HW3

Isabelle Svendsen

2025-02-11

Isabelle Svendsen EID: iks294 Github Link: https://github.com/isvendsen06/HW3

## Problem One

Part A.
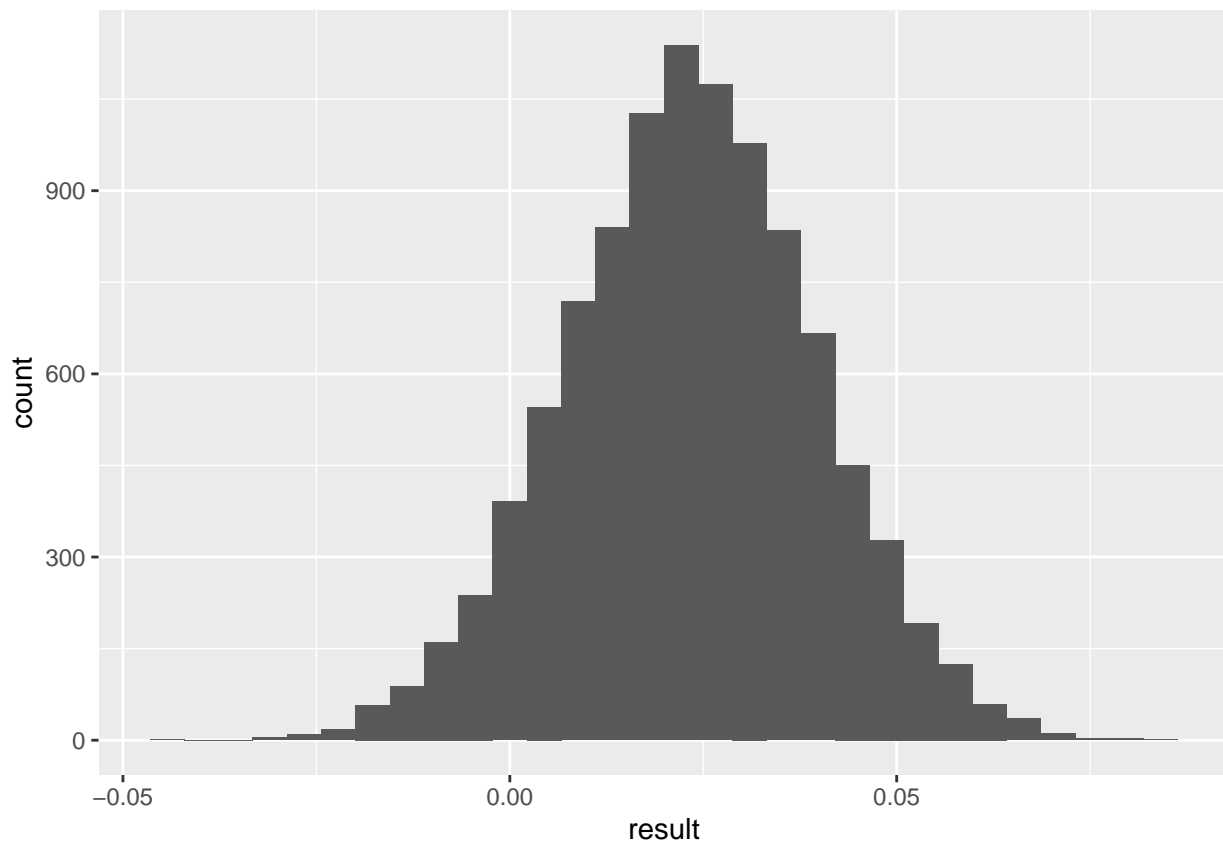

```
##     name        lower      upper level     method   estimate
## 1 result -0.009012059 0.05496147  0.95 percentile 0.02348235
```

Claim: Gas Stations charge more if they lack direct competition in sight. Evidence: When looking at the bootstrap that compares the means of the prices with a competitor vs. no competitor, there is a 95% confidence interval of around -0.0075 to 0.05 with the estimate around 0.0235. Conclusion: Our findings show that after bootstrapping 10000 times to compare the prices of Gas at Stations with and without competitors there is not enough evidence to prove and increase in prices. Since 0 is in the interval we are not able to tell if there is a difference between the prices.

