030 (which contains 0), with 95% confidence. Thus, the claim is not supported.

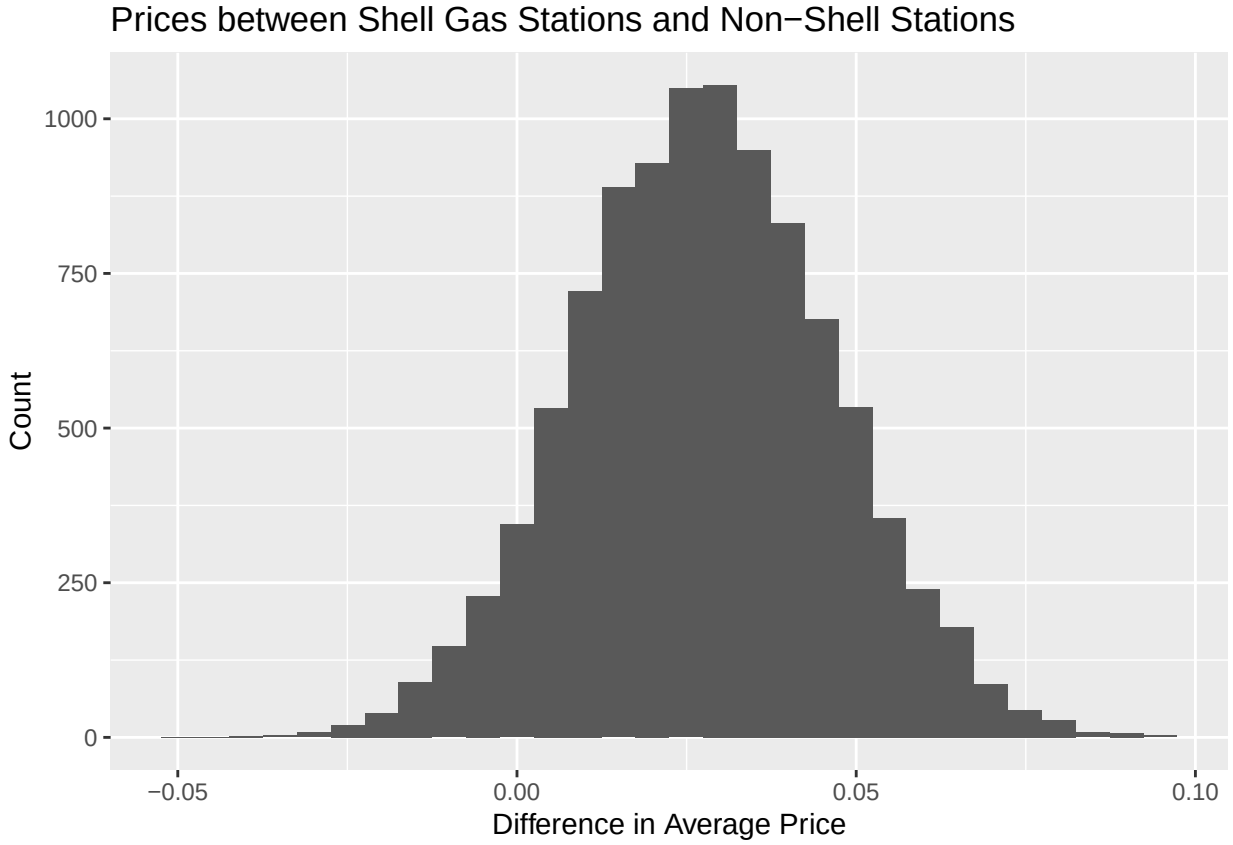
Part D.




```
##       name       lower     upper level     method  estimate
## 1 diffmean 0.008641723 0.0807914  0.95 percentile 0.0456962
```

The claim was that gas stations with direct highway access charge more than gas stations without direct highway access. The bootstrapped histogram of the difference in means between gas stations with and without direct highway access, as well as the 95% confidence interval, show that a difference is likely. More specifically, the difference in price between gas stations with and without direct highway access is somewhere between 0.009 and 0.081 (which does not contain 0), with 95% confidence. Thus, the claim is supported.

Part E.

