Homework 2

Trent Latz

2025-02-09

Name: Trent Latz **UT EID:** tjl2597 **GitHub Link:** https://github.com/trentjlatz/SDS315-HW3

```r
knitr::opts_chunk$set(echo = FALSE,
                      eval = TRUE,
                      warning = FALSE,
                      message = FALSE,
                      fig.align = "center",
                      R.options = list(max.print=50))
suppressPackageStartupMessages(suppressWarnings({
  library(tidyverse)
  library(mosaic)
}))
gas <- read.csv("gasprices.csv")
sclass <- read.csv("sclass.csv")
nbc <- read.csv("nbc_pilotsurvey.csv")
ebay <- read.csv("ebay.csv")
```

Problem 1:

Which of these theories seem true, and which are unsupported by data? Take each theory one by one and assess the evidence for the theory in this data set.
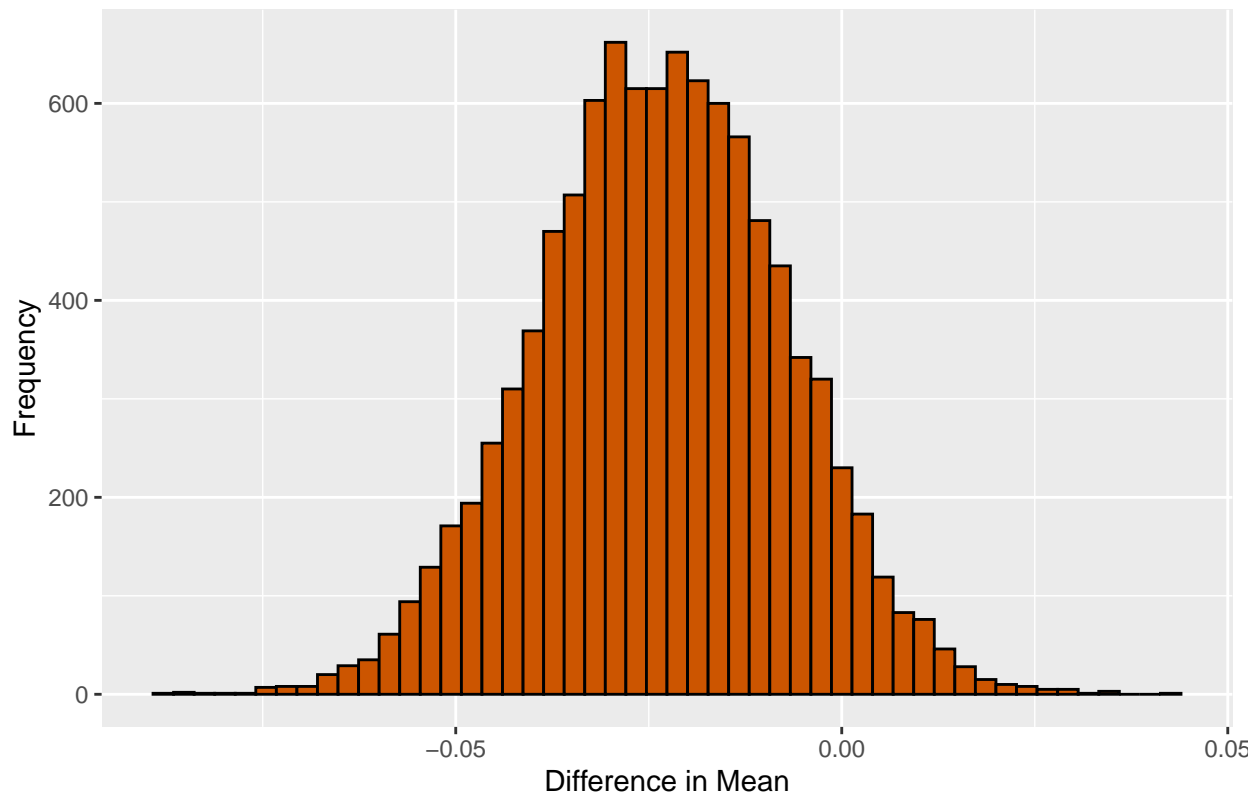
Theory A:

Claim:

Gas stations charge more if they lack direct competition in sight.

