

Karthik_Kasa_SDS315Homework4

Karthikeya Kasa

2025-02-18

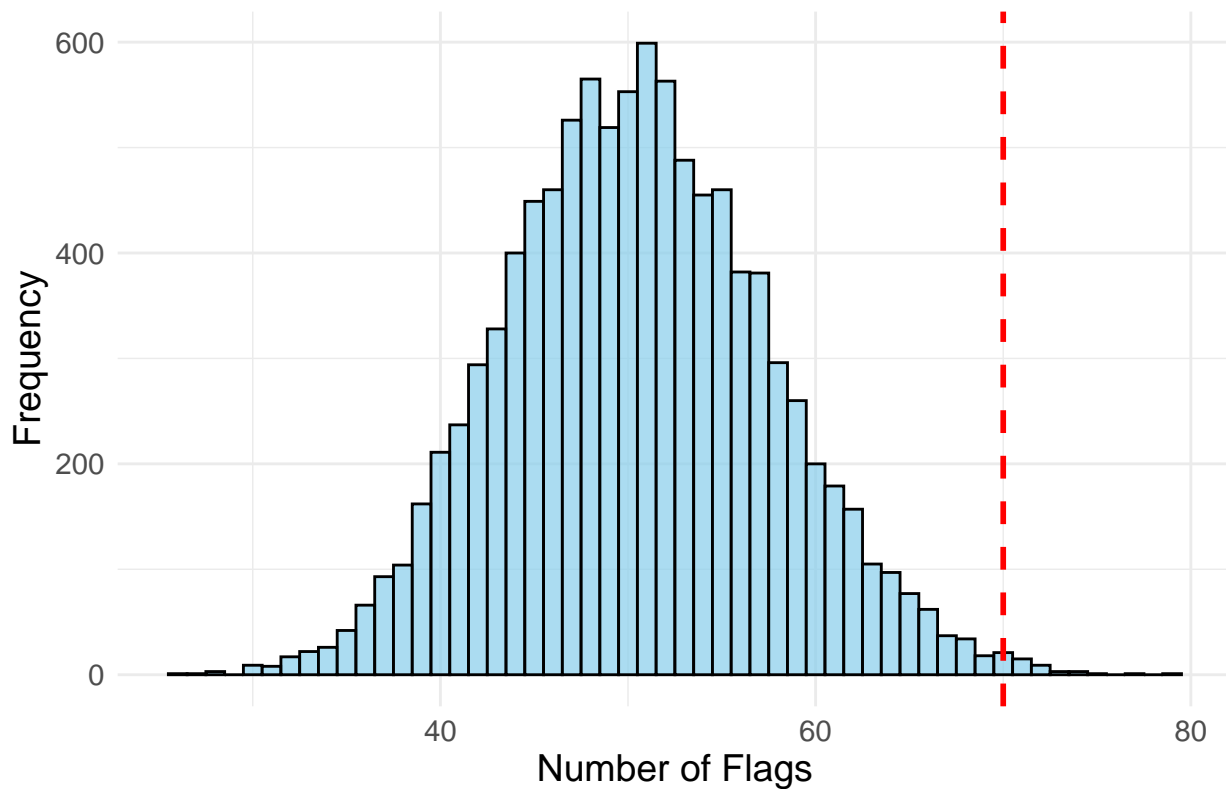
UTEID: kk38378

Github Link:

<https://github.com/karthikeyakasa/SDS2025Homewok4>

Problem 1) Iron Bank

Distribution of Simulated Flags



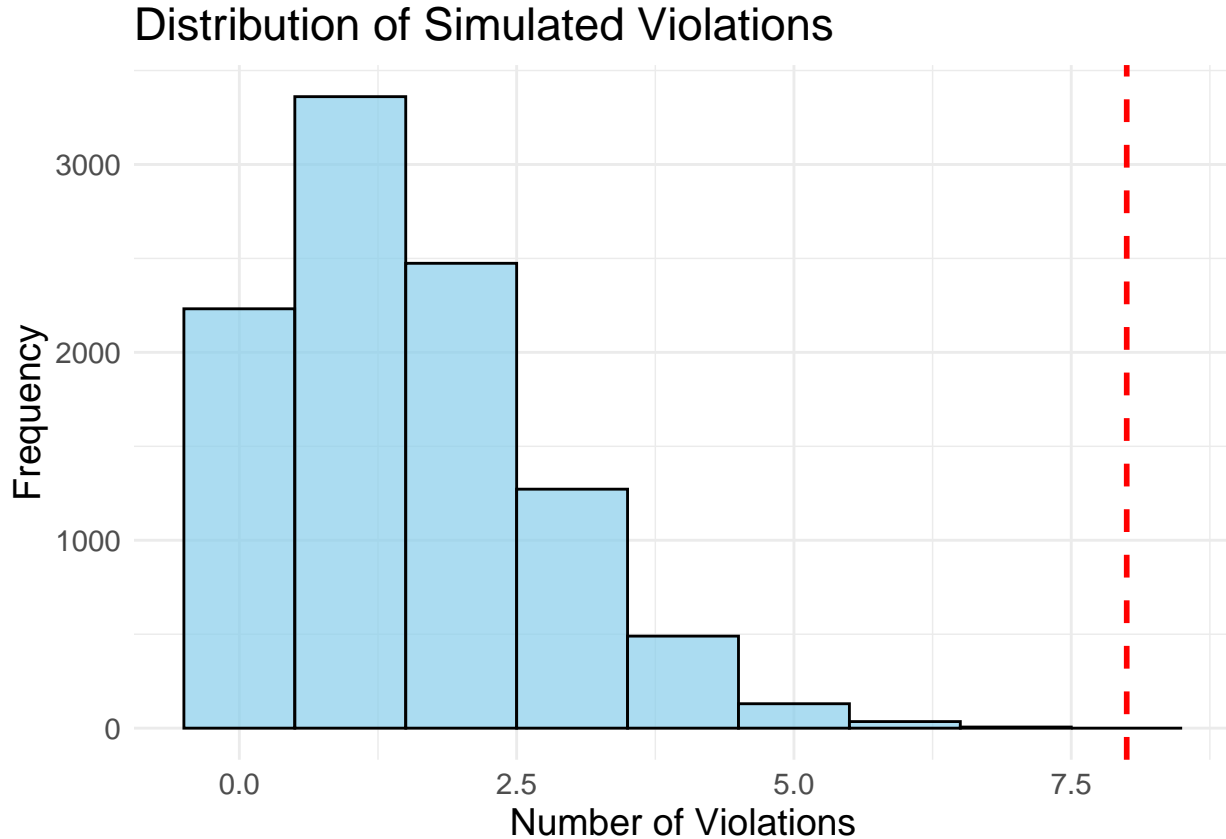
```
## [1] "P value: 0.0054"
```

The Null Hypothesis or H_0 states that the rate of which Iron Bank employees' trades are flagged is the same as the baseline rate of 2.4% observed among all traders, which means that any deviations from this rate are due to random chance rather than systematic differences in trading behavior. The test statistic is the number of flagged trades in a given simulation, and in the real data the value is 70. To test this, I used Monte Carlo simulation with 10,000 trials, where I simulated 2,100 trades per trial and recored how many times 70 trades were flagged. The p-value ended up being 0.0038, meaning that if Iron Bank employees were being flagged at a normal rate, this extreme result would only happen 0.38% of the time. Since this a very

small possibility, I reject the null hypothesis and conclude that Iron Bank's flagged trade rate is significantly higher than expected, making it reasonable for the SEC to investigate.

