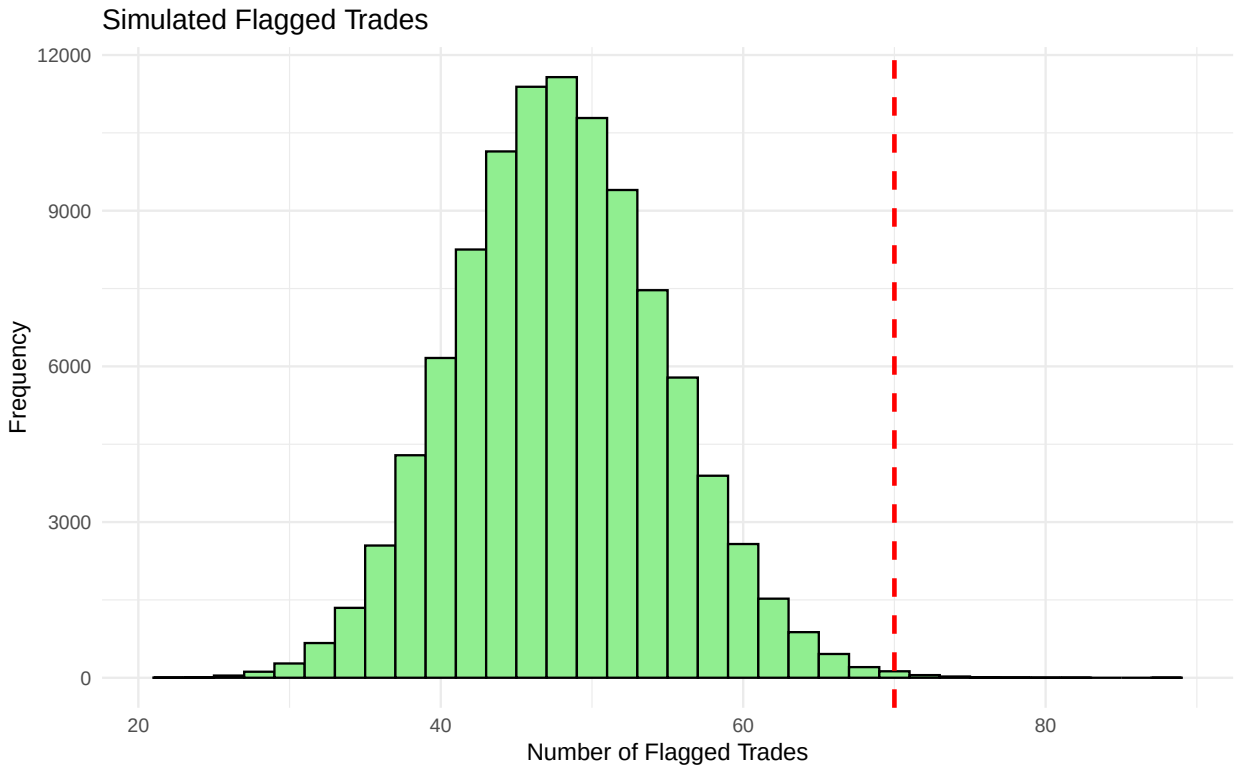
Homework 4

Pranav Belligundu(psb898) - SDS 315 UT Austin

https://github.com/pranav-B21/SDS-315/tree/main/HW%203

Problem 1: Iron Bank




```
## P-value: 0.00214
```

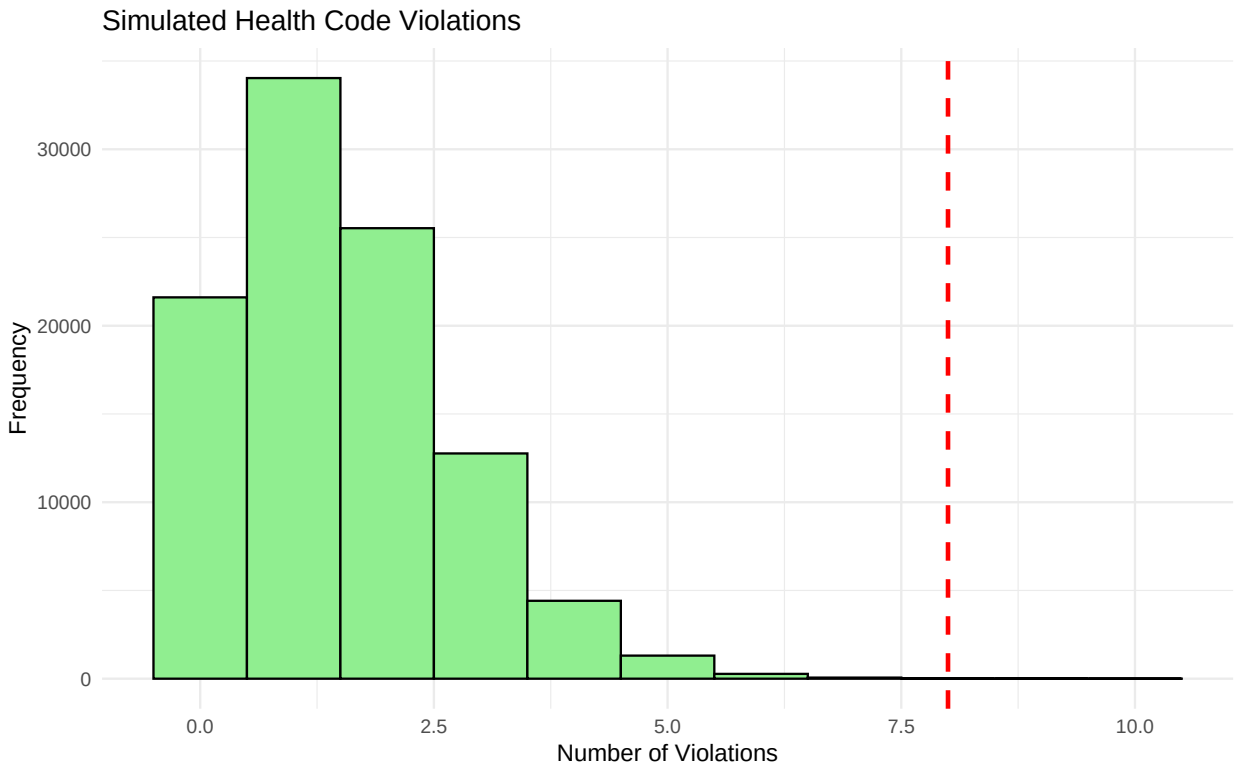
Null Hypothesis: the proportion of flagged trades actually follow the baseline provided of 2.4%

Test statistic: The number of flagged trades(observed 70), p-value test

Simulation details: Simulates 100,000 trials of 2021 trades. Each trade has 2.4% probability of being flagged. Using nflip() from mosaic package for binomial simulations with the given n value and the prob.

Conclusion: since the p_value = 0.00208 $<$ 0.05, we can reject the null hypothesis. This suggests that the trading activity at Iron Bank is not consistent with normal market behavior and the number of flagged trades(70 out of 2021) is way more than actual.

Problem 2: Health Inspections




```
## P-value: 0.00012
```

Null Hypothesis: the true rate of health code violations at Gourmet Bites is equal to the citywide baseline rate of 3%

Test statistic: The number of inspections (observed 50), p-value test

Simulation details: Simulates 100,000 trials of 50 inspections Each inspection has 3% probability of being labeled as a health code violation. Using nflip() from mosaic package for binomial simulations with the given n value and the prob.

Conclusion: since the p_value = 0.00012 $<$ 0.05, we can reject the null hypothesis. This suggests that the rate of health code violations at Gourmet Bites is significantly higher than the citywide baseline rate of 3%. The observed number of violations (8 out of 50 inspections) is highly unlikely to occur due to random chance alone.