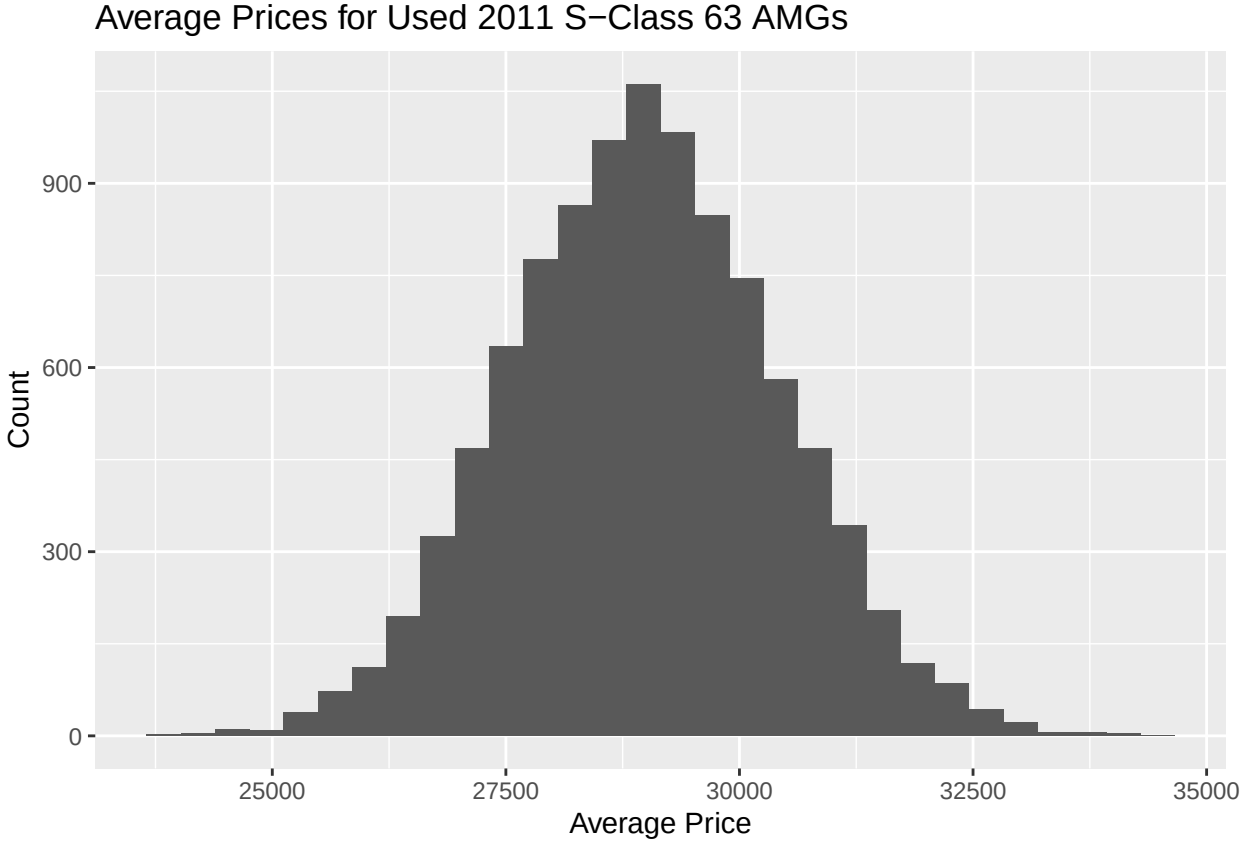



```
##       name        lower      upper level     method   estimate
## 1 diffmean -0.009128443 0.06503172  0.95 percentile 0.02740421
```

The claim was that Shell charges more than all other non-Shell brands. However, the bootstrapped histogram of the difference in means between Shell gas stations and non-Shell gas stations, as well as the 95% confidence interval, show that no difference is likely. More specifically, the difference in price between Shell gas stations and non-Shell gas stations is somewhere between -0.009 and 0.065 (which contains 0), with 95% confidence. Thus, the claim is not supported.

