# Homework 4
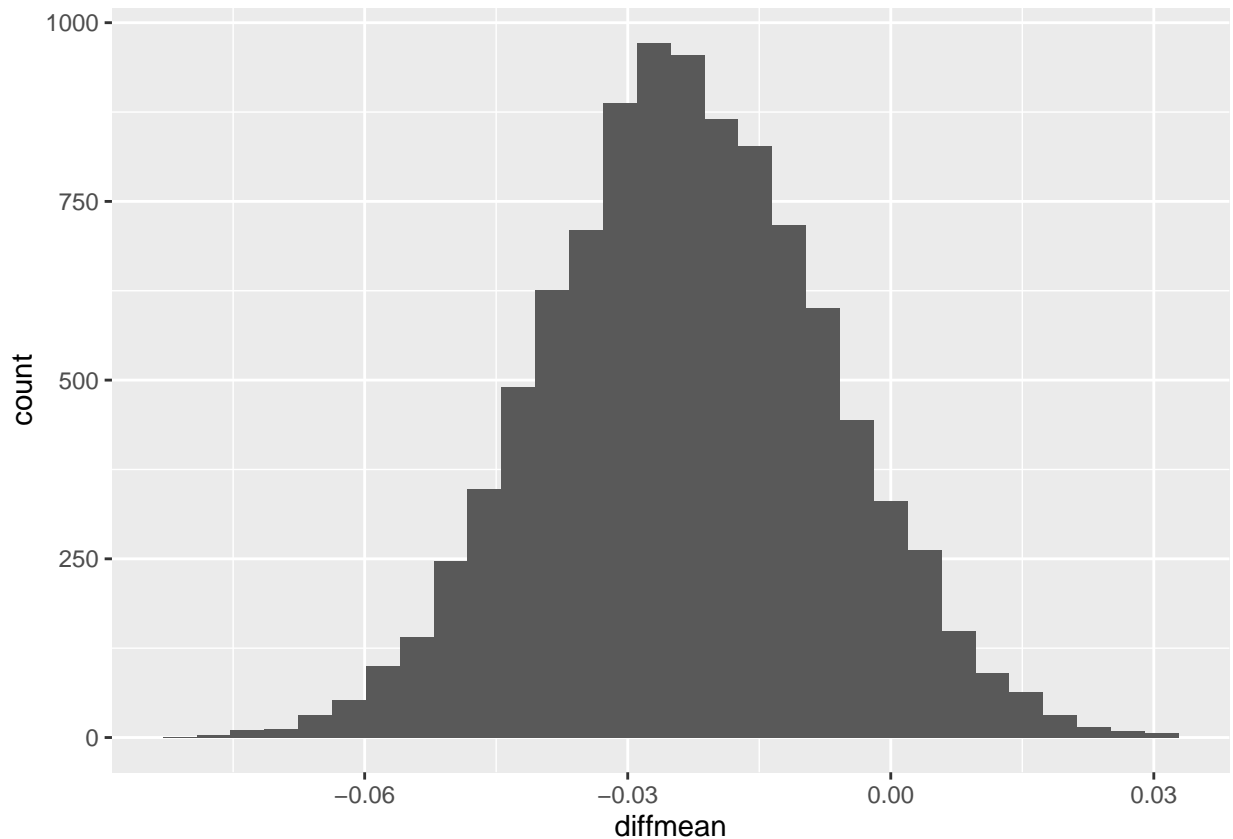
## Aditya Prasad - arp4759

### 2024-02-20

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



```
##       name      lower      upper level      method     estimate
## 1 diffmean -0.05496507 0.008424789  0.95 percentile -0.02348235
```

Evidence: We are 95 percent confident that mean difference between price of gas at gas stations with competition and without competition is between -0.05 and 0.007.

Conclusion: Because 0 is in the interval, it can be concluded that the mean difference between price of gas at gas stations with competition and without competition is insignificant, proving the theory incorrect with a 95% confidence level.

Claim: The richer the area, the higher the gas prices.