Part B.
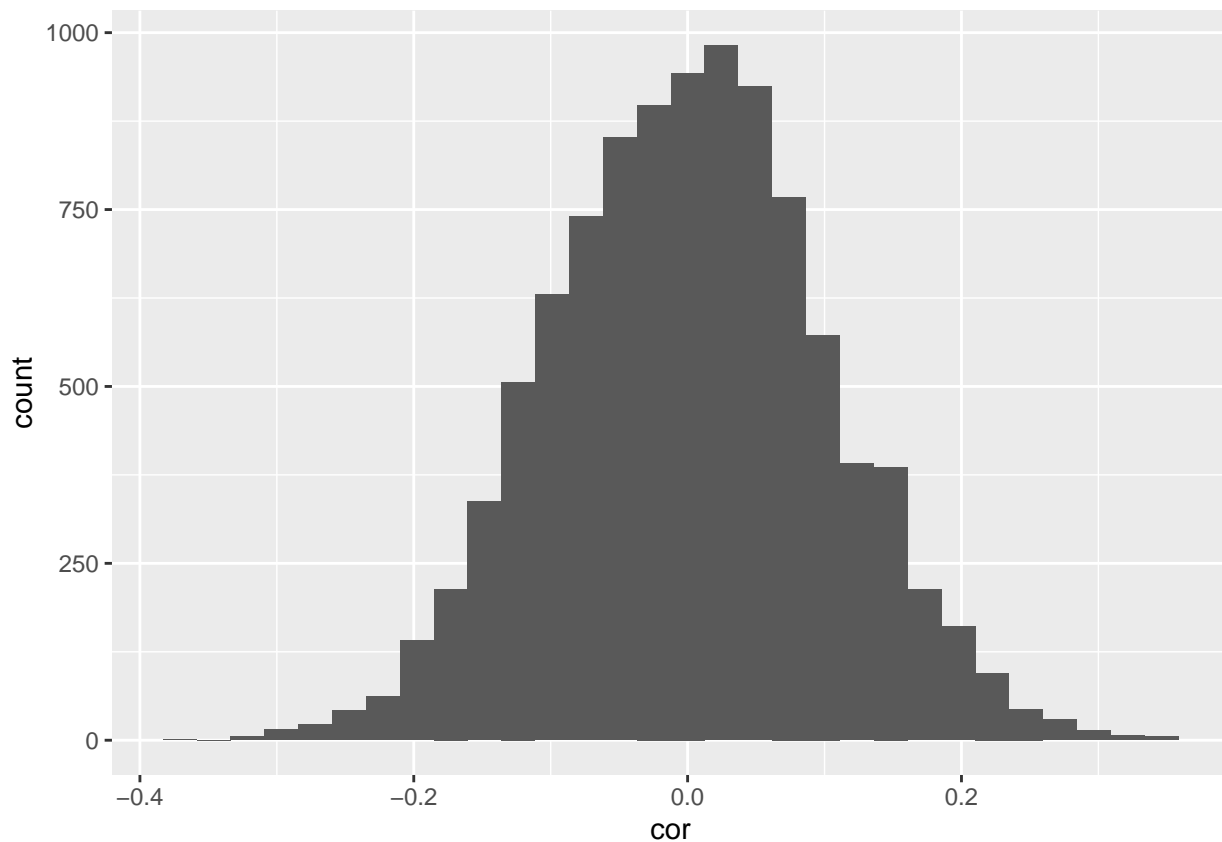

```
##   name      lower      upper level     method  estimate
## 1  cor -0.1917517 0.1991507  0.95 percentile 0.3961546
```

