Hw_4

Pranita

2025-02-20

Name: Pranita Chaudhury

UT EID: pc28377

Github: https://github.com/PranitaChau/Hw_4.git

Problem 1

Simulated Distribution of Flagged trades
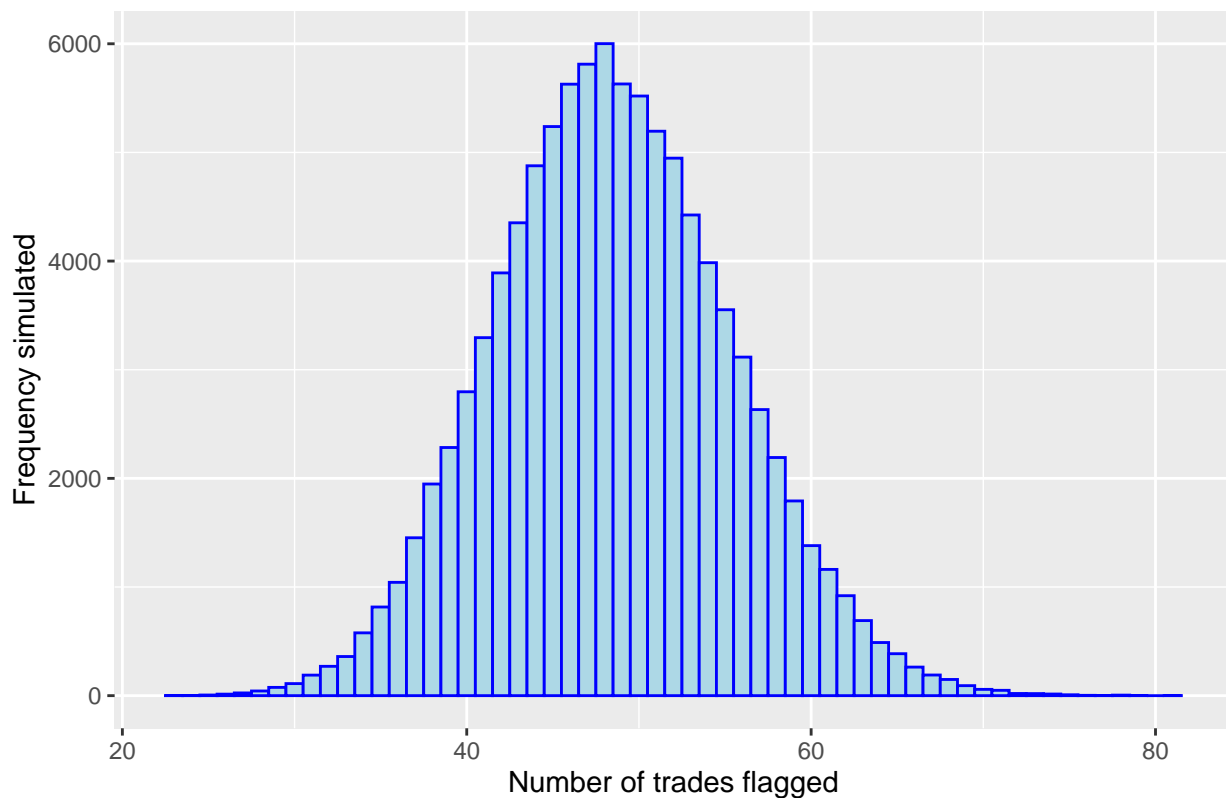

```
## [1] 0.0182
```

Null Hypothesis: The Iron Bank is not flagged more than the baseline rate. Test Statistic: The number of trades that are flagged belonging to Iron Bank (70). (Graph shown above) P-value: 0.018 Conclusion: Since the p value is very low there is no evidence supporting the null hypothesis, and we reject the hypothesis that the number of flagged trades at Iron Bank is under the baseline trade flagging rate of 2.4%.