Problem 2

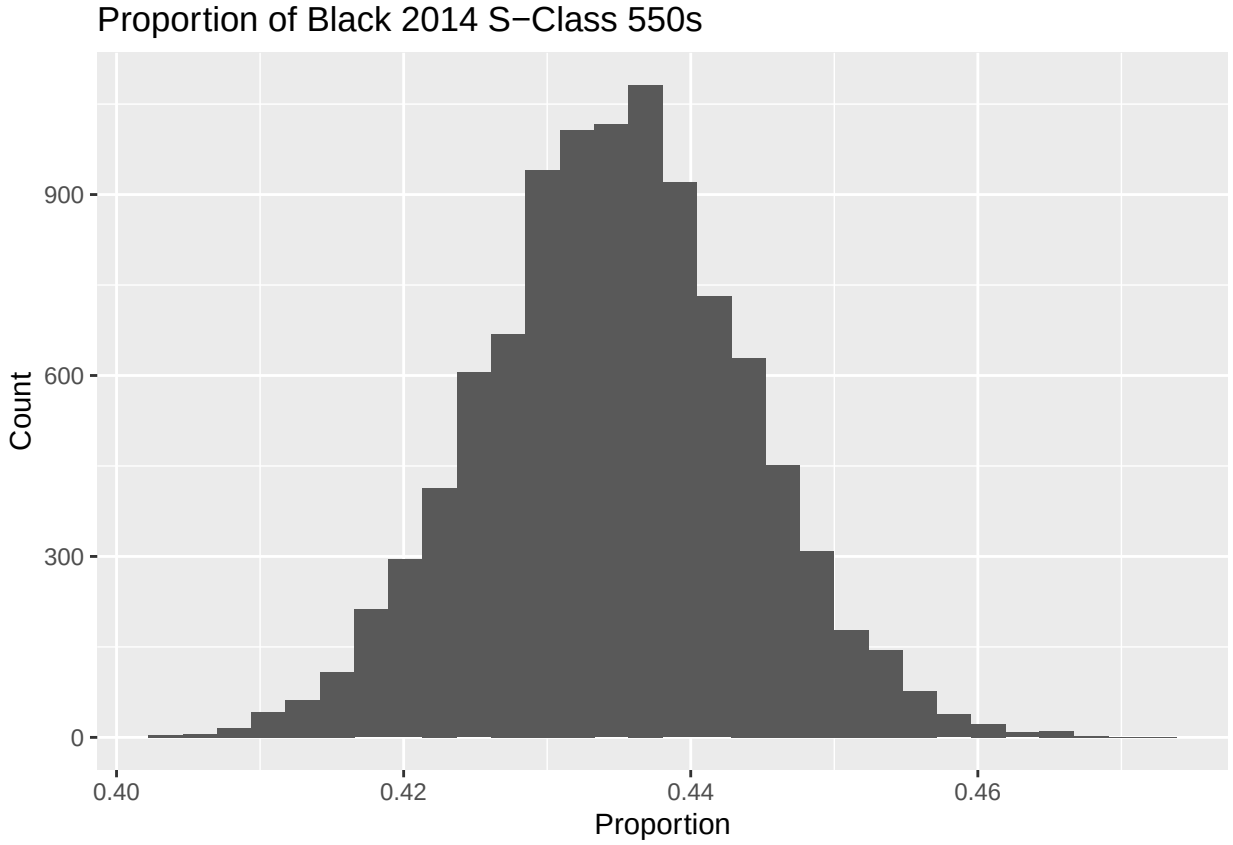
Part A.




```
##   name    lower    upper level     method estimate
## 1 mean 26232.64 31800.62  0.95 percentile 28997.34
```

The average mileage of 2011 S-Class 63 AMGs in the used-car market lies somewhere between $26,326.79 and $31,797.17, with 95% confidence.

Part B.

