HW4-SDS315

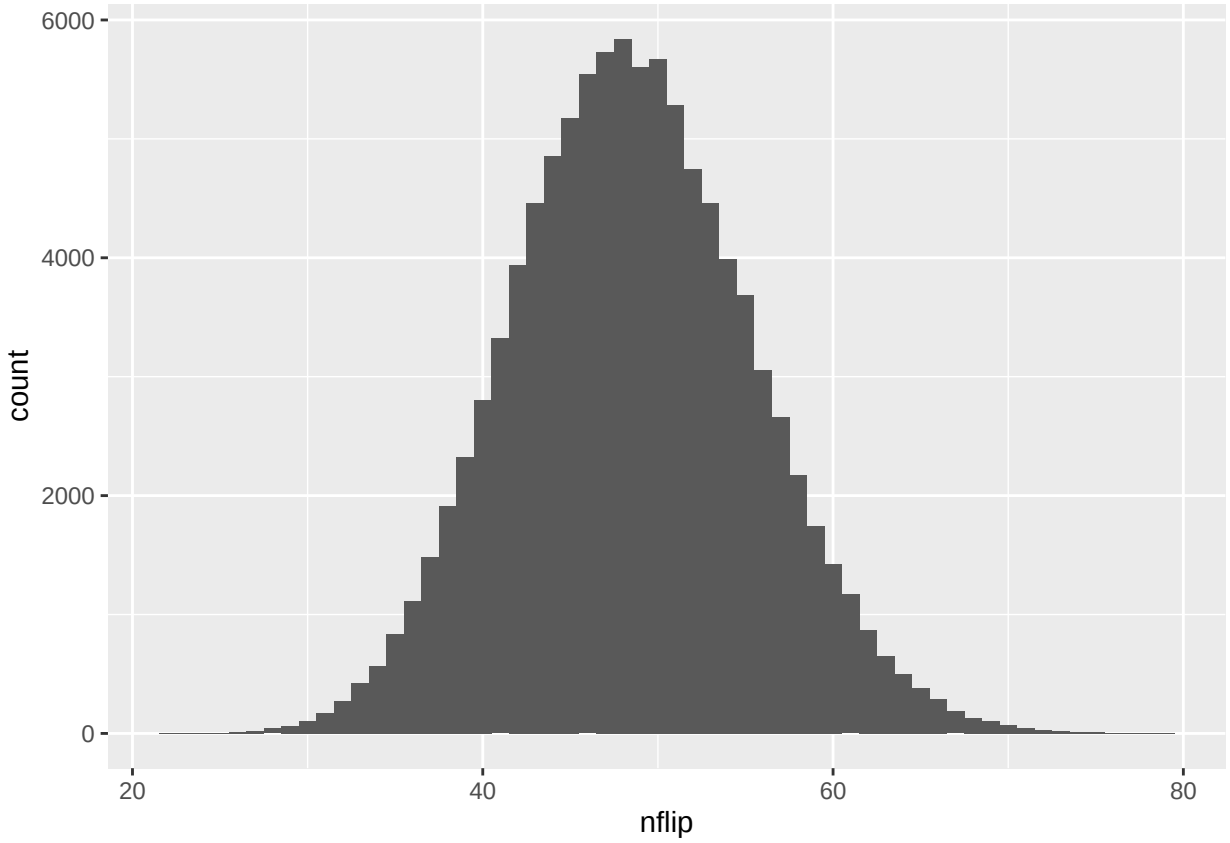
Seraphym Ignacio

2025-02-20

EID: sri346

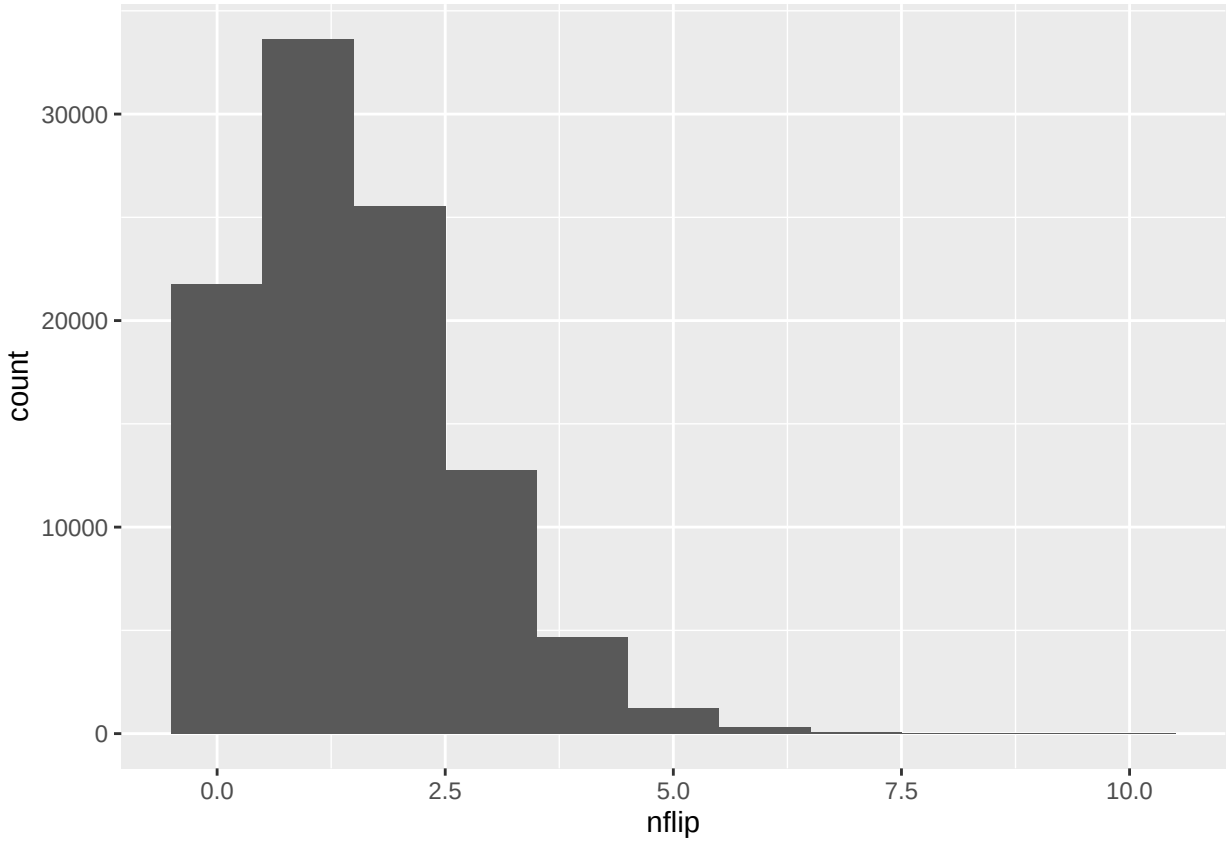
GitHub Link: https://github.com/SeraphymIgnacio/SDS315_HW4.git

Problem 1



The null hypothesis being tested is "over the long run, securities trades from the Iron Bank are flagged at the same 2.4% baseline rate as that of other traders". The test statistic used to measure evidence against the null hypothesis was the number of trades flagged by the algorithm, or 70/2021. The plot above displays the distribution of this test statistic, assuming that the null hypothesis is true. The p-value is 0.0019.
Conclusion: The p-value is less than 0.05, so we reject the null hypothesis. This means that trades are NOT flagged at the same 2.4% rate.

