

# Hw\_3

Pranita

2025-02-12

Name: Pranita Chaudhury

UT EID: pc28377

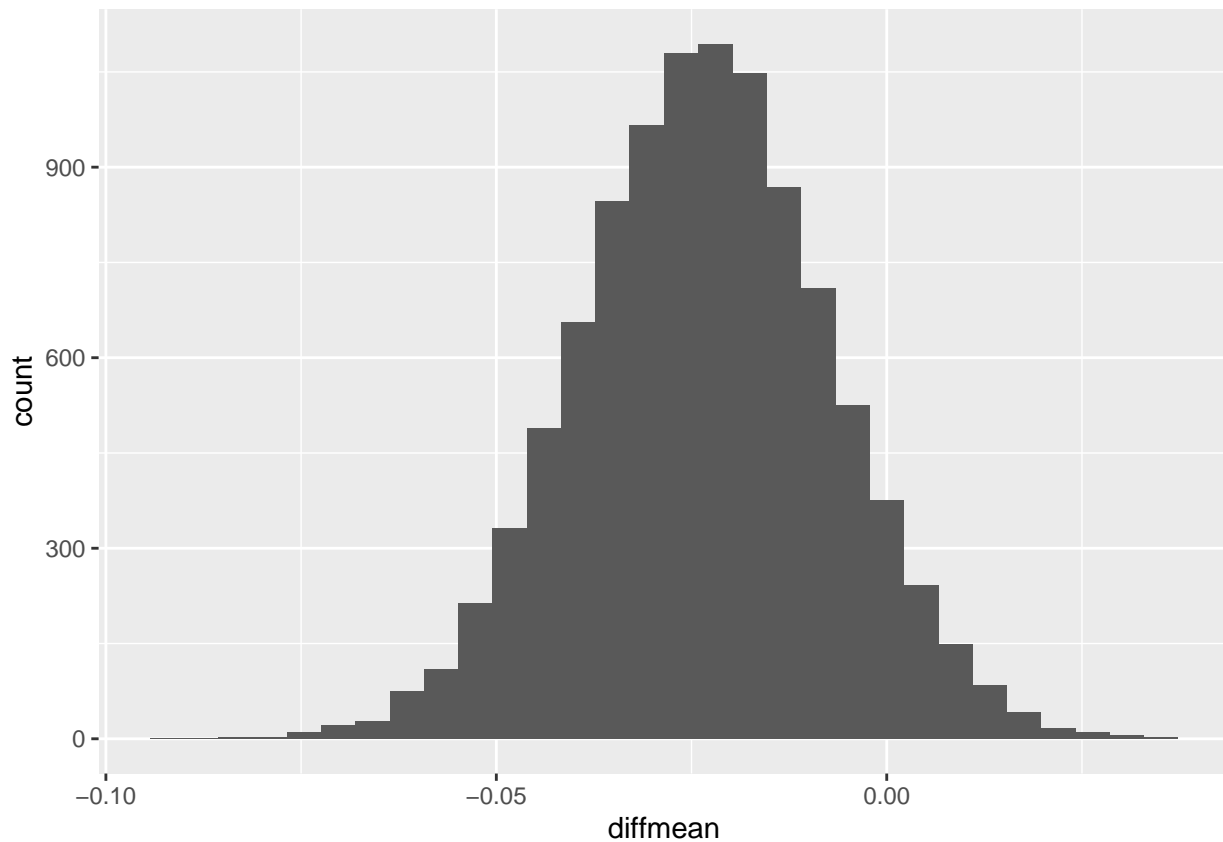
Github: [https://github.com/PranitaChau/Hw\\_3.git](https://github.com/PranitaChau/Hw_3.git)

**to do later- make graphs look pretty, fix 1b 3c, write up all conclusions for ALL**

## Problem 1

### Theory A

The first theory is that gas stations charge more if they lack direct competition in sight.



