homework4

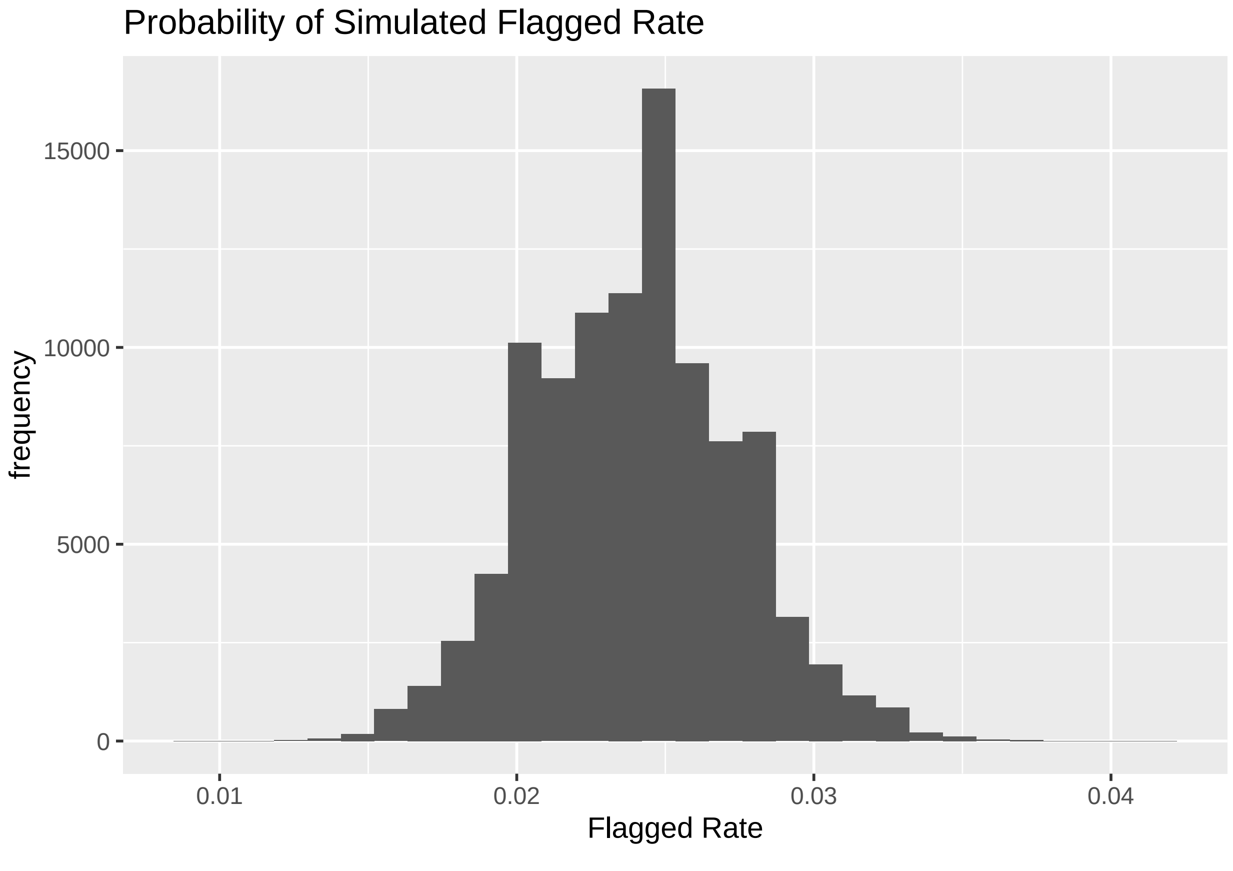
Emma Chung

2025-02-18

github link:

