Homework 4

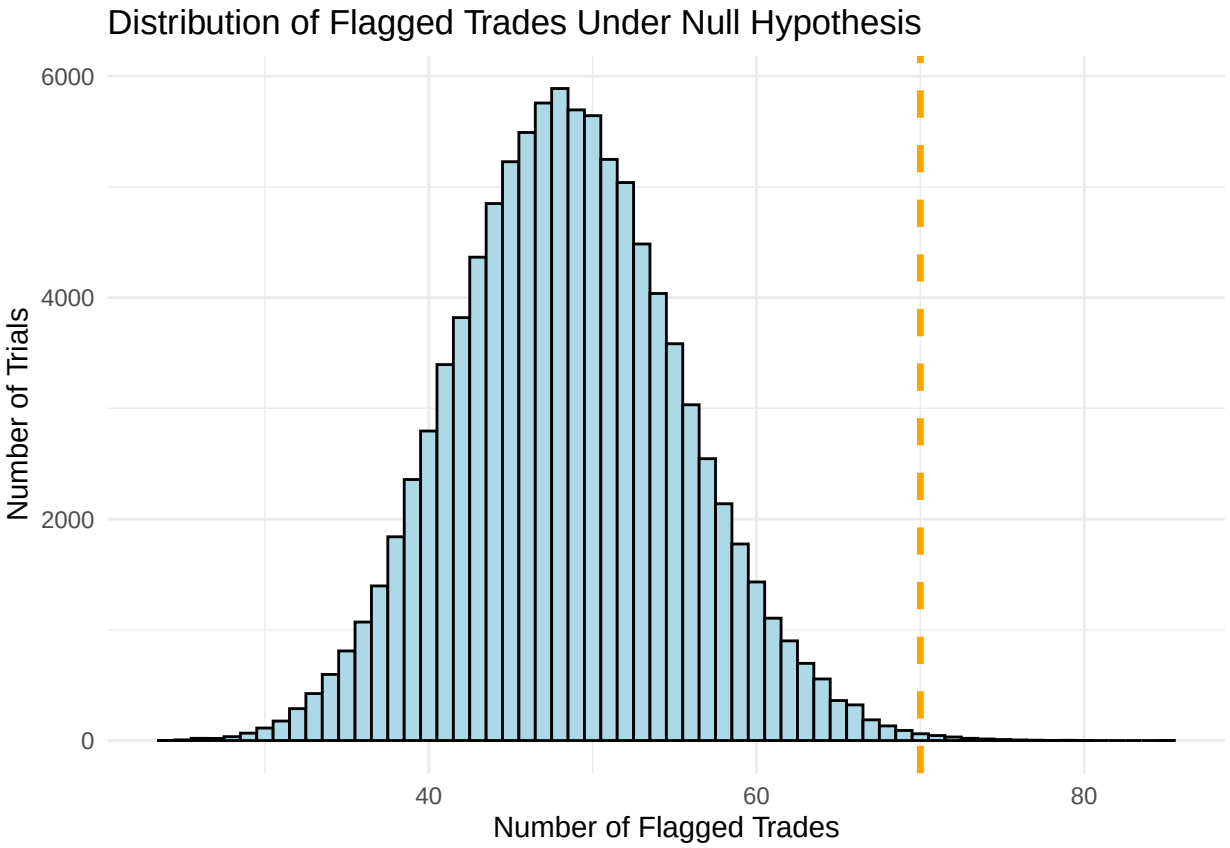
Arun Mahadevan Sathia Narayanan

2025-02-18

GitHub Link:
To GitHub

GitHub Link (Text Format):
https://github.com/arunmsn/SDS315/tree/main/Week_06/Homework_4