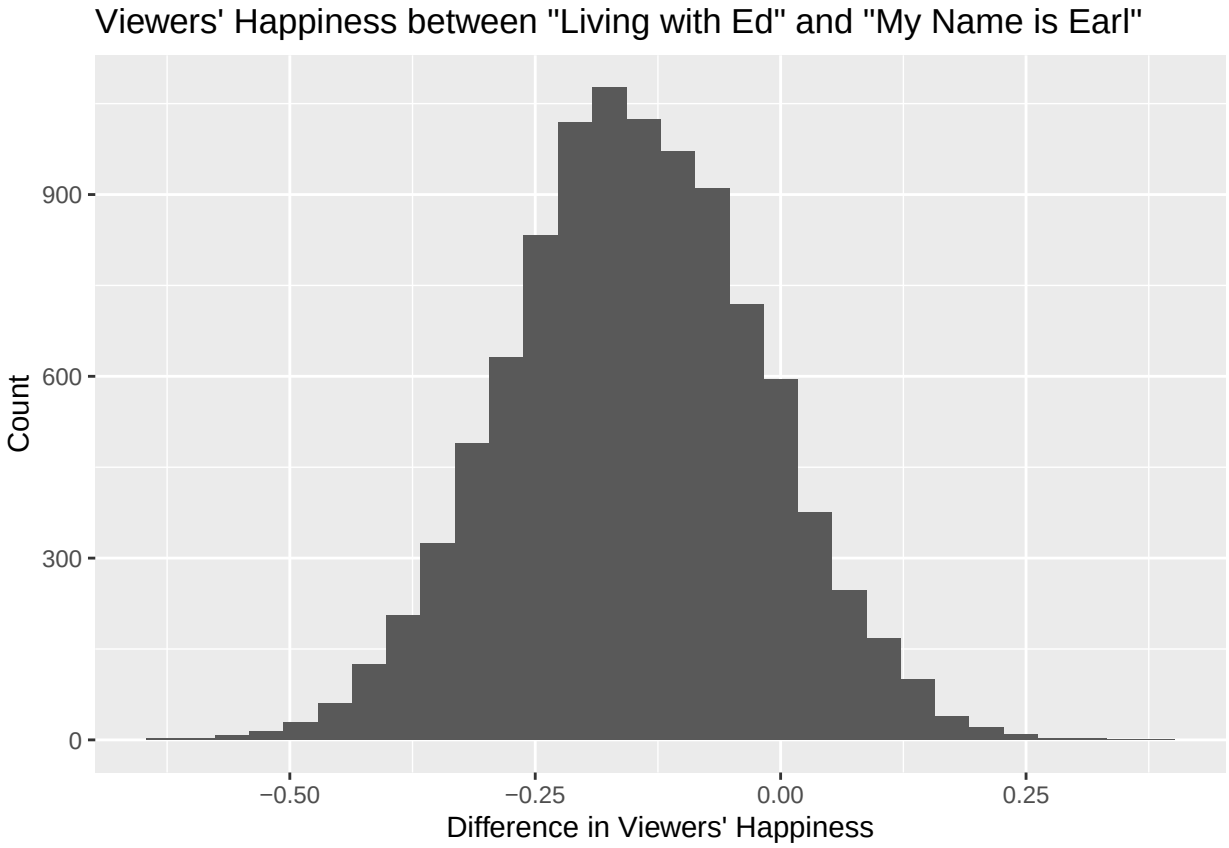



```
##        name     lower    upper level     method  estimate
## 1 prop_TRUE 0.4167532 0.453098  0.95 percentile 0.4347525
```

The proportion of black 2014 S-Class 550s in the used-car market lies somewhere between 0.417 and 0.453, with 95% confidence.