Evidence:

We will compare average gas price between stations with and without competitors, then bootstrap for accuracy.

```
##        N        Y
## 1.875882 1.852400
```

Bootstrapped Distribution of Difference in Mean Gas Prices


```
##       name      lower      upper level     method     estimate
## 1 diffmean -0.05513981 0.007615941  0.95 percentile -0.02348235
```

Conclusion:

The confidence interval for the difference in mean gas prices between those with and without competition is (-5.41 cents, .87 cents) with 95% confidence. Since the interval includes zero, we cannot reject the null hypothesis. The presence of visible competition does not appear to have a significant effect on the mean gas prices based on this analysis.

Theory B:

Claim:

The richer the area, the higher the gas prices.

Evidence:

```
##     12786    31192    36698    37396    37690    41279    42615    44476
## 1.882000 1.783333 1.856667 1.850000 1.811250 1.861667 1.815000 1.927143
##     52306    54526    60856    63750    70095    79315    81903    87306
## 1.906250 1.770000 1.843333 1.906000 1.798333 1.868571 1.951667 1.900000
##    128556
## 1.983333
```

Bootstrapped Distribution of Difference in Mean Gas Prices
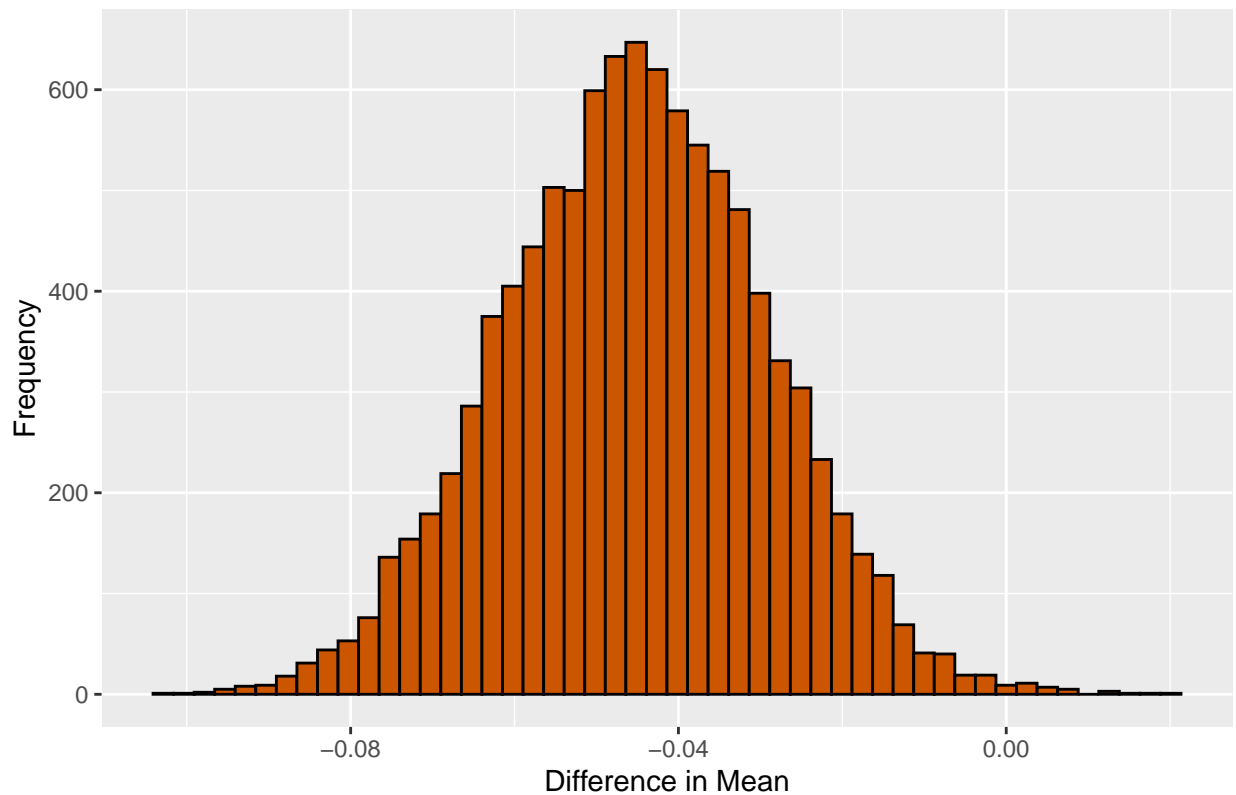

```
##       name      lower      upper level    method    estimate
## 1 diffmean -0.07647577 -0.01438238  0.95 percentile -0.04552049
```