Problem 3: Evaluating Jury Selection for Bias

```
##
##  Chi-squared test for given probabilities
##
## data:  observed_counts
## X-squared = 12.426, df = 4, p-value = 0.01445
```

Null Hypothesis: the true rate of health code violations at Gourmet Bites is equal to the citywide baseline rate of 3%

Test statistic: Chi-Square Goodness-of-Fit Test

Simulation Details: chisq.test solves the simulation for us so that there is no need to manually calcualte the expected and original counts. We get a chisq value of 12.43 and the degrees of freedom(df) df = 5 − 1 = 4. Using these 2 values we can determine the p-value for this statistic.

Conclusion: since the p-value = 0.015 < 0.05. There is significant evidence that the distribution of empaneled jurors does not match the county's population. This suggests potential bias.

Problem 4: LLM Watermarking

Sentence ID	P-Value
1	0.879
2	0.513
3	0.981
4	0.866
5	0.981
6	0.132
7	0.964
8	0.981
9	0.981
10	0.513

Given the chi-squared results and the p-value, I believe the sentence that is likely produced by an LLM with a watermark is Sentence 6: "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."

The sentence has the highest chi-squared value of 39.04543. This suggests that the letter frequency distribution deviates more from the expected natural English distribution. The sentence also has a relatively low The p-value for this sentence is 0.132, meaning that only about 13.2% of the naturally occurring sentences in the Brown Corpus have such extreme letter frequency distributions. Even thought it isnt an extreme outlier (< 0.05), it is still much lower than most other sentences.