Problem 2) Health Inspections

```
## [1] 2
```

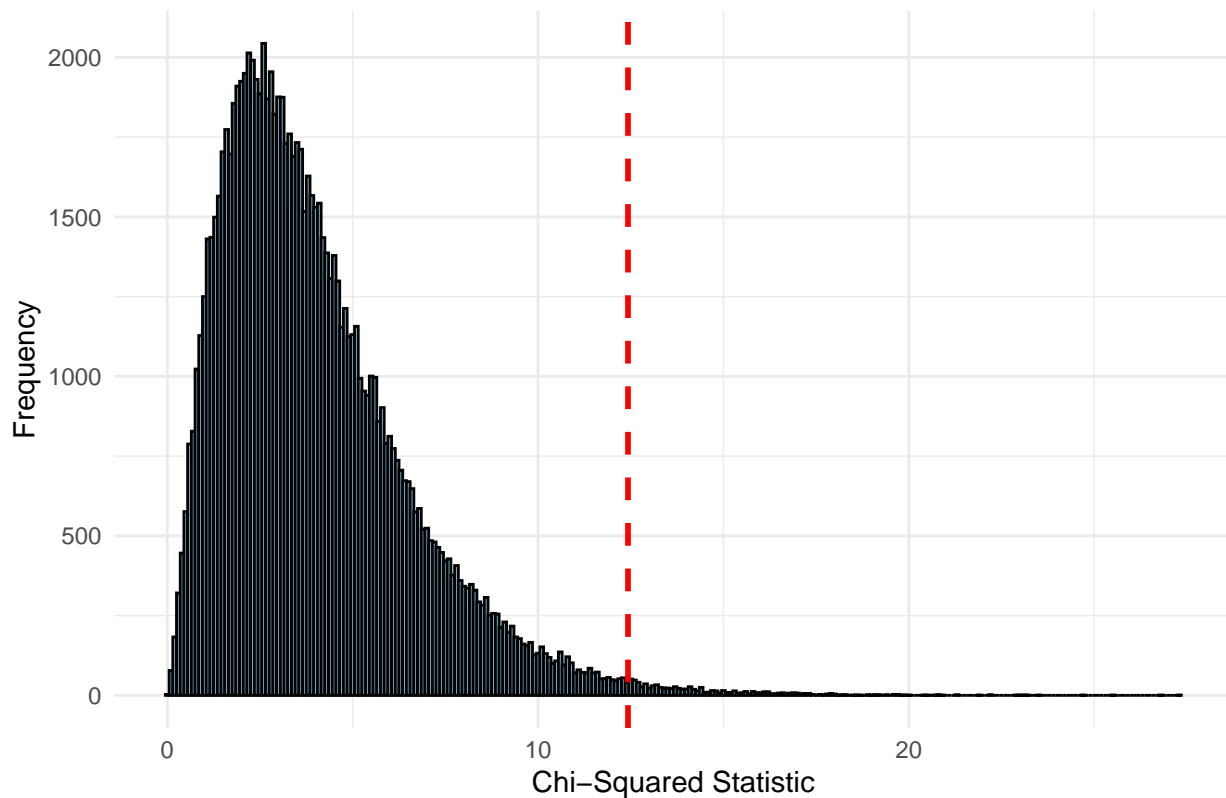


```
## [1] "P-value:0"
```

The null hypothesis is that the Gourmet Bites' restaurants receive health code violations at the same 3% rate as all other restaurants in the city, meaning that any increase in violations is just due to random chance. The test statistic is the number of violations in a given sample, and in the real data, this value is 8 out of 50 inspections. To test this, I used Monte Carlo simulation with 10000 trials, where I simulated 50 inspections per trial and counted how often at least 8 violations were observed. The p-value came out to be 0.0001 meaning that if Gourmet Bites truly had a violation of only 3%, we'd expect to see 8 or more violations in just 0.01% of cases. Since this probability is far below the common significance thresholds, I reject the null hypothesis and conclude that Gourmet Bites' violation rate is significantly higher than expected, providing strong justification for a further investigation by the Health Department

Problem 3) Evaluating Jury Selection for Bias

Distribution of Chi-Squared Statistics (Under Null Hypothesis)



```
## [1] "P-value:0.00889"
```

The null hypothesis here is that the jurors picked by this judge have the same racial/ethnic makeup as the county's eligible jury population, meaning any differences are just by chance. To test this, I used the chi-squared statistic, which compares the actual juror counts (85, 56, 59, 27, and 13) to the expected counts based on the county's population. I ran a Monte Carlo simulation with 100,000 trials, randomly picking jurors based on those proportions and calculating the chi-squared statistic each time. The p-value came out to 0.00869, which means that if jurors were picked based on the county's proportions, we'd only expect to see this big of a difference about 0.87% of the time. Since the p-value is much lower than the usual cutoff, I can reject the null hypothesis. This suggests that the jury selection process under this judge might be different from what we'd expect based on the county's demographics, which could mean there's some bias. However, other factors like eligibility or juror excusals could be involved, so we'd need more data to really understand what's going on. The histogram of chi-squared values from the simulations shows that the observed value is a big outlier, which backs up this conclusion.