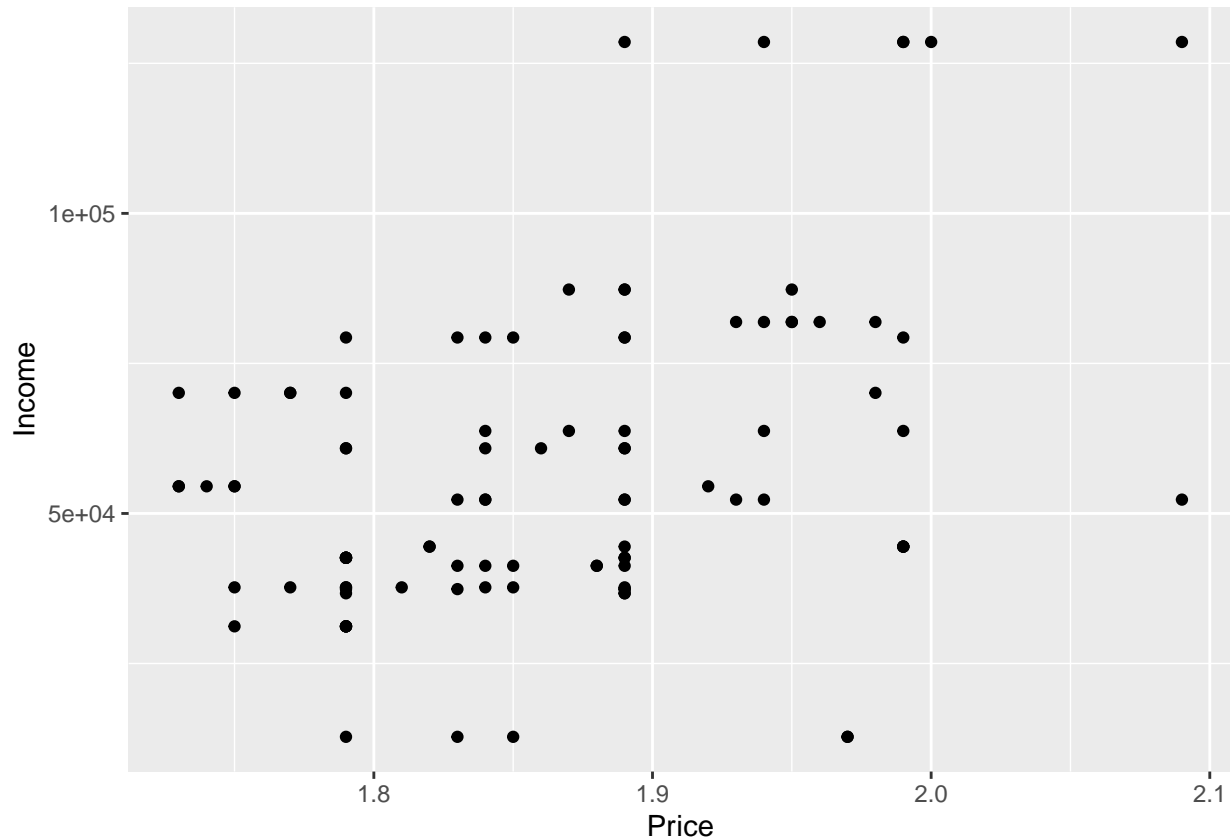
##	name	lower	upper	level	method	estimate
----	------	-------	-------	-------	--------	----------

```
## 1 diffmean -0.05485557 0.007865986 0.95 percentile -0.02348235
```

The claim is gas stations that lack direct competition in sight charge higher prices for gas compared to those that are located with competitors in sight. We can collect evidence for this claim by comparing the mean price for gas stations with and without competitors in sight. Using a 95% confidence interval, there is a -0.054 to 0.008 difference in prices. Since 0 is included in this interval, it should be noted that there is no significant difference in prices for gas stations that do and do not have competitors nearby. Therefore the claim that gas stations with competitors in sight cannot be supported with our data using a 95% confidence interval.