Problem 2

Distribution of Health code violations
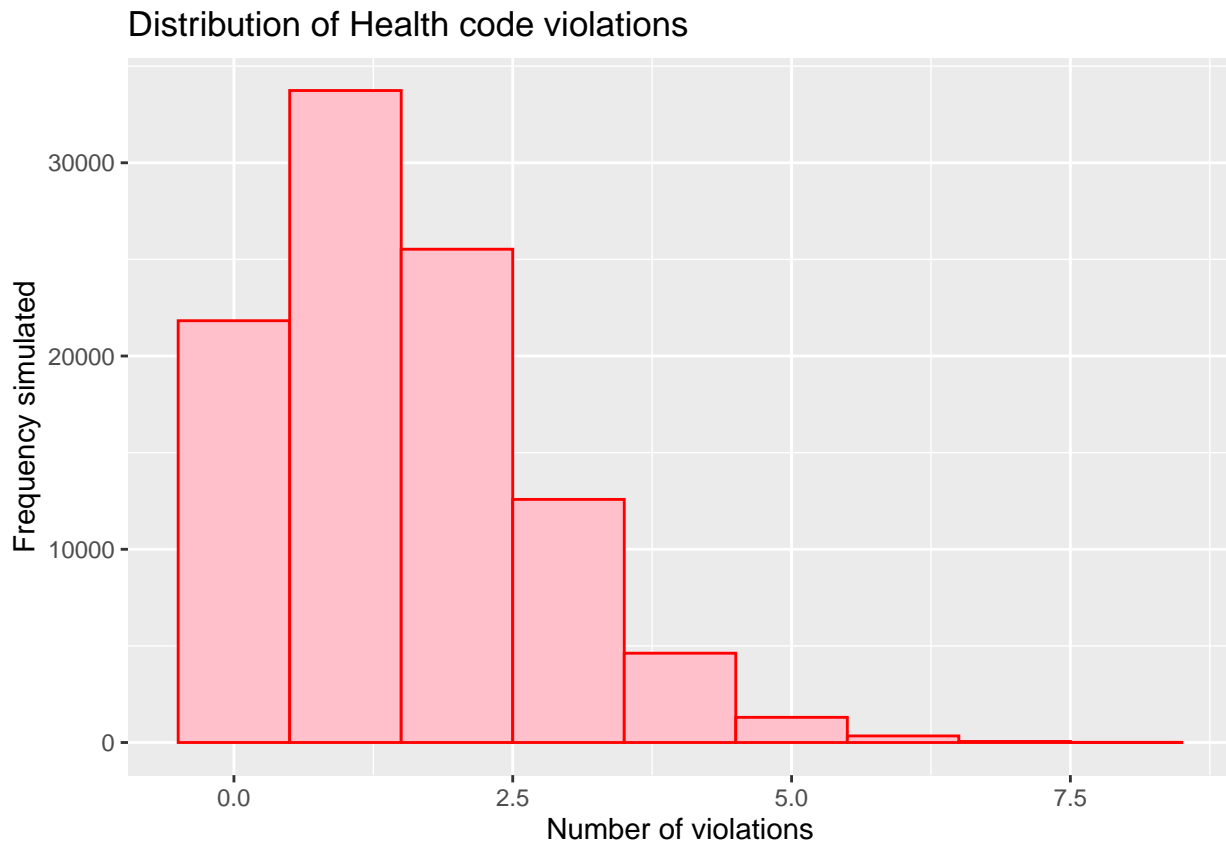

```
## [1] 1e-04
```

Null Hypothesis: The local chain Gourmet Bites is not receiving higher than usual health code violations compared to the city wide average of 3%. Test Statistic: The number of health code violations reported from Gourmet Bites (50). (Graph shown above) P-value: 2e-4 Conclusion: Since the p value is very low there is no evidence supporting the null hypothesis, and we reject the hypothesis that the Gourmet Bites is not receiving higher than usual health code violations than the citywide average of 3%.