Problem 1: Iron Bank

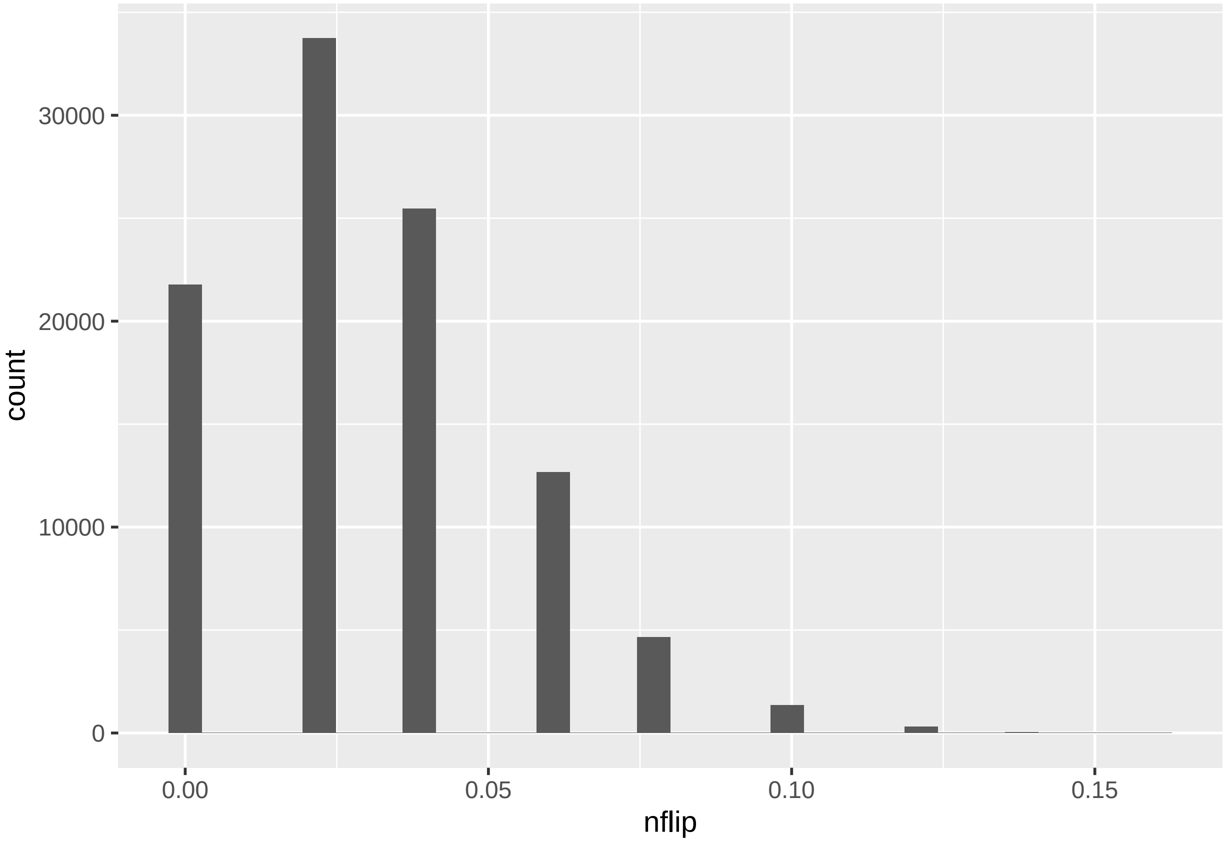
The null hypothesis I am testing here is that true probability of flagged trade is 2.4%. I perform a hypothesis test using the 100000 simulation following the 2.4% probability.



Based on the p value of 0.002, the number is significant enough to reject the null hypothesis. In conclusion, those 70 trades found is inconsistent with the SEC's null hypothesis, it might not just detection error, the baseline rate might be wrong.