Claim: The richer the area, the higher the gas prices. Evidence: I used bootstrapping to look at the correlation of the prices of gas and then the average incomes of that area. Then from that I found the 95% confidence interval of the data and it is -0.193 to 0.197 with an estimate of 0.396. These is shown in the table above. Conclusion: In conclusion there is a nonexistent or very week correlation between the claim of the richer the area, the higher the gas prices. Since the confidence interval contains 0 in it, that means there is a possibility that there is no correlation. This proves that the ricer area you live in, that their gas prices are the same as other areas.
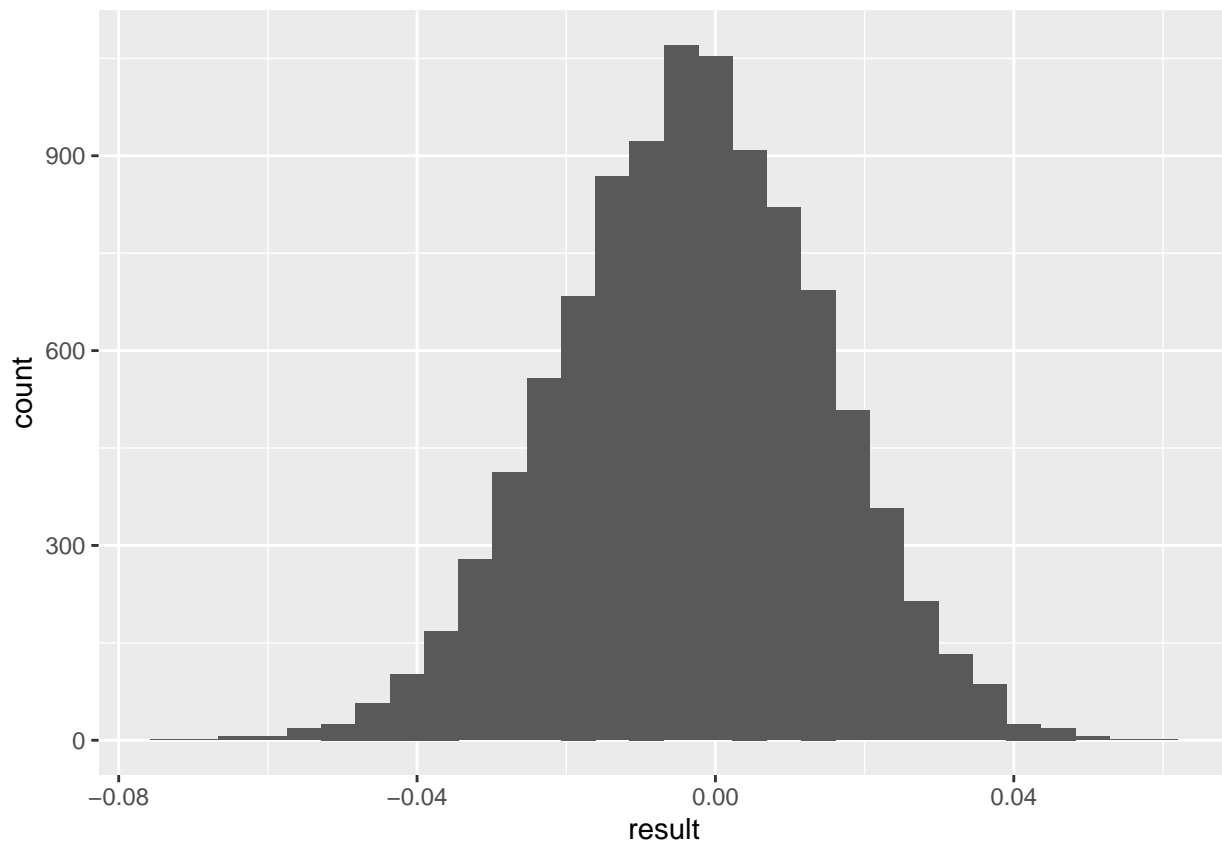
Part c.
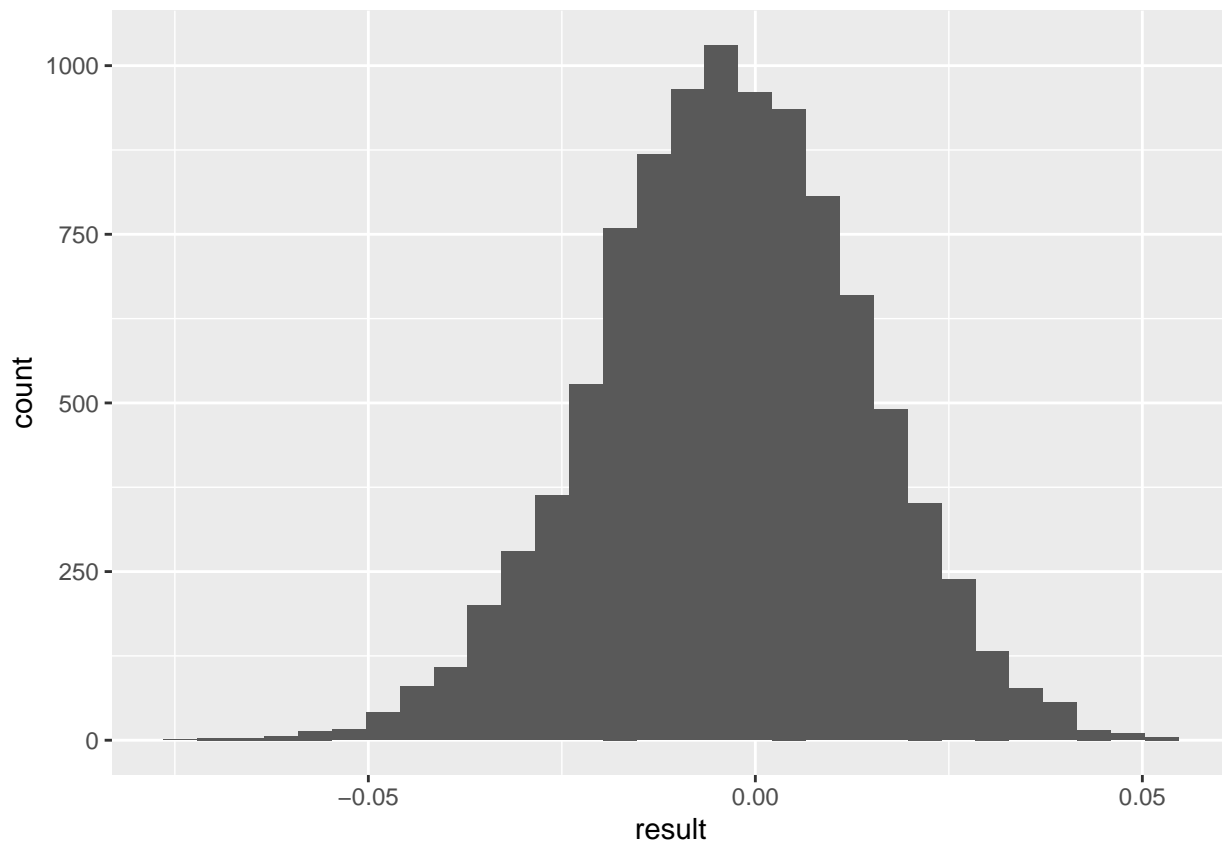

```
##     name      lower      upper level    method     estimate
## 1 result -0.0381203 0.02954887   0.95 percentile -0.003299916
```