Conclusion

The confidence interval for the difference in mean gas prices between those in low versus high income groups (grouped by the mean of incomes) is (-7.65 cents, -1.44 cents) with 95% confidence. Since the interval is negative, we can reject the null hypothesis. Gas prices seem to correlate with income level, with higher prices in higher-income areas based on this analysis. However if grouped by median of incomes this might differ.
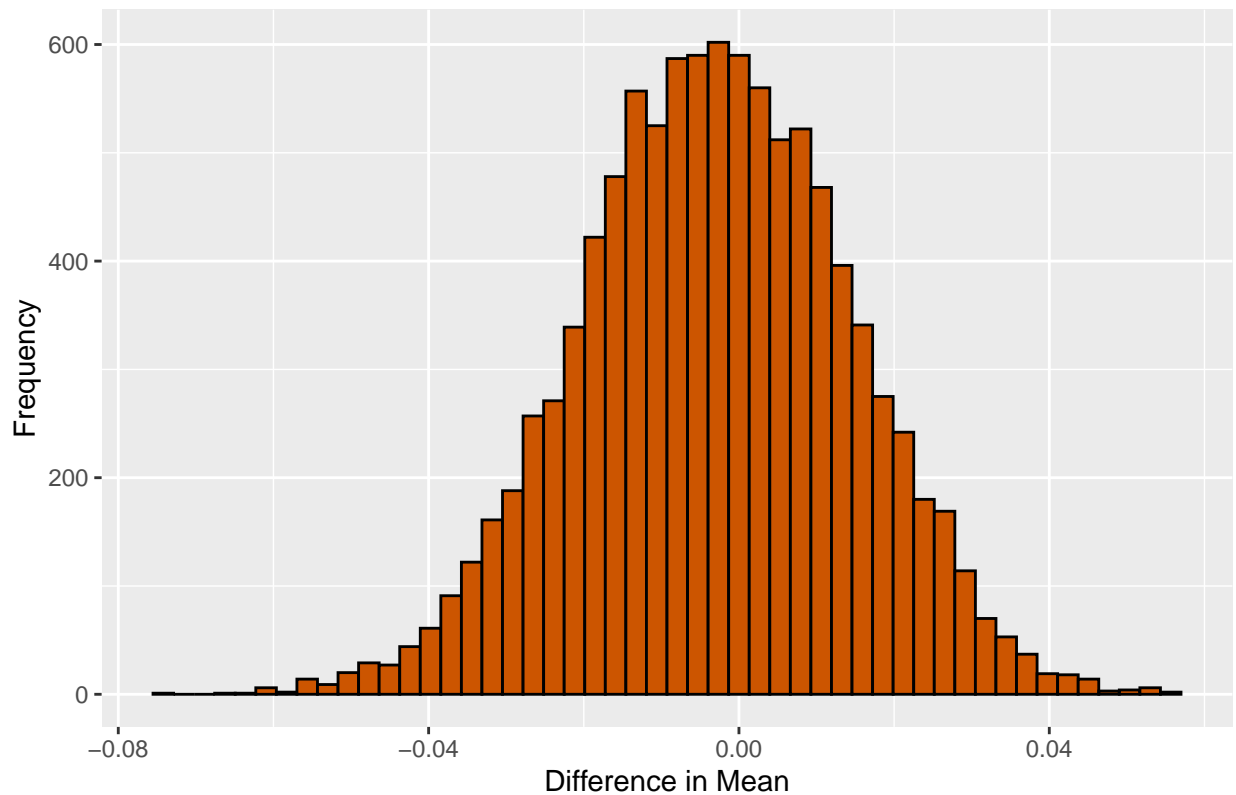
Theory C:

Claim:

Gas stations at stoplights charge more.