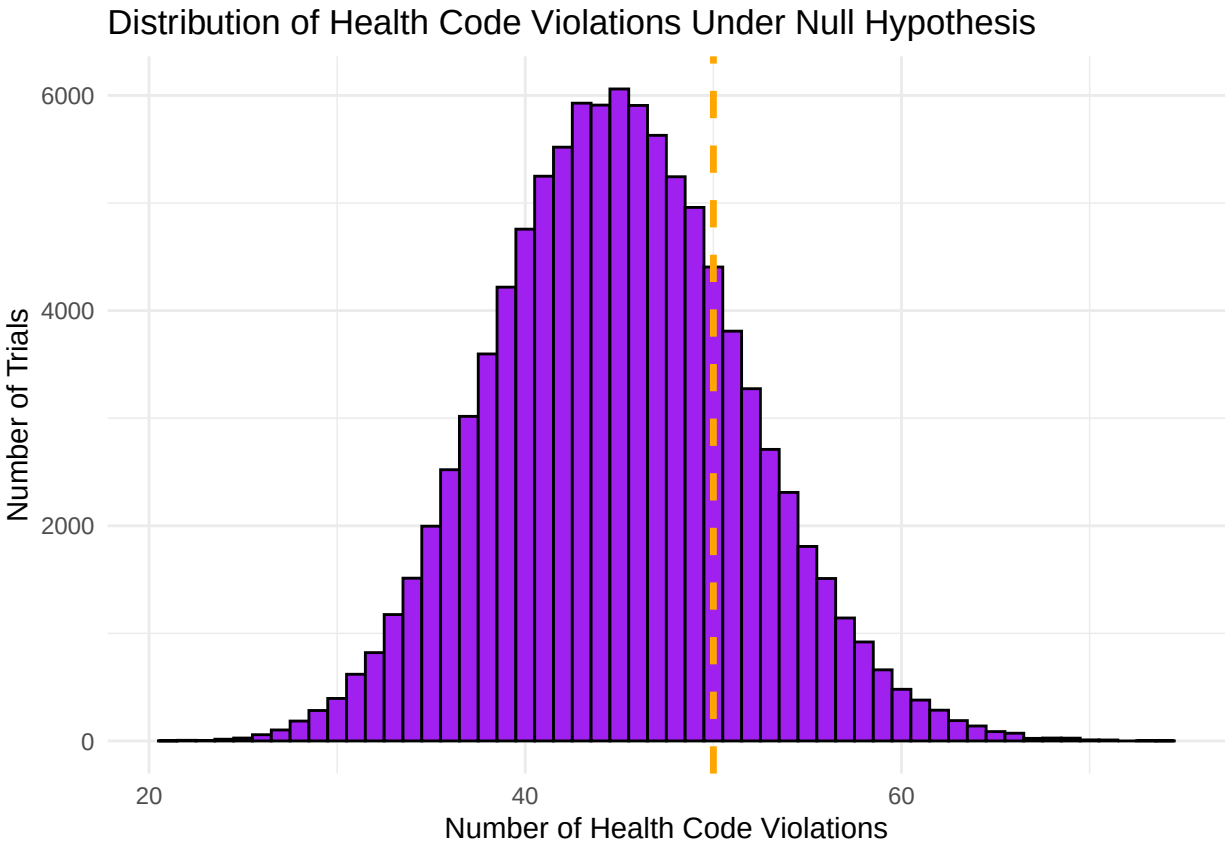
Problem 1 - Iron Bank




```
## P-value: 0.00194
```

Null Hypothesis: the flagged trades at Iron Bank follow the usual 2.4% flagging rate.
Test Statistic: 70 (the observed value of flagged trades)
The above distribution was made from 100,000 Monte Carlo Bootstrap Trials. The vertical line is where 70 flagged trials are. As seen from the result above, the p-value (~ `0.002`) is way less than the statistically significant value of p = 0.05 (AKA the flagging rate here is higher than the expected 2.4% (~ 48 flags)), suggesting that the flagged trades occur at a significantly higher rate than expected by chance alone.