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```
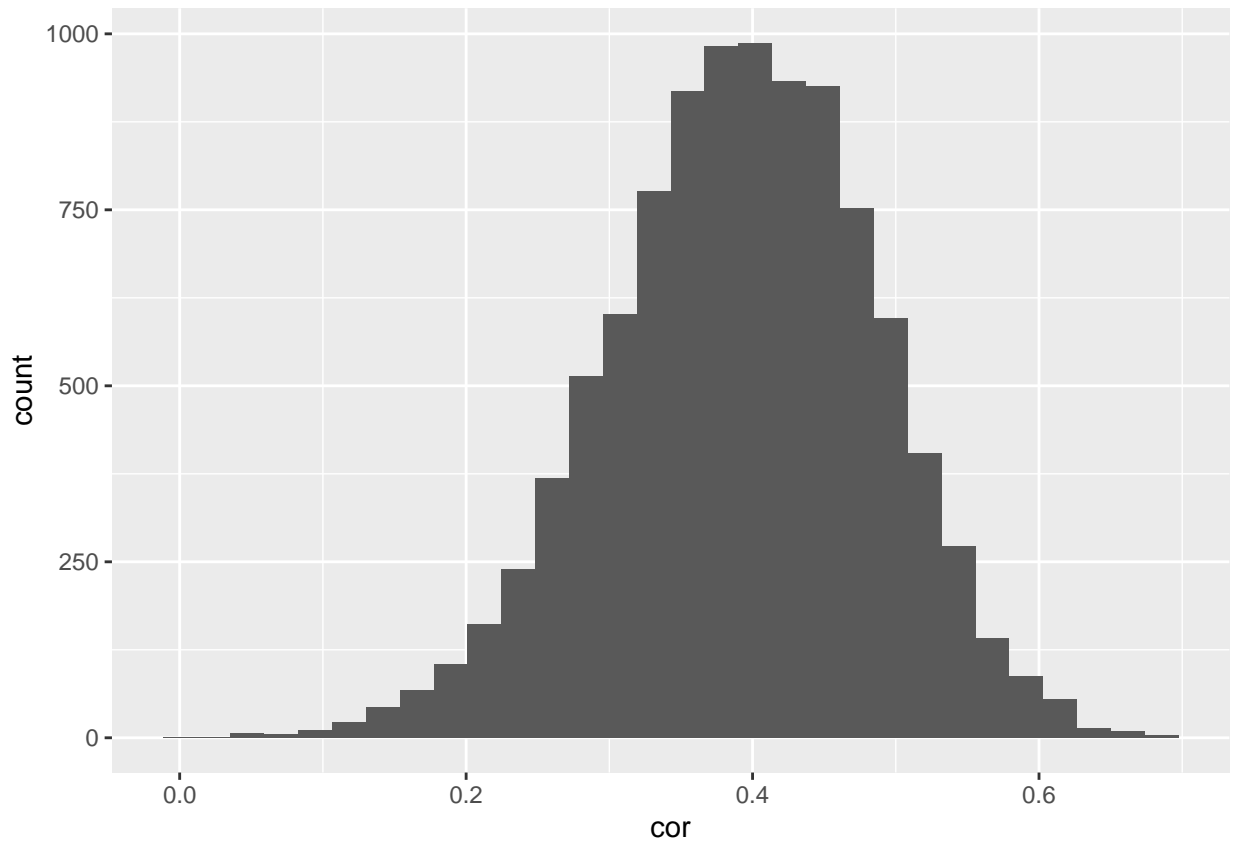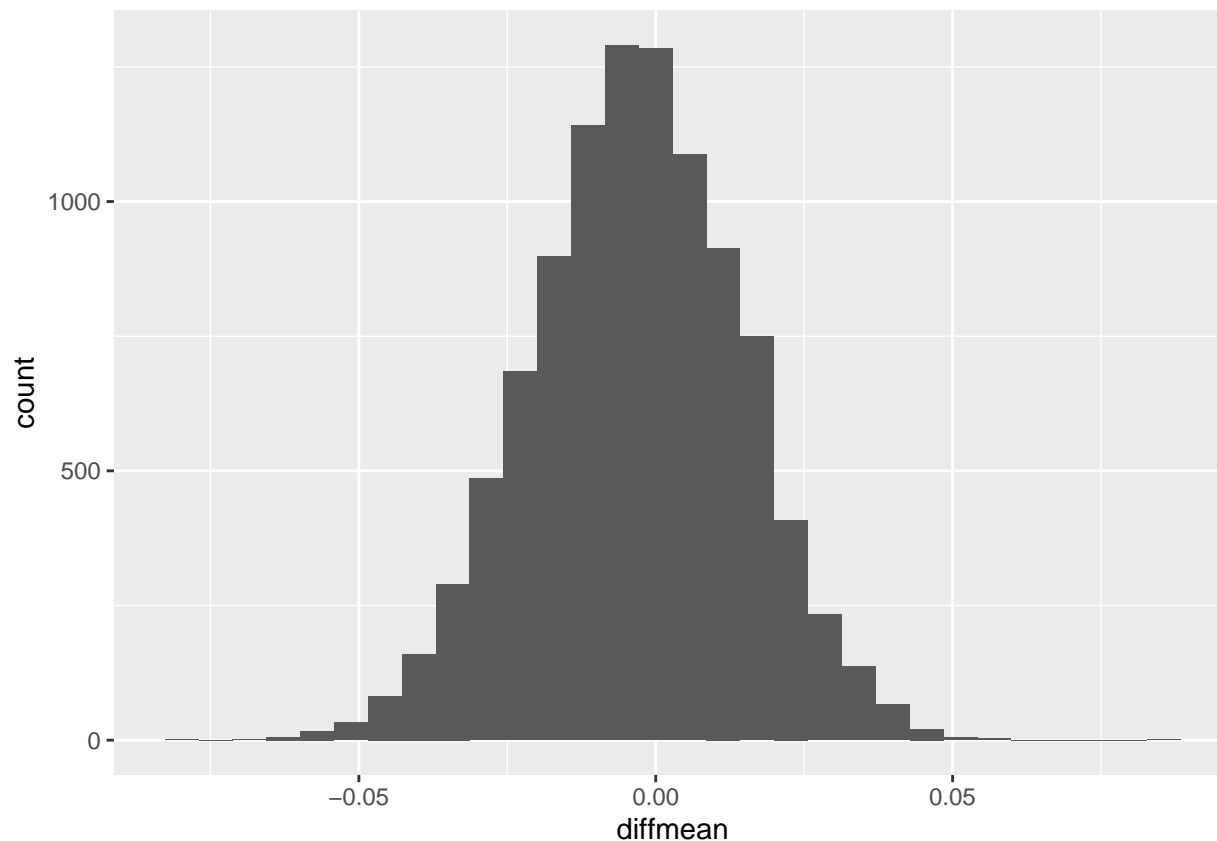


```
##   name    lower    upper level    method  estimate
## 1  cor 0.198394 0.5643546  0.95 percentile 0.3961546
```