Evidence:

```
##        N        Y
## 1.866316 1.863016
```

Bootstrapped Distribution of Difference in Mean Gas Prices


```
##      name      lower      upper level    method     estimate
## 1 diffmean -0.03735241 0.02979502  0.95 percentile -0.003299916
```

Conclusion:

The confidence interval for the difference in mean gas prices between those with and without a stoplight is (-3.74 cents, 2.98 cents) with 95% confidence. Since the interval includes zero, we cannot reject the null hypothesis. The presence of a stoplight does not appear to have a significant effect on the mean gas prices based on this analysis.
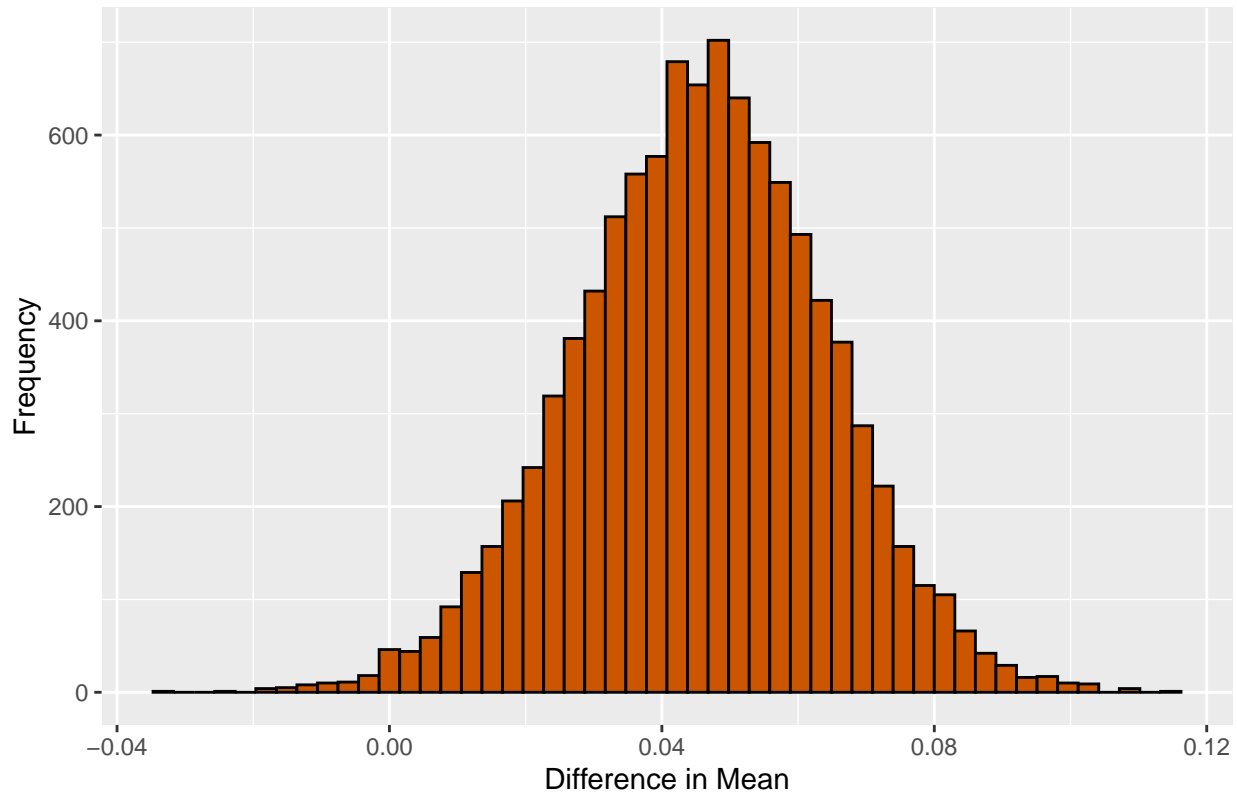
Theory D:

Claim:

Gas stations with direct highway access charge more.