Claim: Gas Stations at stoplights charge more. Evidence: The 95% confidence interval as shown above shows that the interval of possible means is from around -0.0380 to 0.0299 with an estimate of -0.0033. The graph above shows this in a clearer was as well. The graph and confidence interval show that 0 is included in the interval. Conclusion: Since 0 is included in the interval there is not enough evidence to prove that gas stations at stoplights charge more. By looking at the the difference in means we are able to tell is there is a difference between the prices and stoplights and the prices without stoplights. When the confidence interval contains 0 that means that there is a likely hood that the there is no difference between the two. Also the interval in itself is really small which makes it hard to prove that there is a relationship. That is why there is no sufficent evident to prove that gas stations at stoplights charge more.

Part D.


```
##     name        lower       upper level    method      estimate
## 1 result -0.03793651 0.03020175  0.95 percentile -0.003299916
```

Claim: Gas stations with direct highway access charge more. Evidence: Using bootstrapping to create the graph and confidence interval of the difference in mean prices between access from the highway and no access from the highway we can see that 0 is included in the interval of possible differences. Conclusion: Since 0 is included in the interval there is not enough evidence to prove that Gas Stations with direct highway access charge more.

Part E.
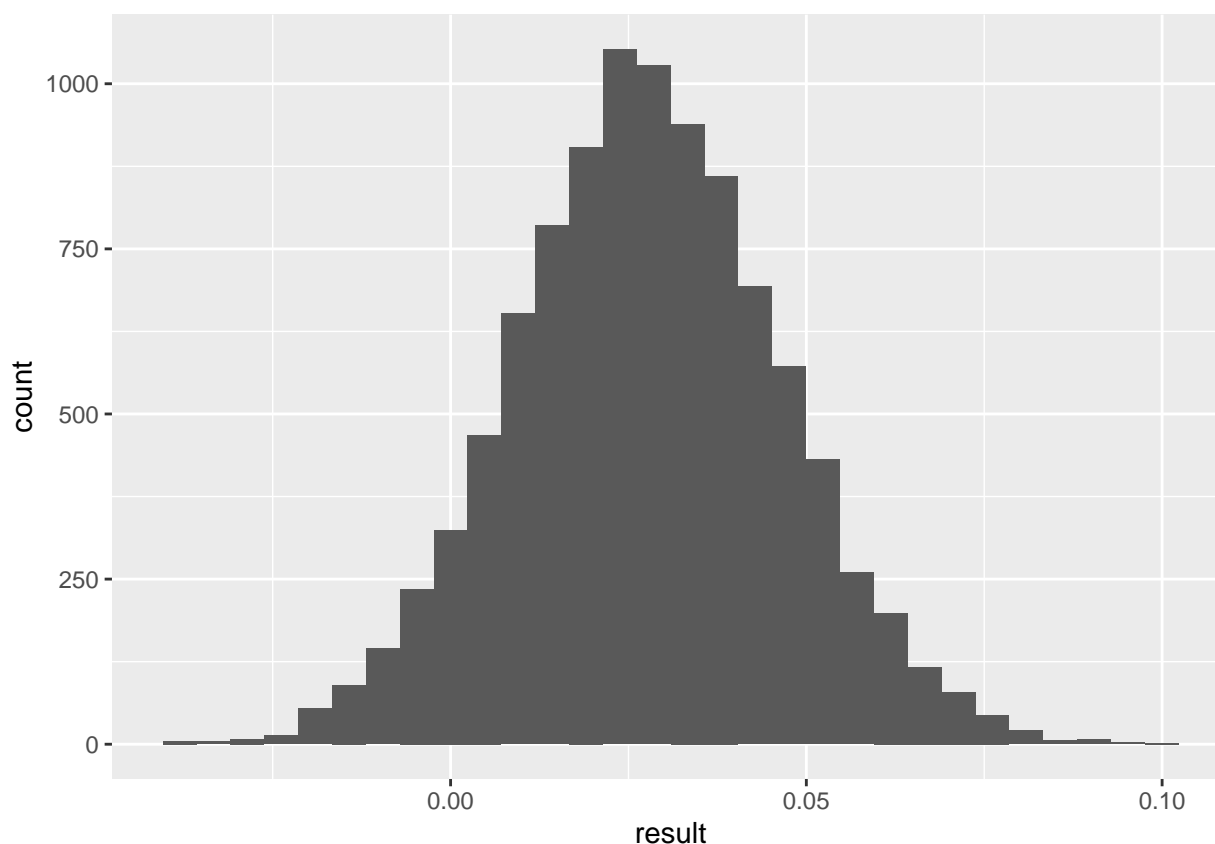

```
##      name        lower      upper level      method   estimate
## 1 result -0.00933465 0.06514966  0.95 percentile 0.02740421
```