Problem 3

```
## [1] 12.42639
```

Distribution of Chi−squared Statistics
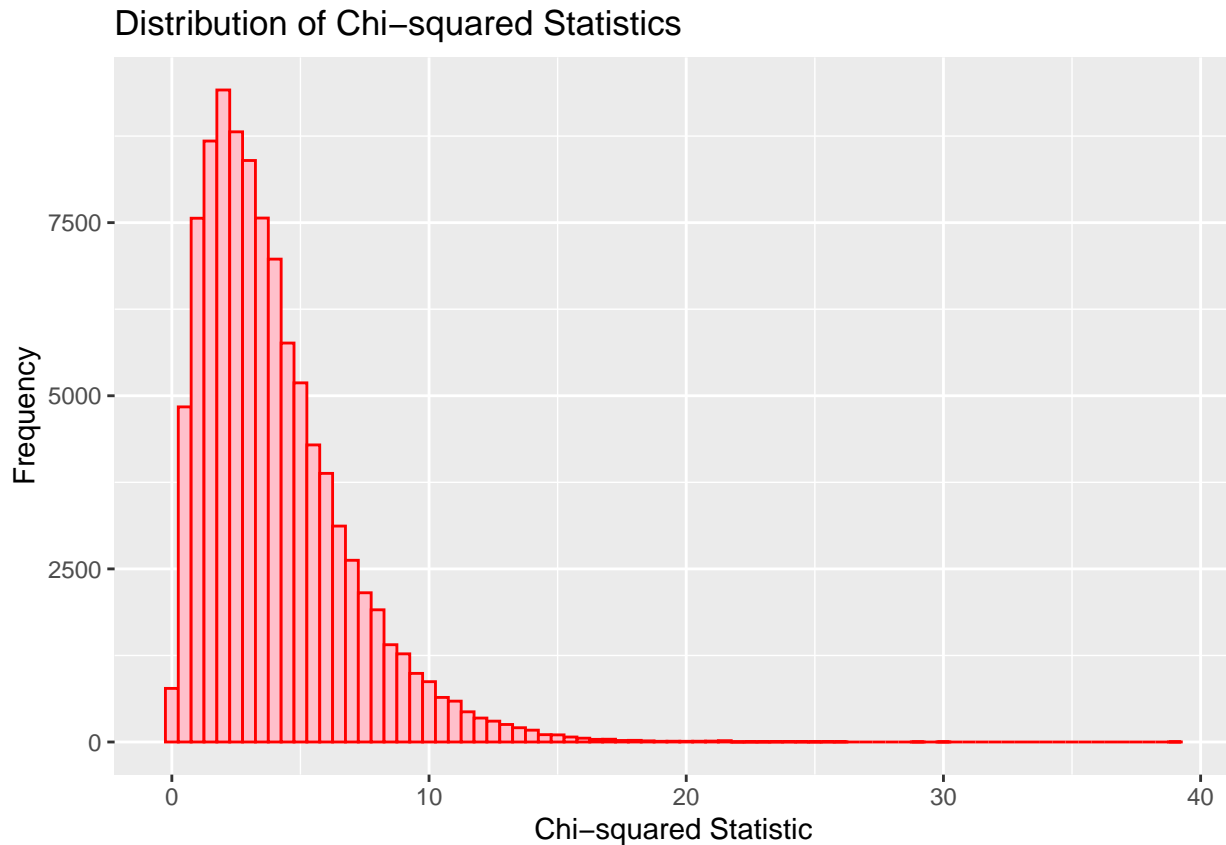

```
## [1] 0.014
```

Ho = The difference in each group of jurors can be explained by chance. Test Statistic = We are using a chi-squared test where the expected counts is 240 multiplied by the proportion for that group (Group 1 = 30%, Group 2 = 25%, Group 3 = 20%, Group 4 = 15%, Group 5 = 10%). We then calculate the p value, which is 0.01462. P(T | Ho) = Since the p value is 0.01462, it is low enough to reject the null hypothesis that the difference in each group of jurors can be explained by chance. Conclusion = Based on the fact hat we rejected the null hypothesis, we can conclude that there is statistical significance to suggest systematic bias in jury selection for this specific judge. Other explanations that may exist to explain this discrepancy include that due to the specific kind of case the judge decided to have more or less of a specific group present or that this area may have a misrepresentation due to the demographics living there. Further investigations could be conducted by comparing this to other judges (other observed counts), increase the same size to more than 20 trials, or analyze patterns across a specific demographic area in order to determine if this discrepancy lies solely with this judge.

Problem 4

Part A

```
## [1] 27.56914
```

Part B

```
## # A tibble: 10 x 1
##    P_Value
##      <dbl>
## 1    0.52
```

```
##  2   0.937
##  3   0.084
##  4   0.496
##  5   0.491
##  6   0.001
##  7   0.335
##  8   0.006
##  9   0.091
## 10   0.066
```

The table shows all the p values in the order the sentences were inputted. Based on the p values given sentence 6 had the lowest p value and was low enough to reject the null hypothesis, therefore we can conclude that sentence 6 was written by an LLM. Sentence 6 was "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."