Problem 2

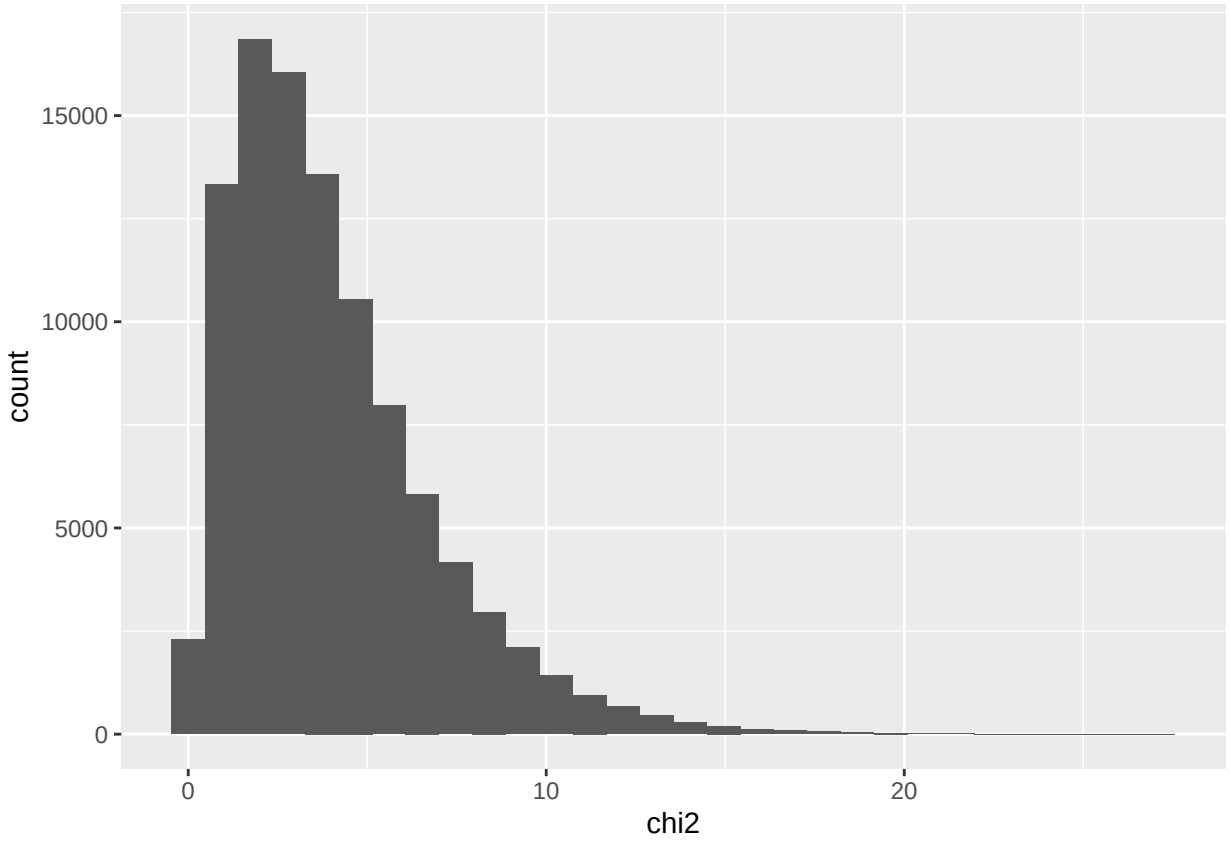


The null hypothesis being tested is "on average, restaurants in the city are cited for health code violations at the same 3% base rate". The test statistic used to measure evidence against the null hypothesis was the number of health code violations, or 8/50. The plot above displays the distribution of this test statistic, assuming that the null hypothesis is true. The p-value is 0.00013.
Conclusion: The p-value is less than 0.05, so we reject the null hypothesis. This means that health code violations are NOT at an average rate of 3%.

Problem 3

```
## [1] 240
```

```
## [1] 5.538889
```

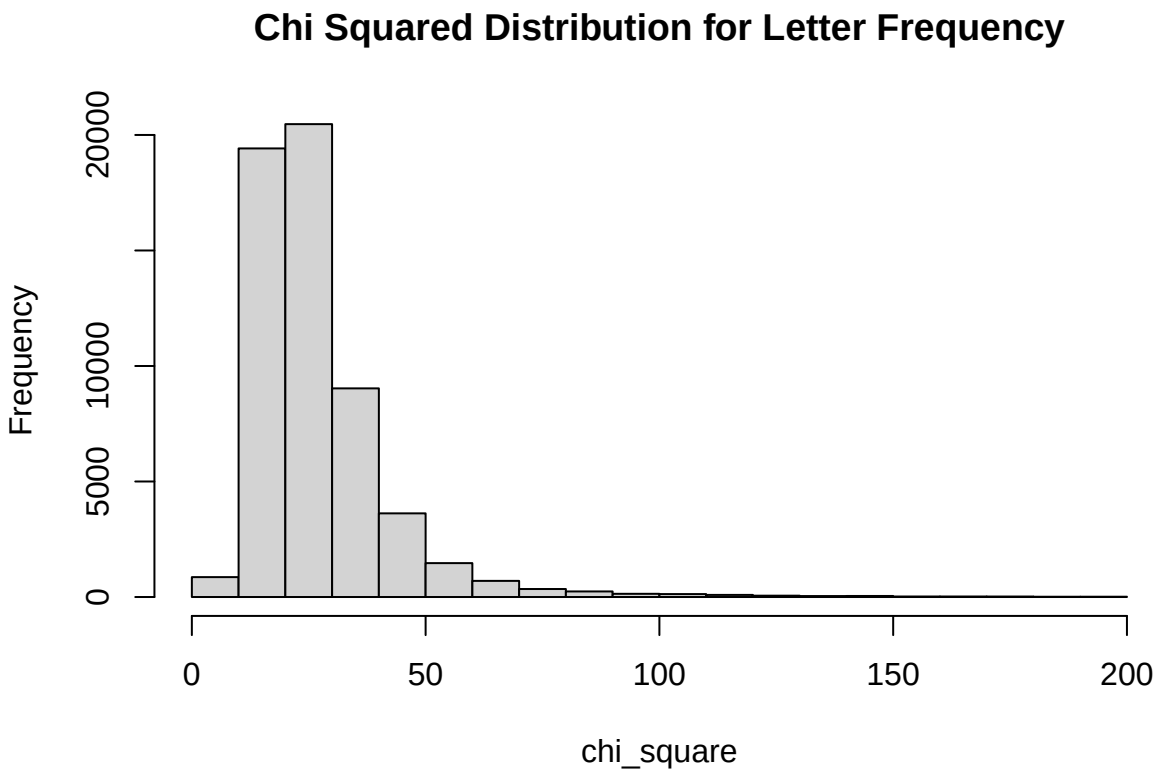


The null hypothesis being tested is "the jury selection process fairly represents the county's eligible population" which means that it is claimed that the observed racial proportions would match the expected proportions. To test this, we compared the expected and observed counts of jurors over 20 trials, created a simulation for 100,000 times each selecting 240 jurors, and calculated the chi-squared statistic for the simulation. From the chi-squared test, The distribution in the plot shows that there is little significance past the value 12.43

The p-value is 0.014.
Conclusion: The p-value is less than 0.05, which means we can reject the null hypothesis that jurors are chosen fairly. This may be due to a shifting demographic, or racial prejudice which undoubtedly still plays a factor in many similar situations. Further investigation across multiple counties may be done in order to determine a more accurate reasoning.

Problem 4

Part A



Part B

the table displays all p-values in the order in which the sentences are inputted. Based on the p-values given sentence 6 had the lowest p-values and is low enough to reject the null hypothesis, concluding that sentence 6 was written with an LLM.

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."

p-val = 0.001