Claim: Shell charges more than all other non-Shell brands Evidence:Based on both the graph and the confidence interval we can see that the possible differences in prices has a higher chance of being positive. However it is also shown that 0 is included in the interval. Conclusion: Since 0 is included there is a chance that they do charge the same price and that there is no difference between them and the other gas stations. However there is also a higher chance that there is a difference, with the majority of the interval being on the positive side, it wouldn't be surprising for a Shell gas station to have a very slight increase in price.

Problem 2.

Part A.

```
##   name    lower    upper level      method estimate
## 1 mean 26280.63 31801.27  0.95 percentile 28997.34
```

The confidence interval above shows that we would expect to see that the average mileage of used 2011 Mercedes S-lass vehicles, with 63 AMG trim, is between 26280.6284 and 31801.2716. This shows that we would expect the average to fall in this set of numbers. If we have a call that doesn't fall between this it shows cars that either would have been driven less then epected or way more then expected.

Part B.

```
##        name     lower     upper level      method  estimate
## 1 prop_TRUE 0.4167532 0.4527518  0.95 percentile 0.4347525
```

The confidence interval above shows the expected values of the proportion of 2014 Mercedes S-Class vehicles, with trim 550, which are the color black. The interval of 0.4168 and 0.4528 shows that we would expect the proportions to fall in this range of around 40% black cars from 2014 with this trim.

Problem 3.

Part A.