## Theory B

The richer the area, the higher the gas price



```
## (Intercept)      Income
## 1.793442e+00 1.248341e-06

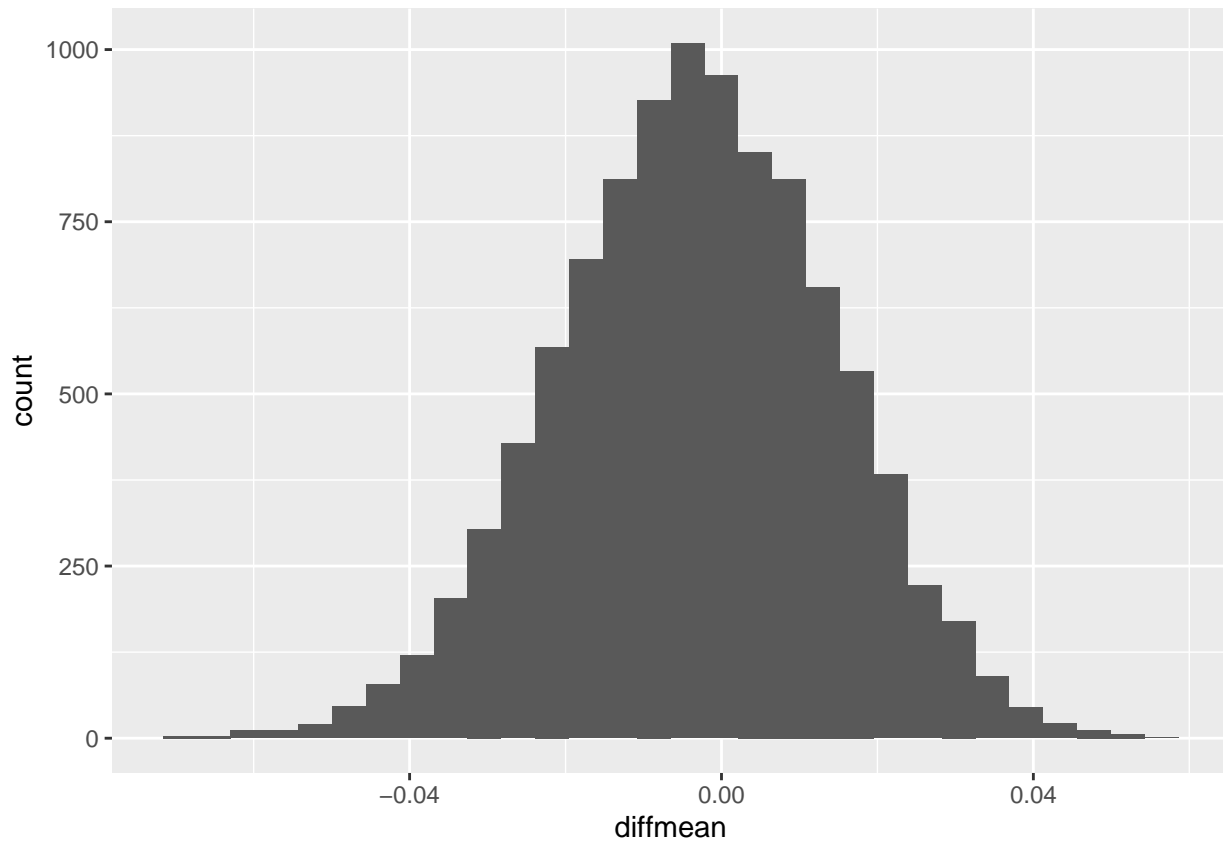
##      name      lower      upper level      method      estimate
## 1 Intercept 1.765583e+00 1.828387e+00 0.95 percentile 1.803386e+00
## 2   Income 6.808302e-07 1.676457e-06 0.95 percentile 1.061923e-06
## 3    sigma 6.233794e-02 8.420965e-02 0.95 percentile 7.954472e-02
## 4 r.squared 4.286526e-02 3.368983e-01 0.95 percentile 9.833010e-02
## 5         F 4.436452e+00 5.030142e+01 0.95 percentile 1.079628e+01

## [1] 0.08151499
```

The claim is that the higher the median income of an area, the price of gas will be more expensive. Evidence can be collected for this by comparing the median income of an area to the price of gas in said area. By plotting the graph there is a slight positive correlation visible, but it is not obvious so we can try to fit a linear model comparing the gas prices and income of the area. The coefficient for income shows a very small positive correlation, suggesting there is a weak correlation between gas prices of 1.248e-6.