Evidence: We are 95% confident that the correlation value between price of gas and income is between 0.197 and 0.566.

Conclusion: Because 0 is not a part of this confidence interval and since moderately positive strong correlation values are included in the interval, that means we can conclude that there is a positive correlation between income in an area and gas prices, estimated at around 0.4.

Claim: Gas stations at stoplights charge more.

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```
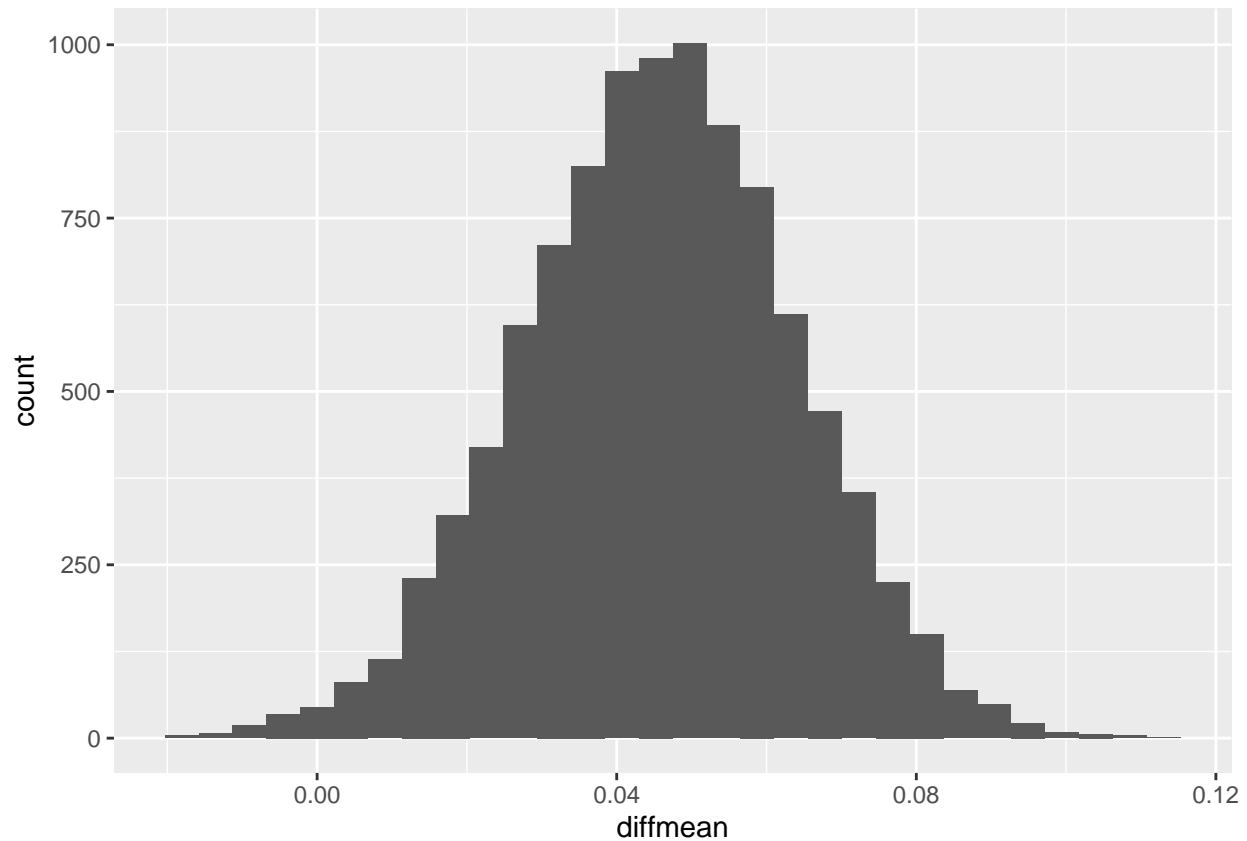
```
##      name        lower      upper level    method      estimate
## 1 diffmean -0.03828556 0.03093106  0.95 percentile -0.003299916
```

Evidence: We are 95% confident that the mean difference between price of gas at locations with stoplights and locations without stopsigns is between -0.03 and 0.03.

Conclusion: Because 0 is in the confidence interval, the mean difference between price of gas at locations with stoplights and locations without stop signs is statistically insignificant, proving that the theory is incorrect.

Claim: Gas stations with direct highway access charge more.