Is there evidence that of the shows, "Living with Ed" and "My Name is Earl", one makes the audience happier? To find this answer, I used 10000 bootstrap samples of the filtered data set of the two shows. In the bootstrap, I found the means for each and subtracted them to find the difference in the means. Then using this I found the confidence Interval.
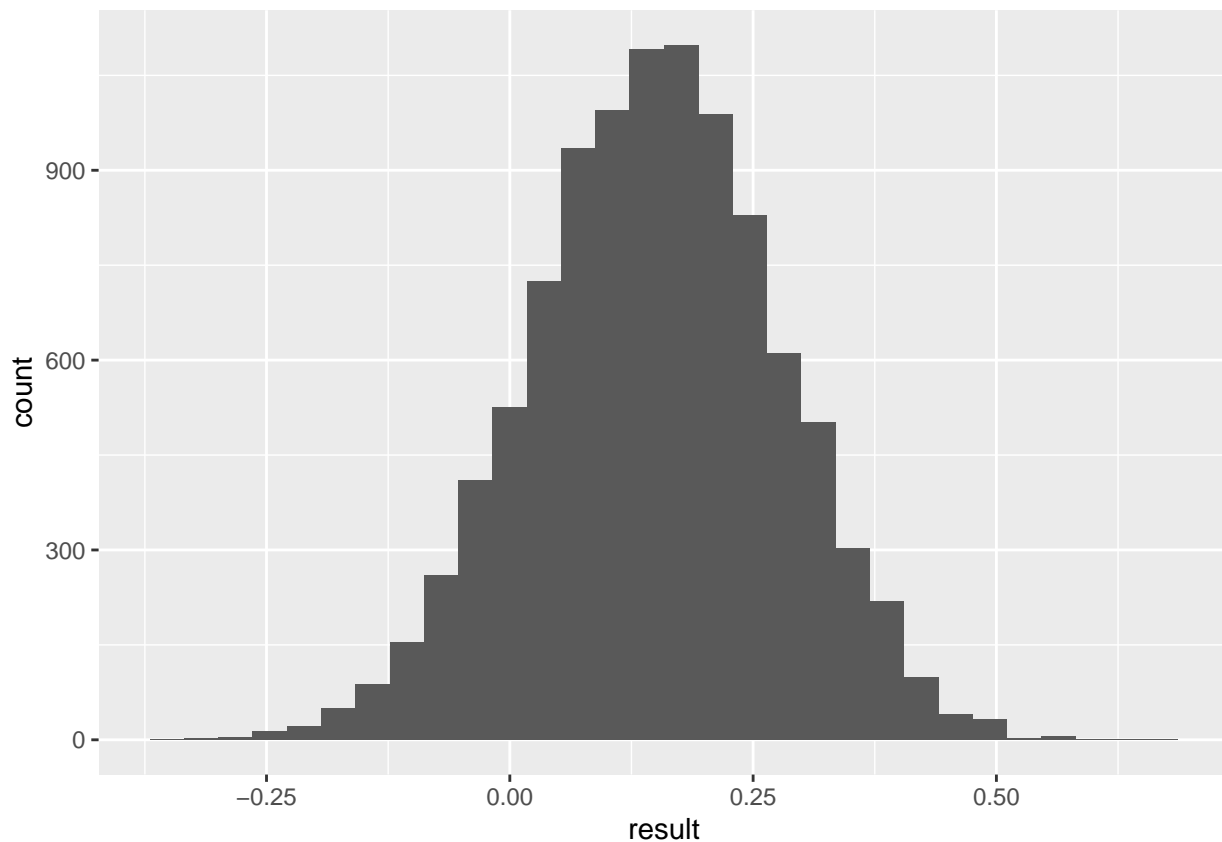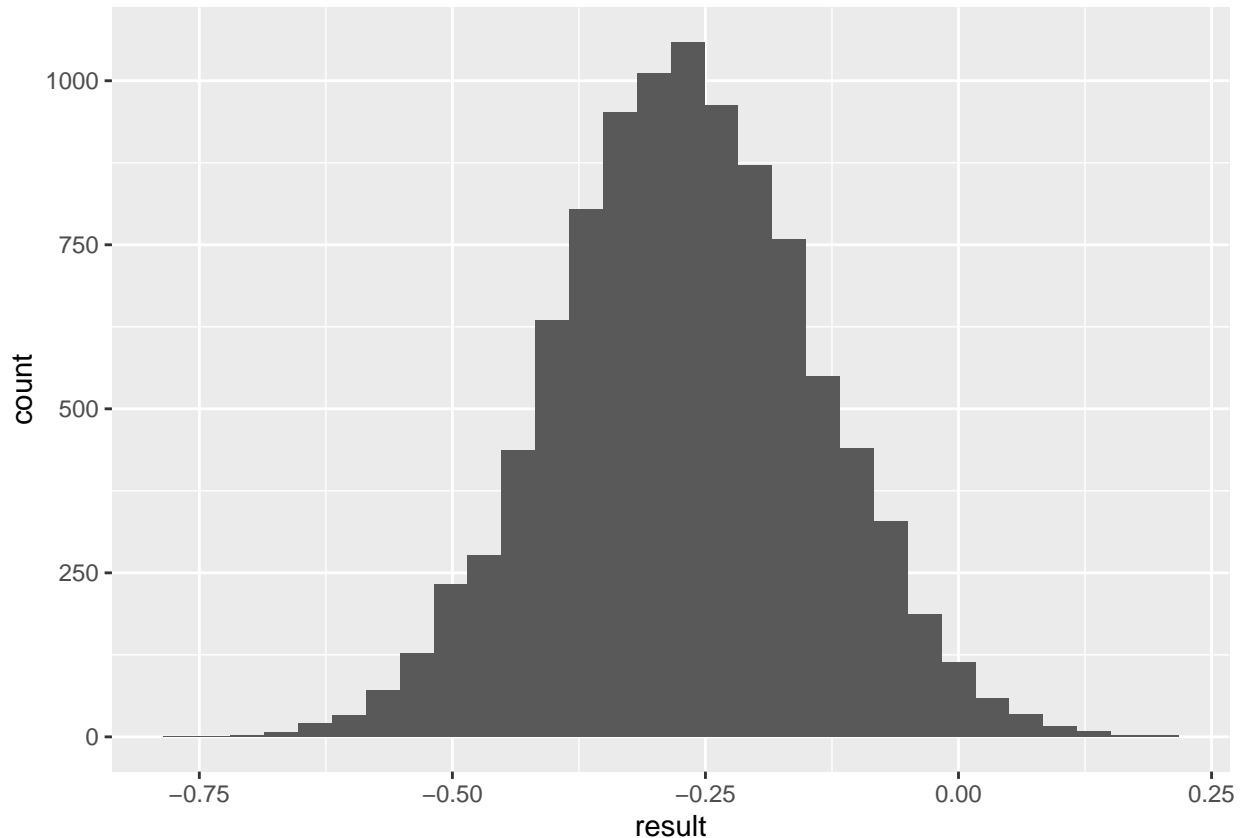


```
##      name      lower      upper level      method  estimate
## 1 result -0.1018751 0.3899251  0.95 percentile 0.1490515
```

The graph and confidence interval above show the distribution of the differences in means for the both shows. The 95% confidence interval and the graph show that the a possible difference is 0, however the 0 is on the left side of the interval with more of the interval on the right. This means we a 95% confident that the difference will fall into this range of numbers. However, since 0 is included in this range, there isn't enough evidence to prove that one show make the audience happier. This is good for the producers because then they know that they aren't doing worse compared to the other show.

Part B.