## Theory C

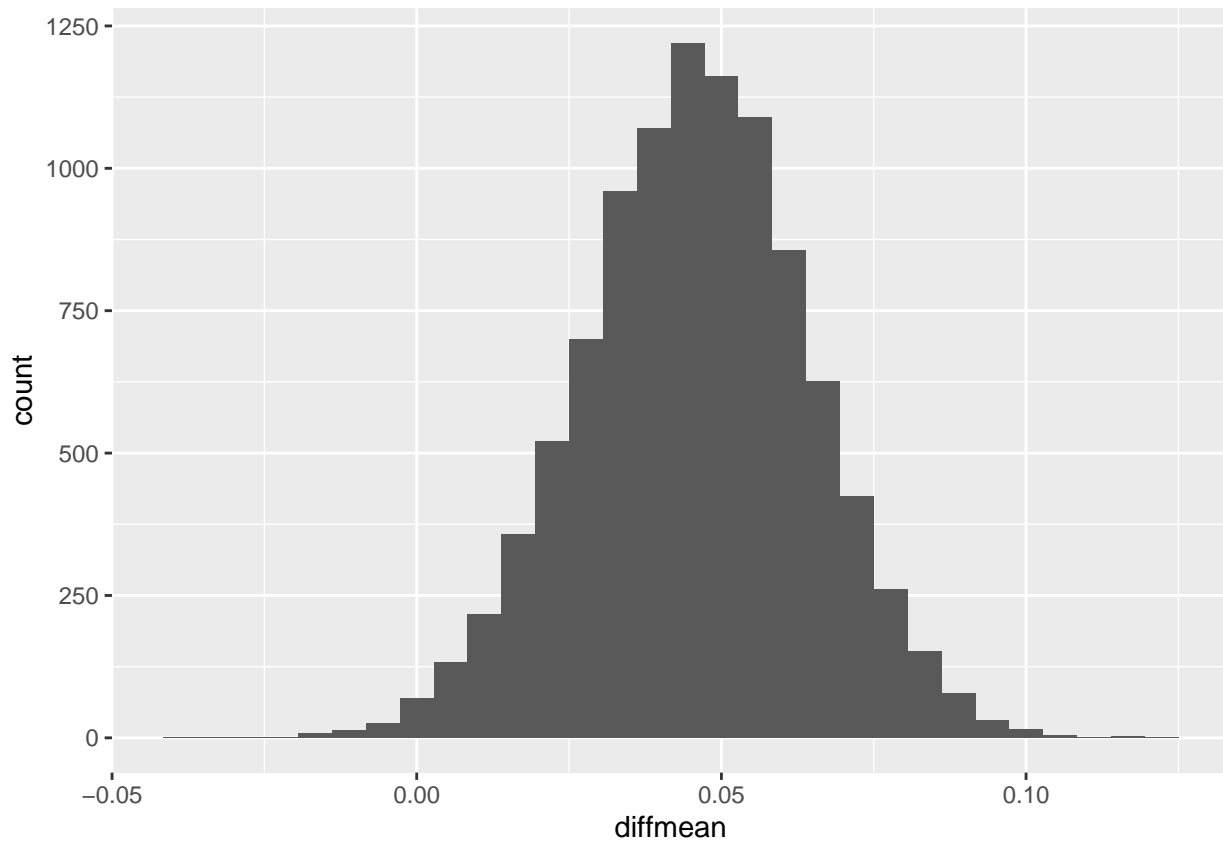
Gas stations at stoplights charge more



```
##      name      lower  upper level  method  estimate
## 1 diffmean -0.03797172 0.03051203  0.95 percentile -0.003299916
```