Problem 4) LLM Watermarking

Distribution of Chi-Squared Statistics (Brown Corpus)

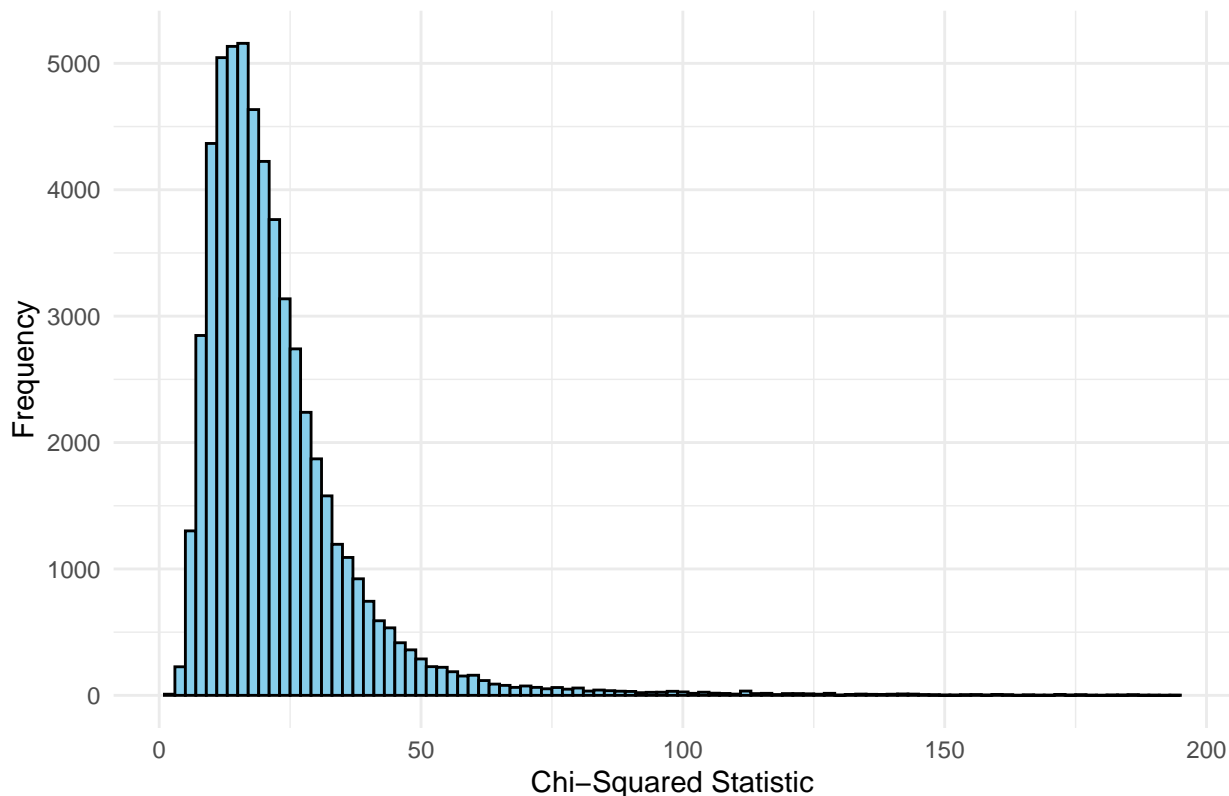


Table 1: P-Values for Each Sentence Based on Chi-Squared Test

Sentence	P-Value
She opened the book and started to read the first chapter, eagerly anticipating what might come next.	0.483
Despite the heavy rain, they decided to go for a long walk in the park, crossing the main avenue by the fountain in the center.	0.775
The museum's new exhibit features ancient artifacts from various civilizations around the world.	0.082
He carefully examined the document, looking for any clues that might help solve the mystery.	0.445
The students gathered in the auditorium to listen to the guest speaker's inspiring lecture.	0.612
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.	0.008
The chef demonstrated how to prepare a delicious meal using only locally sourced ingredients, focusing mainly on some excellent dinner recipes from Spain.	0.342
They watched the sunset from the hilltop, marveling at the beautiful array of colors in the sky.	0.913
The committee reviewed the proposal and provided many points of useful feedback to improve the project's effectiveness.	0.071
Despite the challenges faced during the project, the team worked tirelessly to ensure its successful completion, resulting in a product that exceeded everyone's expectations.	0.051

The sentence most likely to have been generated by AI is sentence number 6. The sentence is:

“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.”

It has the smallest p-value, which basically means its letter pattern is way different from what you'd usually see in normal English text. This could mean the letters were tweaked a little, like how AI-generated text sometimes has small changes in letter frequencies. The low p-value just shows that this sentence doesn't really match the usual way English sentences look.