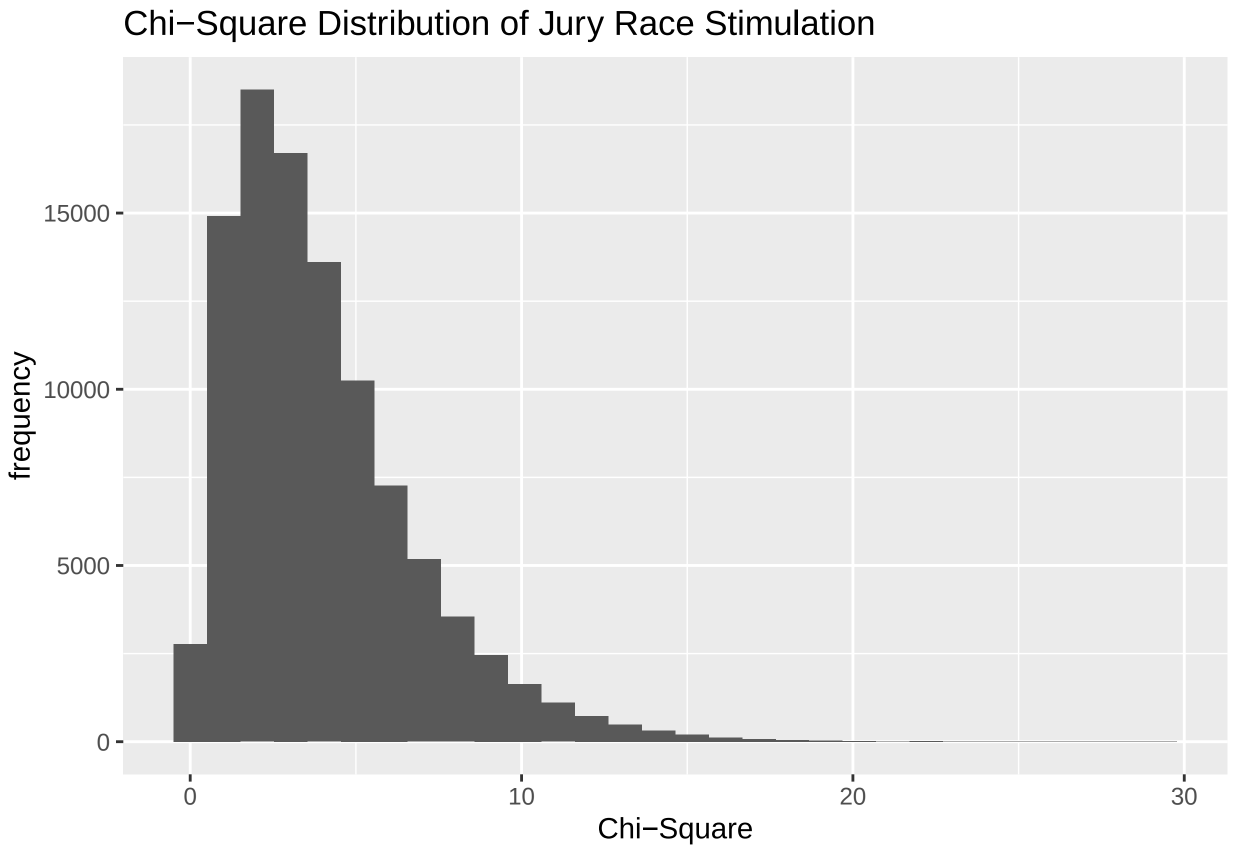
Problem 2: health inspection

The null hypothesis the Health Department is testing Gourmet Bites's rate of health code violations is 3%, not higher than other chains. The test statistics were a 100000 simulation following the 3% probability.



The p value is 0 which is significant enough to reject the null htpothesis, meaning the 8 reports are inconsistent with the hypothesis of baseline being 3%. Therefore we can conclude that the baseline of misreported and is not 3%.