Problem 3

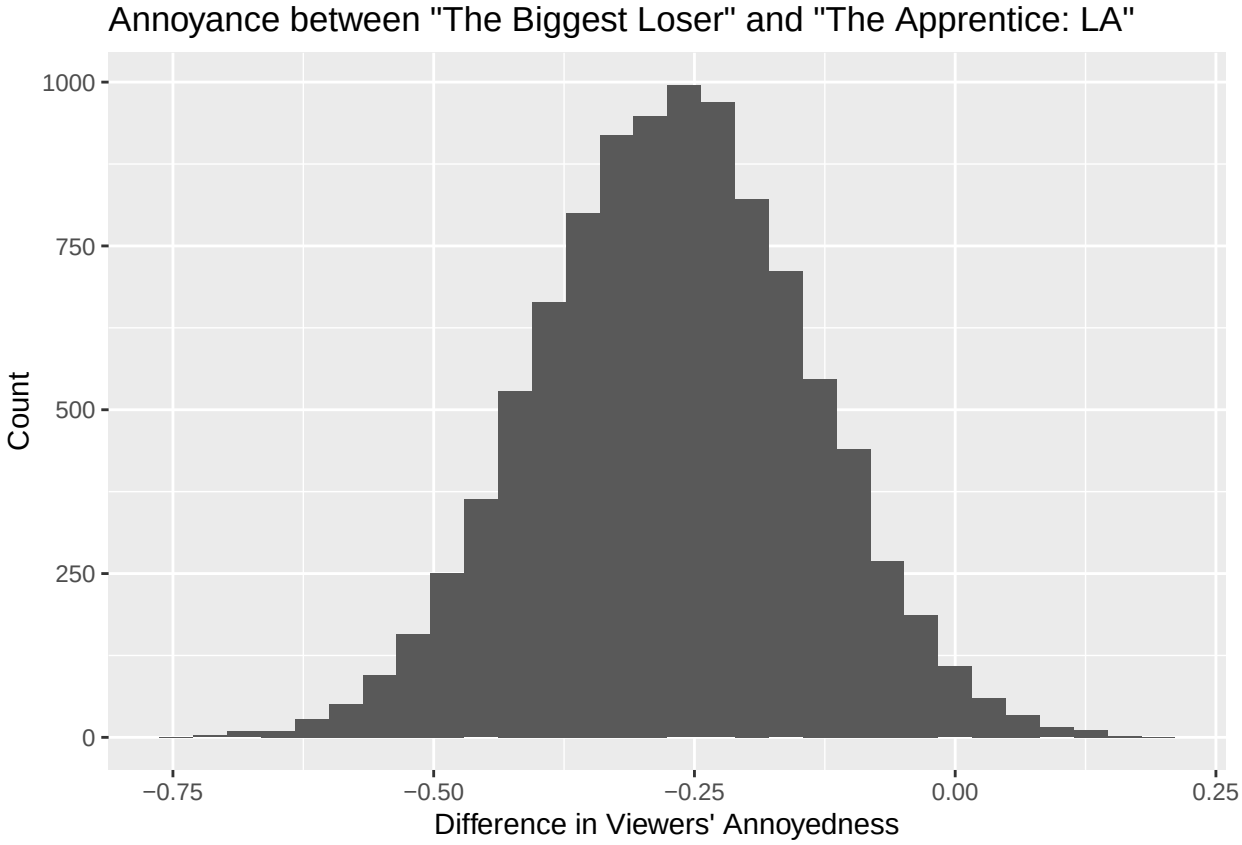
Part A.




```
##       name      lower     upper level     method   estimate
## 1 diffmean -0.3991679 0.1037121  0.95 percentile -0.1490515
```

The question was whether "Living with Ed" or "My Name is Earl" consistently produces a higher mean Q1_Happy response (on a scale of 1-5, with 5 being very happy) among viewers. I used a difference in mean bootstrap as well as a 95% confidence interval for the difference in means to answer this question. Both the bootstrapped histogram and the confidence interval show that no difference is likely. More specifically, the difference in Q1_Happy response between "Living with Ed" and "My Name is Earl" is somewhere between -0.400 and 0.106 (which contains 0), with 95% confidence. Thus, the claim is not supported: between the two shows, one of them does not consistently produce a higher mean Q1_Happy response among viewers at the 95% confidence level.