Problem 2 - Health Inspections

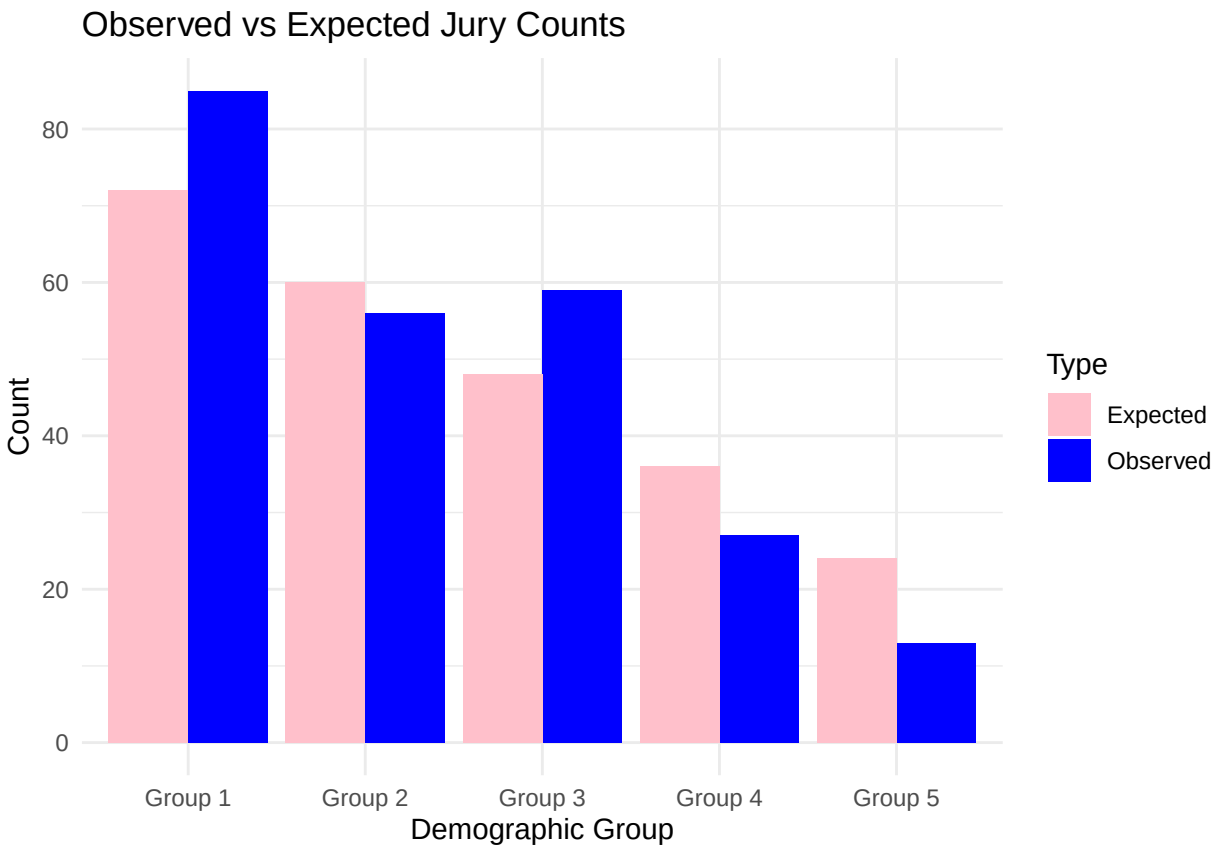



```
## P-value: 0.24282
```

Null Hypothesis: the health code violations follow the usual 3% violation rates.
Test Statistic: 50 (the observed value of health code violations)
The above distribution was made from 100,000 Monte Carlo Bootstrap Trials. The vertical line is where 50 health code violations are. As seen from the result above, the p-value (~ `0.24`) is more than the statistically significant value of p = 0.05 (which indicates that the trials were consistent with the null hypothesis), meaning there is no strong evidence of an unusually high rate of health code violations. Also, when doing a count comparison, 3% of 1500 is 45, and having 50 violations is not statistically significant in comparison.

Problem 3 - Evaluating Jury Selection for Bias

```
##     Group Observed Expected
## 1 Group 1       85       72
## 2 Group 2       56       60
## 3 Group 3       59       48
## 4 Group 4       27       36
## 5 Group 5       13       24
```




```
## Chi-Squared Test: 12.42639
```