## Theory D

Gas stations with direct highway access charge more

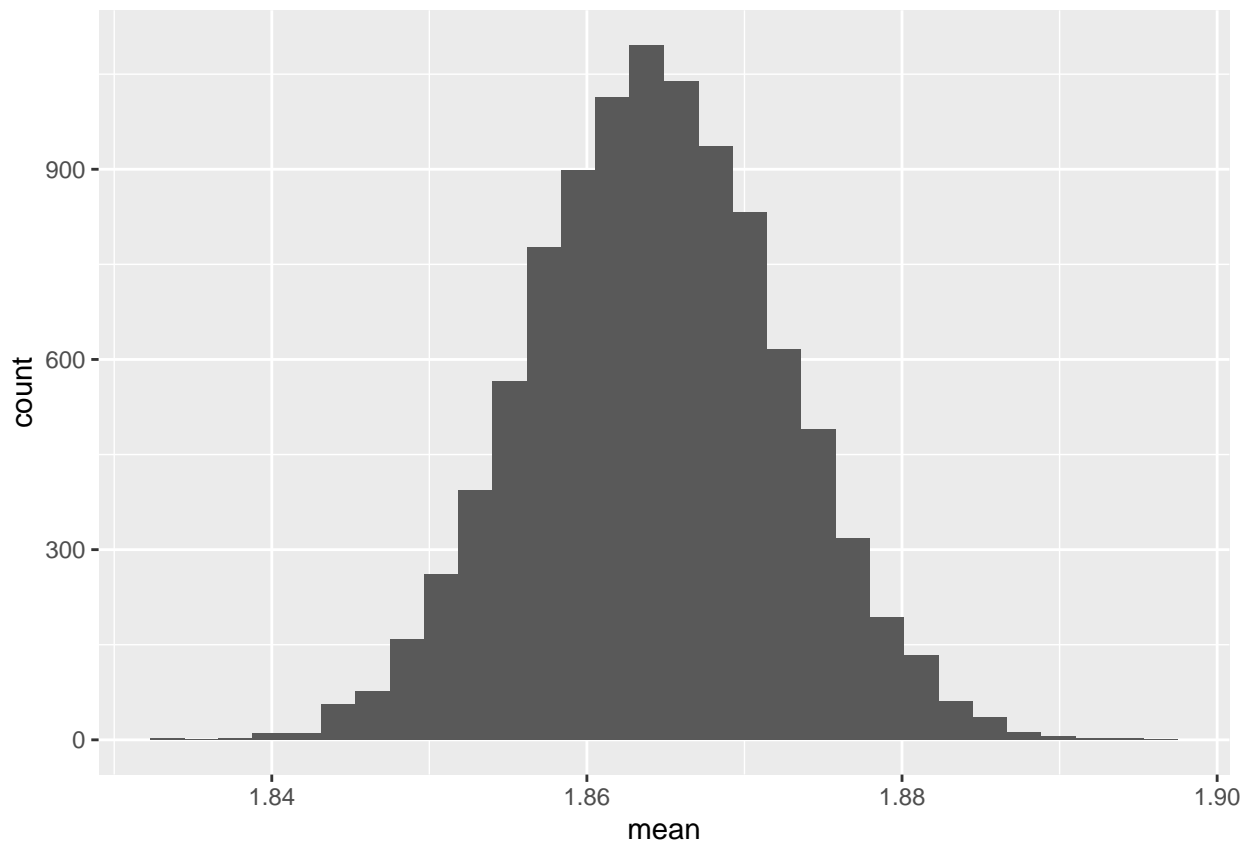


```
##      name      lower      upper level      method estimate
## 1 diffmean 0.008330804 0.08126375  0.95 percentile 0.0456962
```

The claim is that gas stations with direct highway access tend to charge more than gas stations further away from the highway. Evidence can be gathered for this by using a 95% confidence interval to compare gas prices for gas stations with and without direct highway access. The result of that procedure shows that there is between a 0.009 and 0.081 price difference between gas stations with and without direct highway access. Therefore, we can conclude with a 95% confidence interval that gas stations with direct highway access do tend to charge somewhere between \$0.009 and \$0.081 more than gas stations without direct highway access.