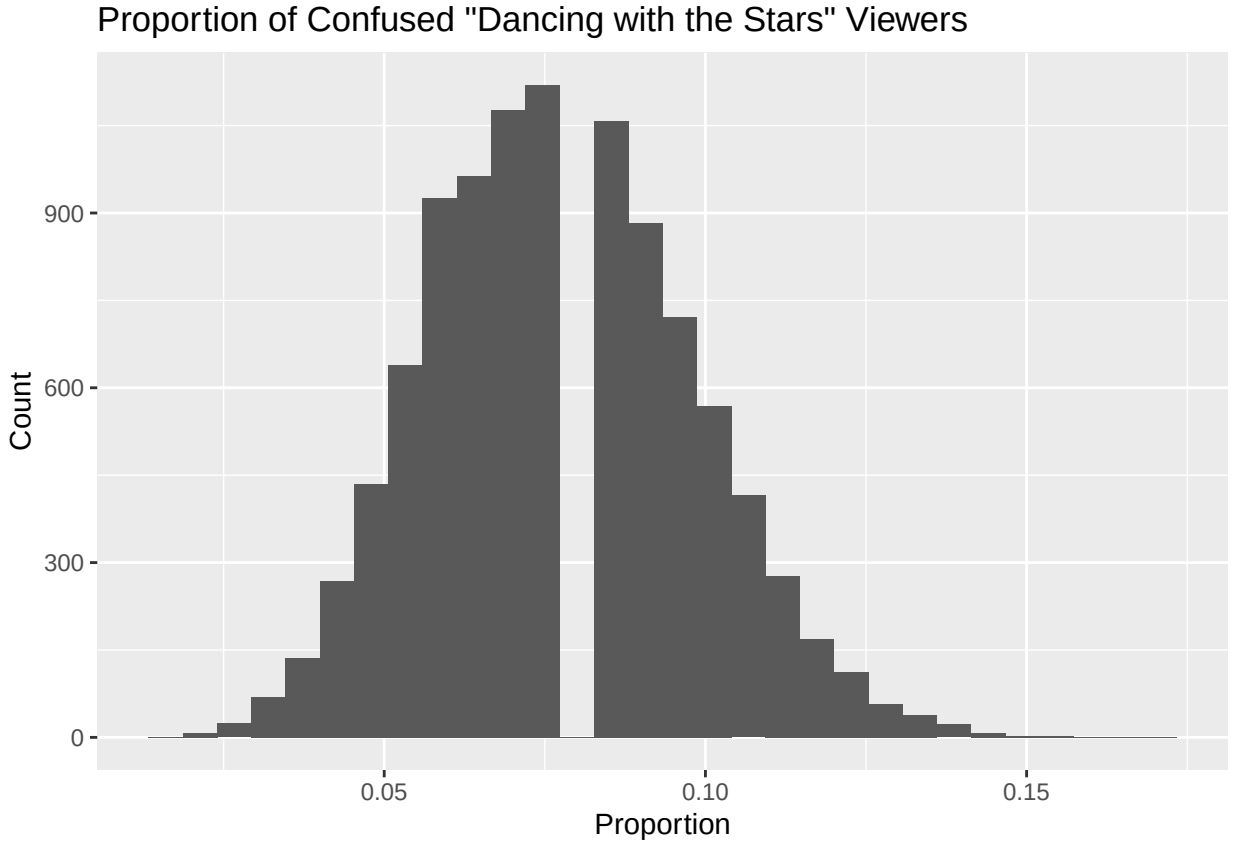
Part B.




```
##       name      lower       upper level     method  estimate
## 1 diffmean -0.5243109 -0.02088424  0.95 percentile -0.270997
```

The question was whether "The Biggest Loser" or "The Apprentice: Los Angeles" consistently produces a higher mean Q1_Annoyed (on a scale of 1-5, with 5 being very annoyed) response among viewers. I used a difference in mean bootstrap as well as a 95% confidence interval for the difference in means to answer this question. Both the bootstrapped histogram and the confidence interval show that a difference is likely. More specifically, the difference in Q1_Annoyed response between "The Biggest Loser" and "The Apprentice: Los Angeles" is somewhere between -0.524 and -0.026 (which does not contain 0), with 95% confidence. Thus, the claim is supported: between the two shows, one of them consistently produces a higher mean Q1_Annoyed response among viewers at the 95% confidence level.