Problem 3: Jury Selection

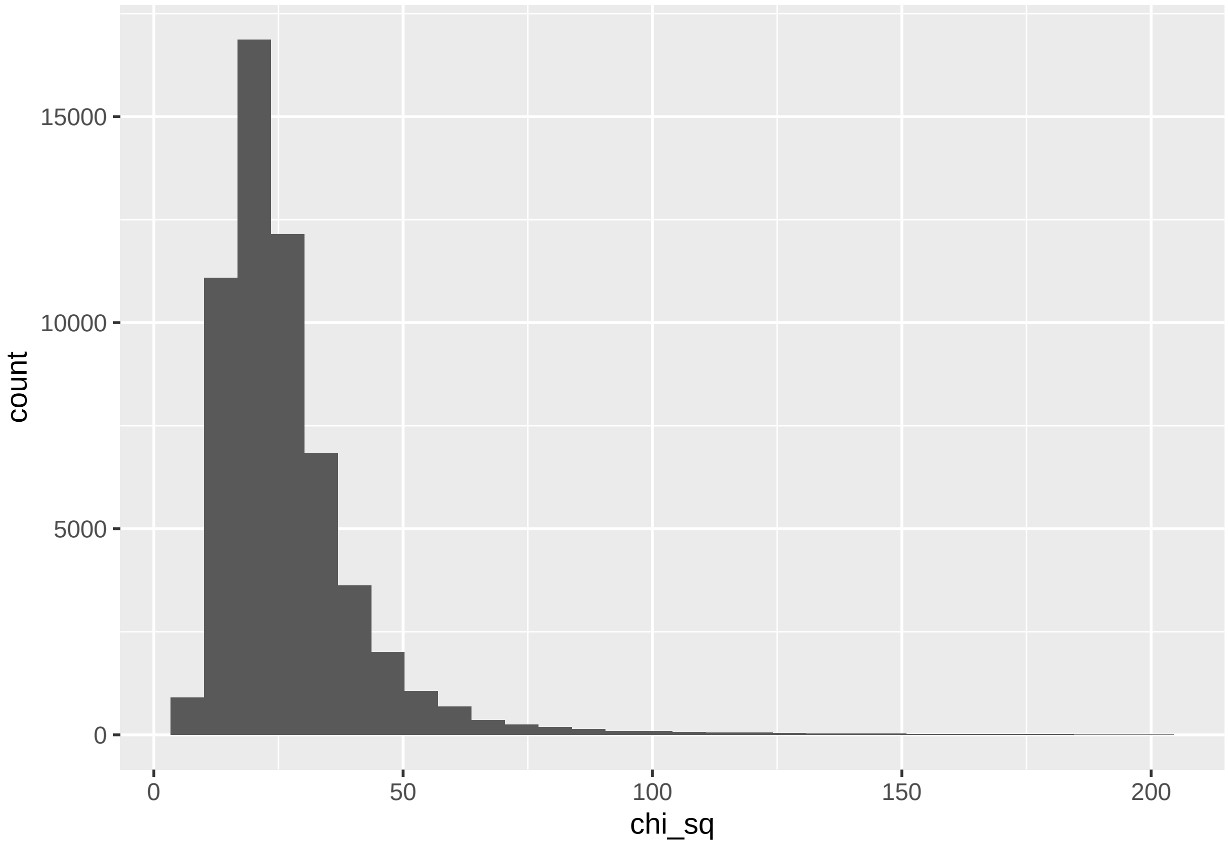


My null hypothesis is that the jury reflect the country's population in terms of race. The test statistic I use is chi-square statistic. I generate 100000 stimulated data and compare with the chi-square of the given data. It has the p value of 0.015, which is significant enough to reject the null hypothesis. In conclusion, the jury selection did not reflect the country's race distribution.

However, it might still be bias due to the random selection of 20 trials, and it could be investigated further by repeatedly stimulate 20 trials and calculate the p value again.

Problem 4: LLM watermark

Part A



The chi-square statistics among human-written sentences is displayed on the above chart.

Part B

```
## # A tibble: 1 x 10
##      p1    p2    p3    p4    p5    p6    p7    p8    p9   p10
##   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1     0     0     0     0     0     0     0     0     0     0
```

The calculation of the p values of the ten sentences was displayed in the above table. Since sentences 6 has a very low p value, I will assume that this sentence was produced by an LLM. Having a low p value means the chi-square statistic of this sentences was extreme and not consistent with other human produced sentences.