## Theory E

Shell charges more than all other non-Shell brands



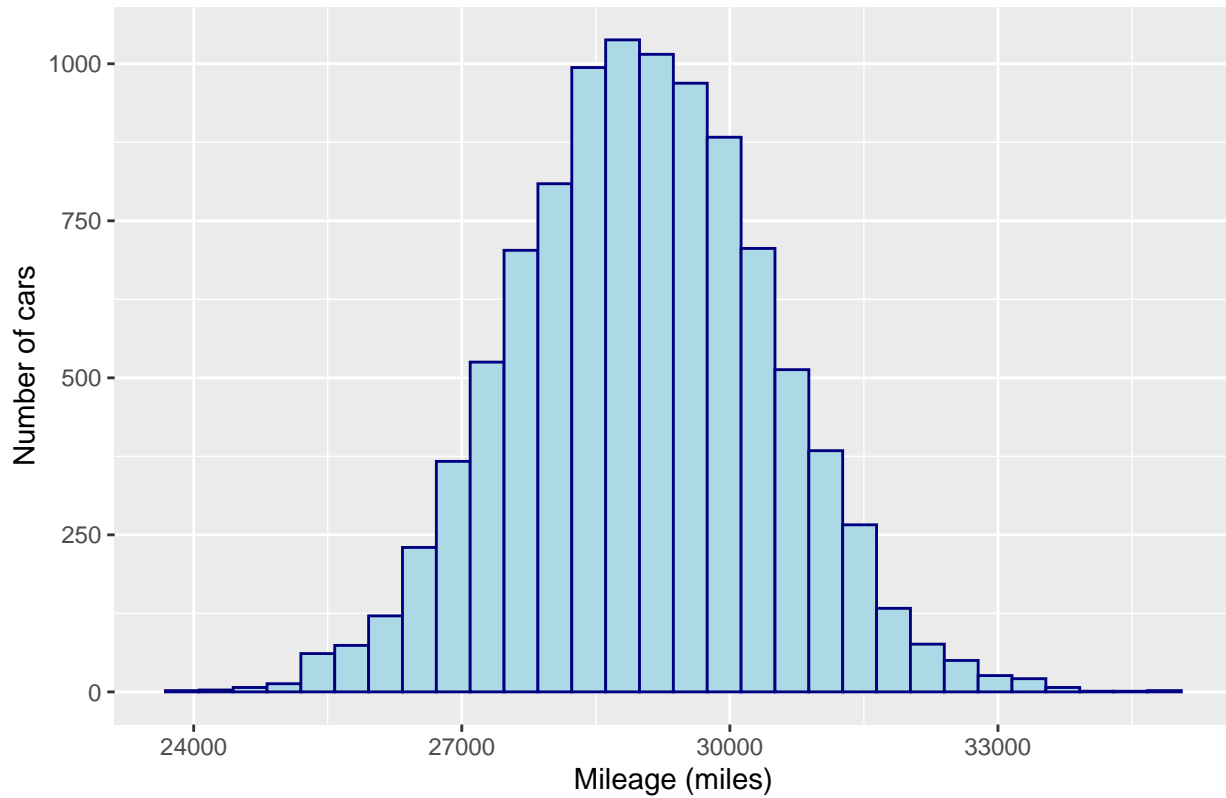
```
##   name   lower   upper level   method estimate
## 1 mean 1.848713 1.880198 0.95 percentile 1.864257
```

The claim is the brand Shell charges more than other brands of gas. This can be tested through comparing the mean prices of the price of gas that Shell charges to all other gas brands. Using a 95% confidence interval it can be determined that there is a 1.85 to 1.88 difference in prices. Therefore, we can conclude that on average the difference in prices between gas from Shell compared to other gas brands is somewhere between \$1.85 and \$1.88 with 95% confidence, and the claim is supported.