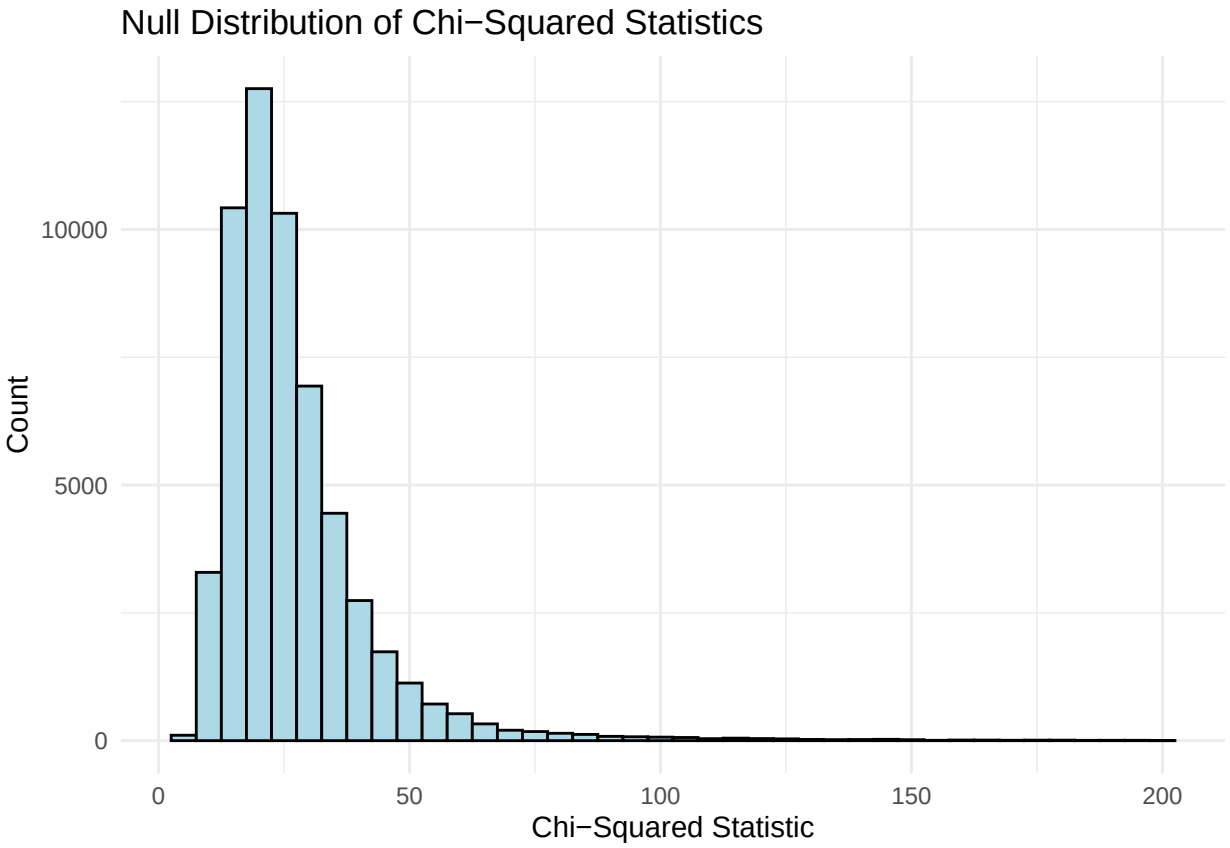
```
## Degrees of Freedom: 4
```

```
## p-value: 0.01398
```

Null Hypothesis (H0): The jurors on the panel follow the same distribution as the county's eligible jury pool, meaning the judge's selection process does not systematically alter group representation.
Test Statistic (T): Using the Chi-Squared Test, we got the observed chi-squared value to be `12.426`.
P(T | H0): This is a p-value calculation. Using a bootstrap sample 100,000 times, I was able to produce a distribution of chi-squared values. With the distribution, it was just a simple comparison to what proportion of the distribution was greater than or equal to the observed chi-squared statistic, and ended up with a p-value ~`0.01`. This p-value is less than p = 0.05, which shows that we should reject the null hypothesis. This shows that the judge's selection failing to alter group representation is not due to chance alone.

Problem 4 - LLM Watermarking

4 - Part A



From the above distribution, we can see that the most common Chi-Squared value is 20 (each bar has a width of 5).

4 - Part B

```
##                    Sentence p_value
## 1  Feeling vexed after  ...   0.009
## 2  Despite the challeng ...   0.059
## 3  The museum's new exh ...   0.076
## 4  The committee review ...   0.084
## 5  The chef demonstrate ...   0.328
## 6  The students gathere ...   0.484
## 7  He carefully examine ...   0.489
## 8  She opened the book  ...   0.513
## 9  Despite the heavy ra ...   0.926
## 10 They watched the sun ...   0.988
```

The sentence watermarked by an LLM is the following:
`"Feeling vexed after an arduous and zany day at work, she hoped for a peaceful and quiet evening at home, cozying up after a quick dinner with some TV, or maybe a book on her upcoming visit to Auckland."`
This is due to this sentence scoring a p-value of `0.009`, which is very low compared to the base p = 0.05, indicating that this was the work of an LLM and not a human.