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```
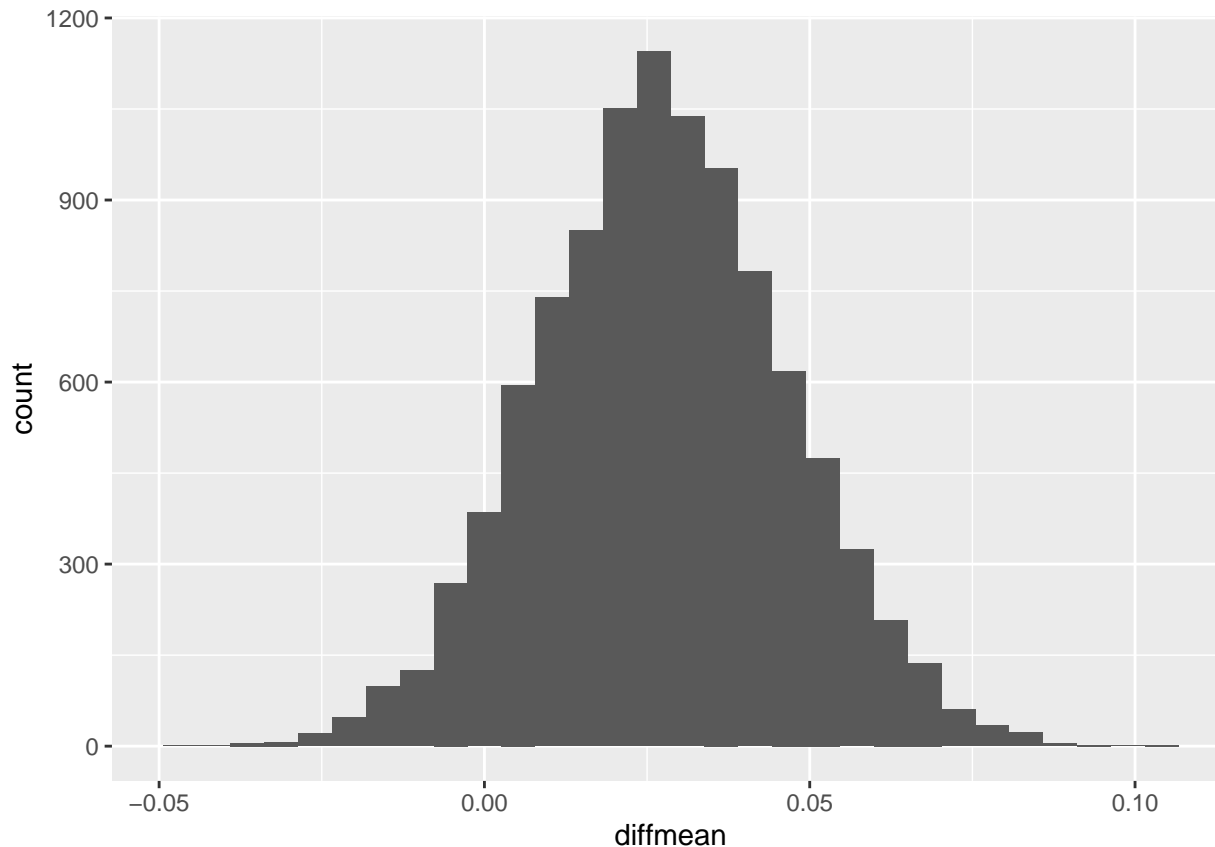
```
##        name        lower       upper  level      method   estimate
## 1 diffmean 0.009485039 0.08058348   0.95 percentile 0.0456962
```

Evidence: We are 95% confident that the mean difference between price of gas with direct access to highways and no direct access to highways is between 0.008 and 0.08.

Conclusion: Even though the interval does not contain 0, the mean difference is practically insignificant since the values of the confidence interval are so small, they are almost 0. This means there is practically no proof that Gas stations with direct highway access charge more.

Claim: Shell charges more than all other non-Shell brands.

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
##       name       lower     upper level     method   estimate
## 1 diffmean -0.009601366 0.0655976  0.95 percentile 0.02740421
```

Evidence: We are 95% confident that the mean difference between price of gas at Shell vs. non-Shell brands is between -0.009 and 0.06.

Conclusion: Because 0 is in the confidence interval, we are 95 percent confident that the mean difference between price of gas at Shell vs. non-Shell brands is statistically insignificant, proving the theory wrong.

```
##   name    lower    upper level     method estimate
## 1 mean 26261.28 31843.18  0.95 percentile 29888.84
```

We are 95% confident that the mean mileage of 2011 63 AMG's hitting the used car market is between 26277.68 and 31764.48 miles.

```
##      name     lower     upper level     method  estimate
## 1 prop_TRUE 0.4164071 0.4527518  0.95 percentile 0.4357909
```

We are 95% confident that the proportion of 2014 S-Class 550's that were painted black is between 0.417 and 0.453.

Question: Who makes people happier: Ed or Earl? Is there evidence that one show consistently produces a higher mean Q1_Happy response among viewers?