## Problem 2

### Part A

Sampling distribution for 63 AMG Mercedes mileages in 2011

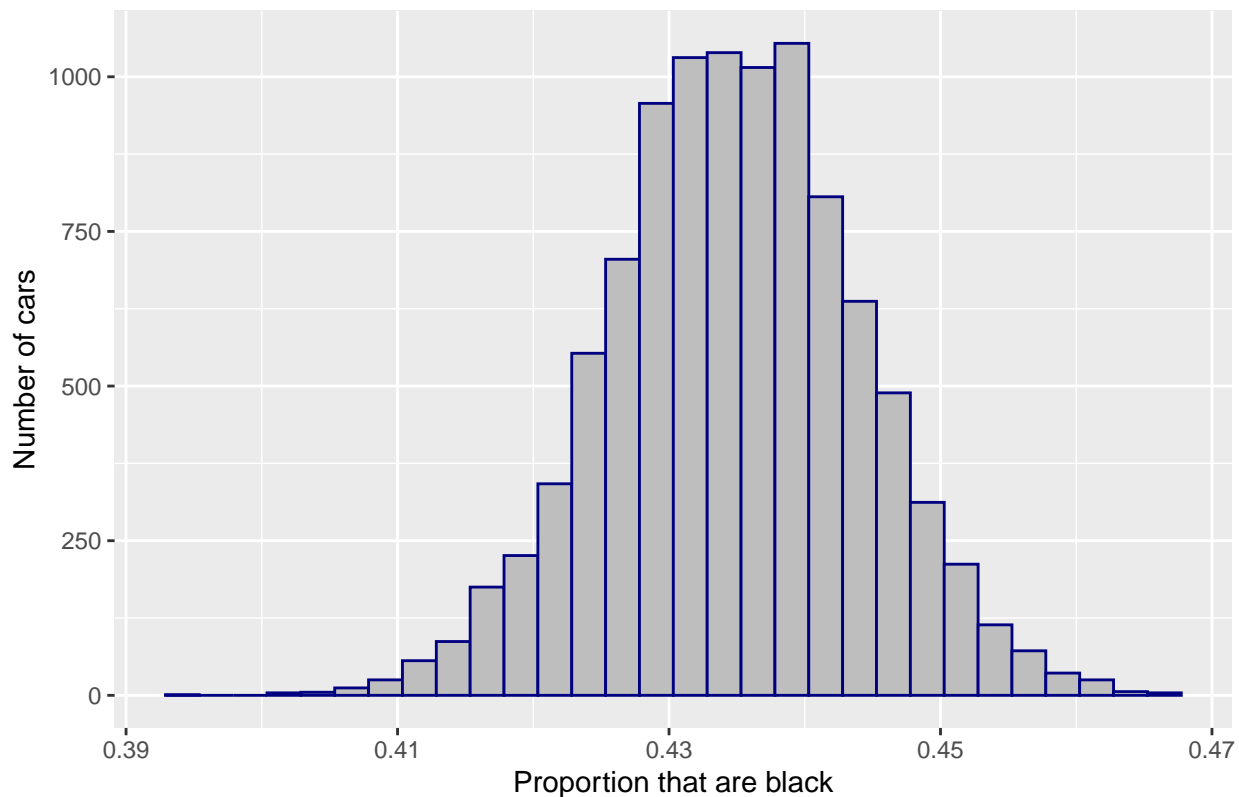


```
## name lower upper level method estimate
## 1 mean 26277.06 31830.67 0.95 percentile 28997.34
```

We are looking at used Mercedes S-Class vehicles sold on cars.com with a trim of 63 AMG in the year 2011. The graph shows the distribution of their mileage. Based on this data we can say with 95% confidence that the average mileage of cars that fit in this category is somewhere between 26293.74 and 31777.48 miles.

## Part B

### Sampling distribution for 550 trim Black Mercedes in 2014



```
##      name      lower      upper level      method      estimate
## 1 prop_TRUE 0.4164071 0.453098 0.95 percentile 0.4347525
```

We are looking at used Mercedes S-Class vehicles sold on cars.com with a trim of 550 in the year 2014. The graph shows the distribution of the proportion of cars colored black. Based on this data we can say with 95% confidence that the average proportion of black cars that fit in this category is somewhere between 0.4164 and 0.4531.

## Problem 3

### Part A

```
##      name      lower      upper level      method      estimate
## 1 result -0.3969811 0.1059561 0.95 percentile -0.1490515
```