Is there evidence that between the two shows, "The Biggest Loser" and "The Apprentice: Los Angeles" that the audience felt more annoyed while watching the show. To find out this answer I used 10000 bootstrap samples of the filtered data set of the two shows. In the bootstrap, I found the means for each and subtracted

them to find the difference in the means. Then using this I found the confidence Interval and the histogram of the distributions.



```
##     name     lower      upper level    method estimate
## 1 result -0.5228741 -0.01885586  0.95 percentile -0.270997
```

The confidence interval and graph shown above show the distribution of the differences in the mean annoyed ratings. The interval doesn't pass through 0 and all of the values are negative. This means that there is enough to show that "The Apprentice: Los Angeles" was more considered more annoying for the audience. The reason we can conclude this is because all of the expected values for the difference and all on one side of 0 meaning that there is an expected difference. This shows the people making the show to try and fix their writing to make it less annoying for the audience. If it is annoying people might stop watching, which is bad for the show producers.

Part C.

What proportion of the audience would we expect to give a response of 4 or greater to the question asking how confusing the show was? I found this by using 10000 bootstrap samples of the filtered data set of the show. I then created a new variable in the data set of whether they rated it a 4 or higher on the confusing question. In the bootstrap, I found the proportion of this variable. Then using this I found the confidence Interval and the histogram of the distributions.

Bootstrapped Proportions of Respondents Giving 4 or Greater on Q2_Con
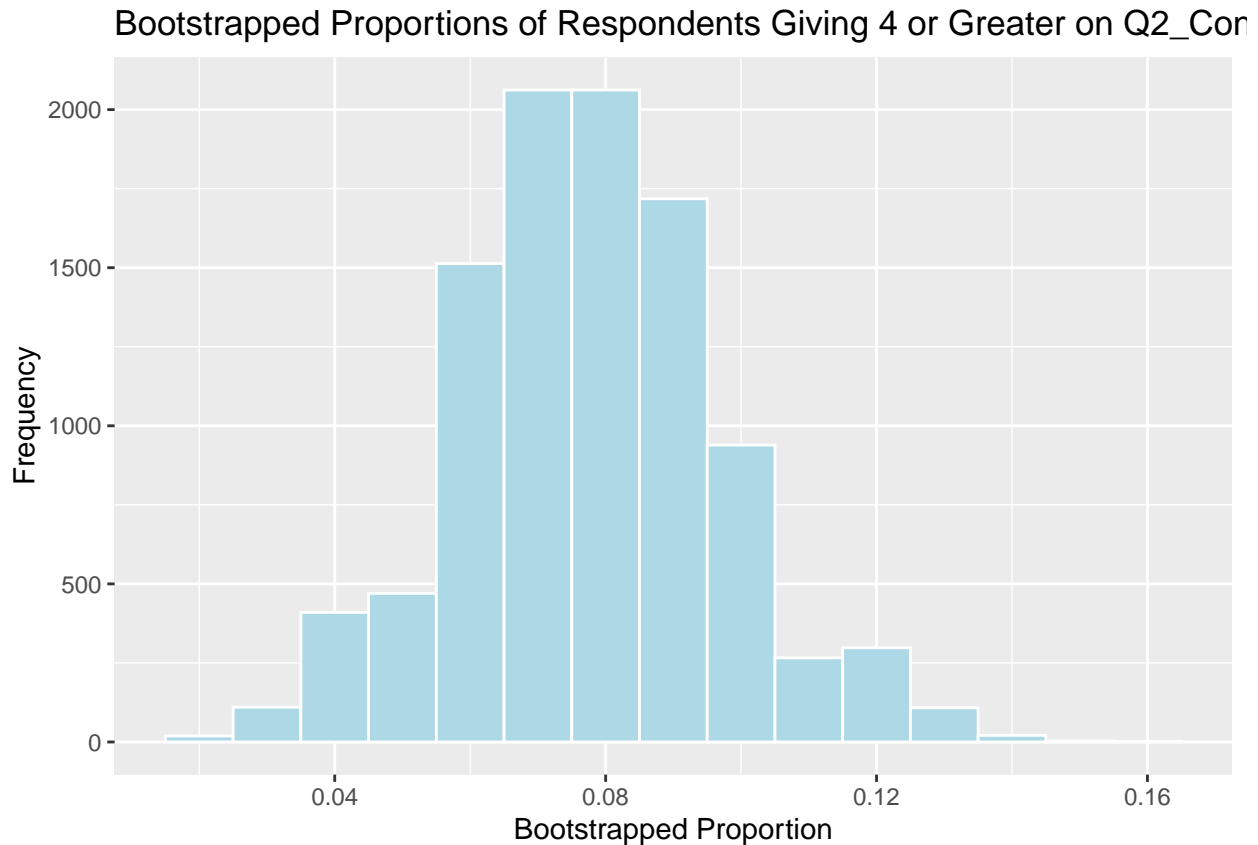


```
##      name       lower      upper level      method    estimate
## 1 prop_TRUE 0.03867403 0.1160221  0.95 percentile 0.07734807
```

The graph and the confidence interval above show the expected proportions to be between 0.0387 and 0.116. The estimate it gave us is 0.0773.This means that we would expect around 7.7% of people to have rated Dancing with the Stars confusing with a 4 or higher. However, I am 95% confident that the proportion will fall in the confidence interval shown above. This is good for the producers because it shows that the proportion of people who are considered pretty confused on the show is a low number.