Evidence:

```
##        N        Y
## 1.854304 1.900000
```

Bootstrapped Distribution of Difference in Mean Gas Prices


```
##       name        lower      upper level     method  estimate
## 1 diffmean 0.009004237 0.08125467  0.95 percentile 0.0456962
```

Conclusion:

The confidence interval for the difference in mean gas prices between those with direct highway access versus those without is (.90 cents, 8.13 cents) with 95% confidence. Since the interval is positive (does not include zero), we can reject the null hypothesis. This indicates that, on average, gas stations with direct highway access have higher prices than those without.
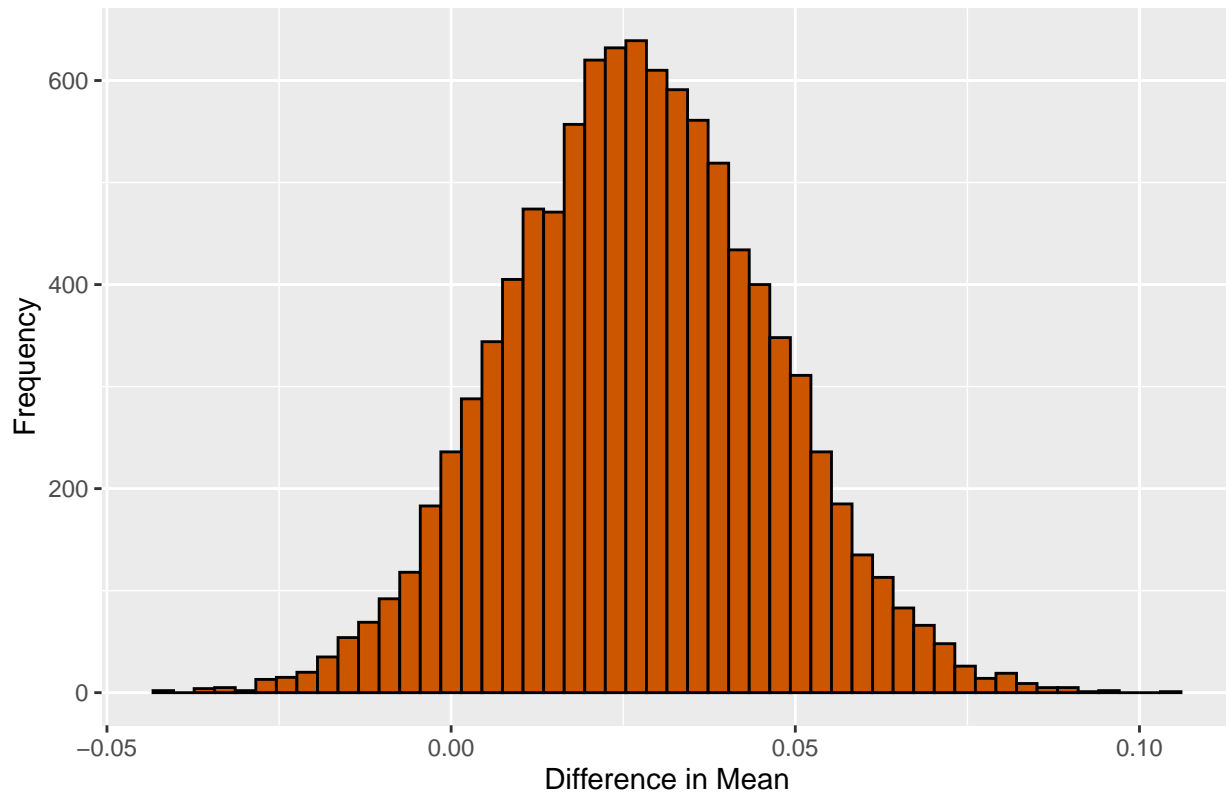
Theory E:

Claim:

Shell charges more than all other non-Shell brands.

Evidence:

Bootstrapped Distribution of Difference in Mean Gas Prices


```
##      name        lower      upper level     method    estimate
## 1 diffmean -0.009320642 0.06526881  0.95 percentile 0.02740421
```

Conclusion:

The confidence interval for the difference in mean gas prices between Shell and any other brand is (-.93 cents, 6.53 cents) with 95% confidence. Since the interval includes zero, we cannot reject the null hypothesis. Whether or not the it is a Shell gas station does not appear to have a significant effect on the mean gas prices based on this analysis.