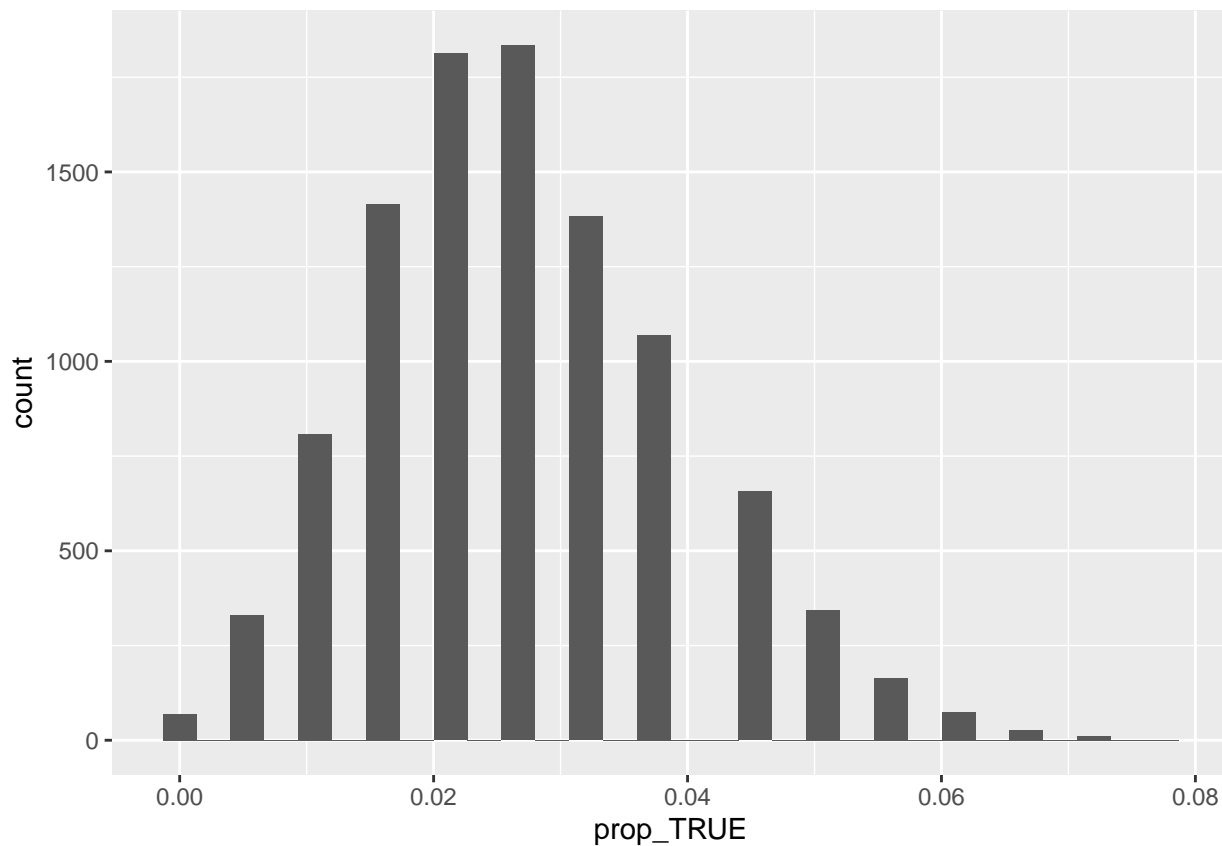
I am answering the question of which show makes people happier: Living with Ed or My Name is Earl? In order to answer this question I used the bootstrapping method. I made a data set only containing the answers

### Part B

```
##      name      lower      upper level      method      estimate
## 1 result -0.393486 0.09447433 0.95 percentile -0.1490515
```

No

## Part C



```
##      name      lower      upper level      method      estimate
## 1 prop_TRUE 0.005524862 0.05524862 0.95 percentile 0.02762431
```

idk if this is right, the #'s are not changing when re sampling

## Problem 4

```
##      name      lower      upper level      method      estimate
## 1 result -0.09111105 -0.01425823 0.95 percentile -0.05228145
```

The question I am trying to answer is whether the revenue ratio is the same in the treatment and control groups, or whether instead the data favors the idea that paid search advertising on Google creates extra revenue for EBay

Question: What question are you trying to answer? 2) Approach: What approach/statistical tool did you use to answer the question? 3) Results: What evidence/results did your approach provide to answer the question? (E.g. any numbers, tables, figures as appropriate.) 4) Conclusion: What is your conclusion about your question? Provide a written interpretation of your results, understandable to stakeholders who might plausibly take an interest in this data set. It is certainly possible in this case for each of these four sections to be only