```
##       name      lower     upper level     method   estimate
## 1 diffmean -0.3993907 0.1017363  0.95 percentile -0.1490515
```

5

Approach: First, I filtered the data to both shows and implemented the bootstrapping method to resample the data 10,000 times and find the difference in means. Then, I used the confint function to output the confidence interval of the mean difference.

Results: We are 95 percent confident that the mean difference of the happiness rating of the shows "Living with Ed" and "My name is Earl" is between -0.39 and 0.101.

Conclusion: Because 0 is in the confidence interval, the mean difference of the happiness rating of the shows "Living with Ed" and "My name is Earl" is statistically insignificant. This means that there is no sufficient evidence that proves one show consistently produces a higher mean Q1_Happy response among viewers.

Question: Consider the shows "The Biggest Loser" and "The Apprentice: Los Angeles." Which reality/contest show made people feel more annoyed? Is there evidence that one show consistently produces a higher mean Q1_Annoyed response among viewers?

```
## # A tibble: 2 x 2
##   Show                        mean_Q1_annoyed
##   <chr>                                 <dbl>
## 1 The Apprentice: Los Angeles            2.31
## 2 The Biggest Loser                      2.04


##      name      lower      upper level     method  estimate
## 1 diffmean -0.5231384 -0.01576133  0.95 percentile -0.270997
```

Approach: First, I filtered out all shows except the reality shows: "The Biggest Loser" and "The Apprentice: Los Angeles". I calculated the mean rating of the response to Q1_Annoyed for both shows. Next I implemented the bootstrapping method where I calculated the mean difference of the responses to Q1_Annoyed for both shows by resampling 10,000 times. Finally, I calculated the 95% confidence interval using the confint function.

Results: We are 95% confident that the mean difference of the response to Q1_Annoyed for "The Biggest Loser" and "The Apprentice: Los Angeles" is between -0.52 and -0.02. The mean of the rating response to Q1_Annoyed for "The Apprentice: Los Angeles" is 2.3 and the mean rating response to Q1_Annoyed for "The Biggest Loser" is 2.03.

Conclusion: Because 0 is not in the confidence interval, this means that mean difference of the response to Q1_Annoyed for "The Biggest Loser" and "The Apprentice: Los Angeles" is statistically significant. In other words, there is evidence that one show consistenly had a higher mean rating for Q1_Annoyed than the other which was "The Apprentice: Los Angeles".

Question: Consider the show "Dancing with the Stars." What proportion of American TV watchers would we expect to give a response of 4 or greater to the "Q2_Confusing" question?

```
##        name      lower     upper level     method   estimate
## 1 prop_TRUE 0.03867403 0.1160221  0.95 percentile 0.06077348
```

Approach: First, I filtered out all the shows except "Dancing with the Stars". Then, I implemented the bootstrapping method, where I calculated the proportion of American TV Watchers who would give Q2_Confusing at least a 4 rating by resampling this data 10,000 times. Finally, I calculated the 95% confidence interval for this proportion.

Results: We are 95% confident that the proportion of American TV Watchers who would give Q2_Confusing at least a 4 rating is between 0.038 and 0.11. The estimated proportion is around 0.07.

Conclusion: Because 0 is not included in this confidence interval, proportion of American TV Watchers who would give Q2_Confusing at least a 4 rating is statistically significant, which is estimated to be around 0.072.

Question: What is the revenue ratio difference between the treatment and control DMAs and provide a 95% confidence interval for the difference? Is the revenue ratio the same in the treatment and control groups?

```
## # A tibble: 2 x 2
##   adwords_pause mean_rev_ratio
##           <int>          <dbl>
## 1             0          0.949
## 2             1          0.897
```

```
##       name       lower        upper level     method     estimate
## 1 diffmean -0.09051878 -0.01322493  0.95 percentile -0.05228145
```

Approach: I implemented the bootstrapping technique, where I calculated the difference in means between the treatment and control group for the revenue ratio. I then calculated the 95% confidence interval using confint.

Results: We are 95% confident that the mean difference of revenue ratio between the treatment and control DMAs is between -0.09 and -0.013. The estimated difference is between -0.05.

Conclusion: Because 0 is not included in the confidence interval, mean difference of revenue ratio between the treatment and control DMAs is statistically significant. This means that the control group consistently produced a higher revenue ratio of 0.05 higher than the treatment group.