Problem 4.

Does the paid search advertising on Google create extra revenue for EBay? To find this answer we want to find the ratio of the revenue for before and after their experiment. Then we want to sort the treatment data and the control data to see if there is a difference between the two. Using a bootstrap to find the differences in the means of the revenue ratios for the reatment and the control, I was able to get a confidence interval.

Bootstrapped Differences of Revenue Ratios
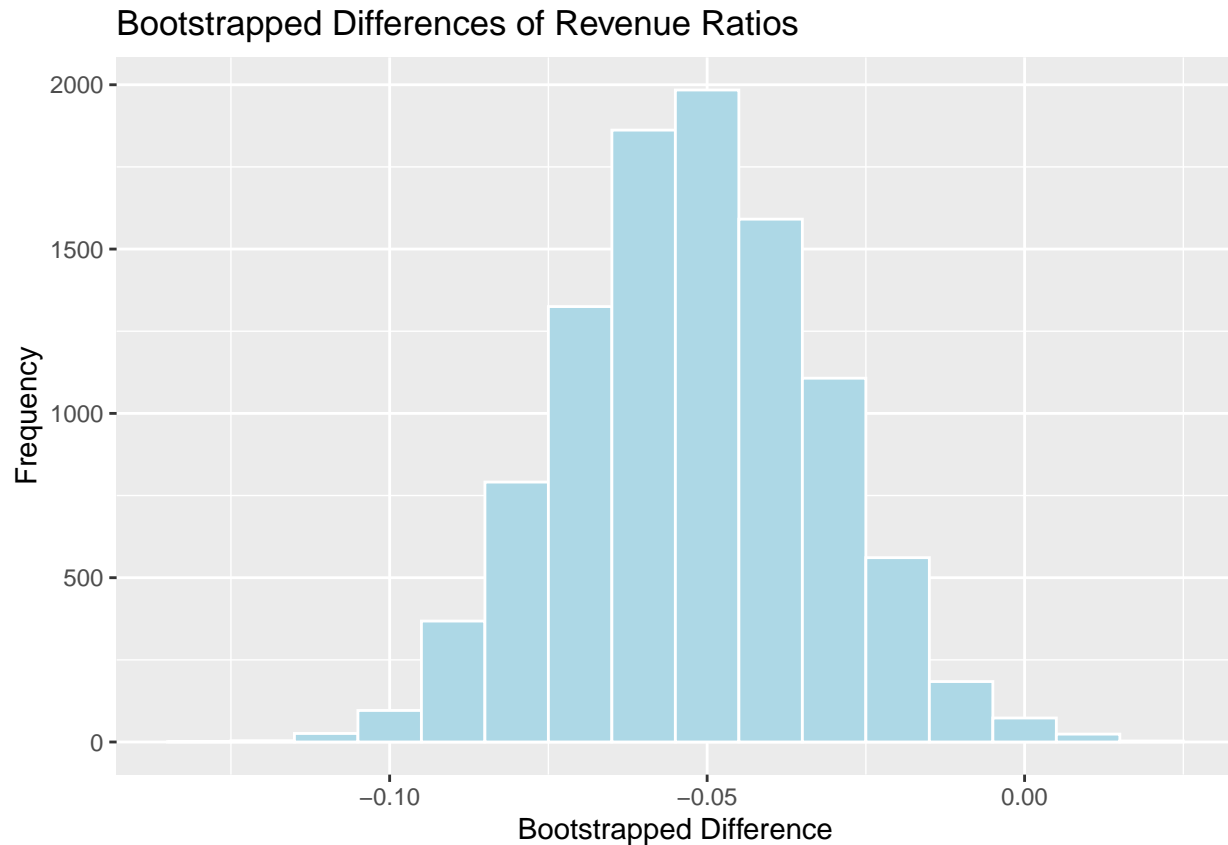

```
##     name      lower      upper level    method     estimate
## 1 result -0.09070234 -0.01370663  0.95 percentile -0.05228145
```

The data the graph and confidence interval show is that the bootstrapped difference is negative and on the interval is shows that, -0.0907 and -0.0137. The estimate it gave us is -0.0523. 0 is not included in the interval which tells us that we do have enough evidence. This shows that there is a difference and that the control samples had repeatedly a higher mean then the treatment group. This is really important to EBay because it shows that the ads do work and help them get revenue. This means that they would want to continue to pay for the Google ads because in the long run it is helping them earn more money even though they are spending more as well.