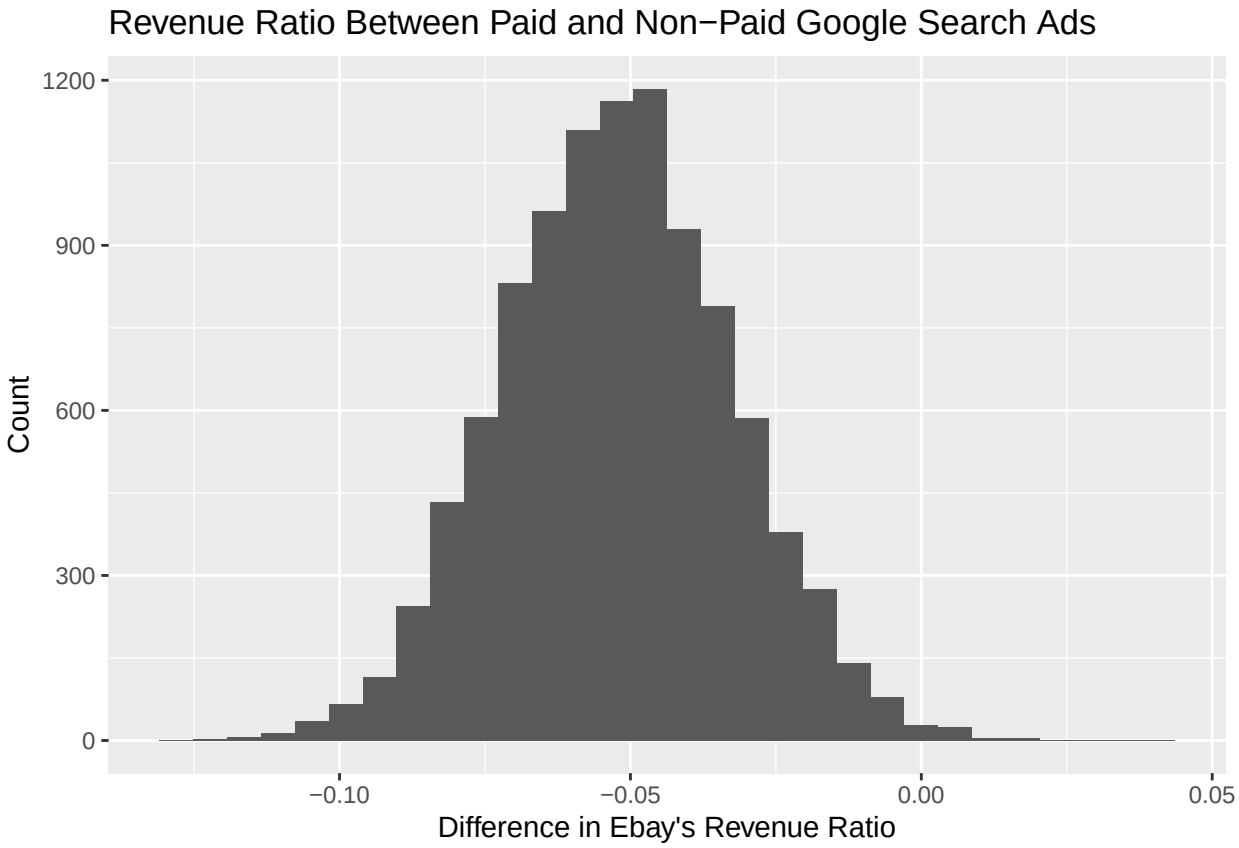
Part C.




```
##        name     lower    upper level     method  estimate
## 1 prop_TRUE 0.4167532 0.453098  0.95 percentile 0.4347525
```

The question was what proportion of American TV watchers would we expect to give a response of 4 or greater to the Q2_Confusing question (on a scale of 1-5, with 5 being very confused) for "Dancing with the Stars." I used a proportion bootstrap as well as a 95% confidence interval for the proportion of responses that are 4 or greater to this question. The proportion of 4 or greater responses to the Q2_Confusing for "Dancing with the Stars" is somewhere between 0.417 and 0.453, with 95% confidence. Thus, we are 95% confident that the proportion of American TV watchers with those responses would lie within that range.

Problem 4




```
##       name       lower       upper level     method    estimate
## 1 diffmean -0.08991054 -0.01340096  0.95 percentile -0.05228145
```

The question was whether or not paid search advertising on Google creates extra revenue for Ebay. I used a difference in mean bootstrap as well as a 95% confidence interval for the difference in means to answer this question. Both the bootstrapped histogram and the confidence interval show that a difference is likely. More specifically, the difference in the revenue ratios between paid search and non-paid search is somewhere between -0.091 and -0.013 (which does not contain 0), with 95% confidence. Thus, paid search advertising on Google likely creates extra revenue for Ebay at the 95% confidence level.