

SDS315_HW5

Neil Sebastian
2024-02-26

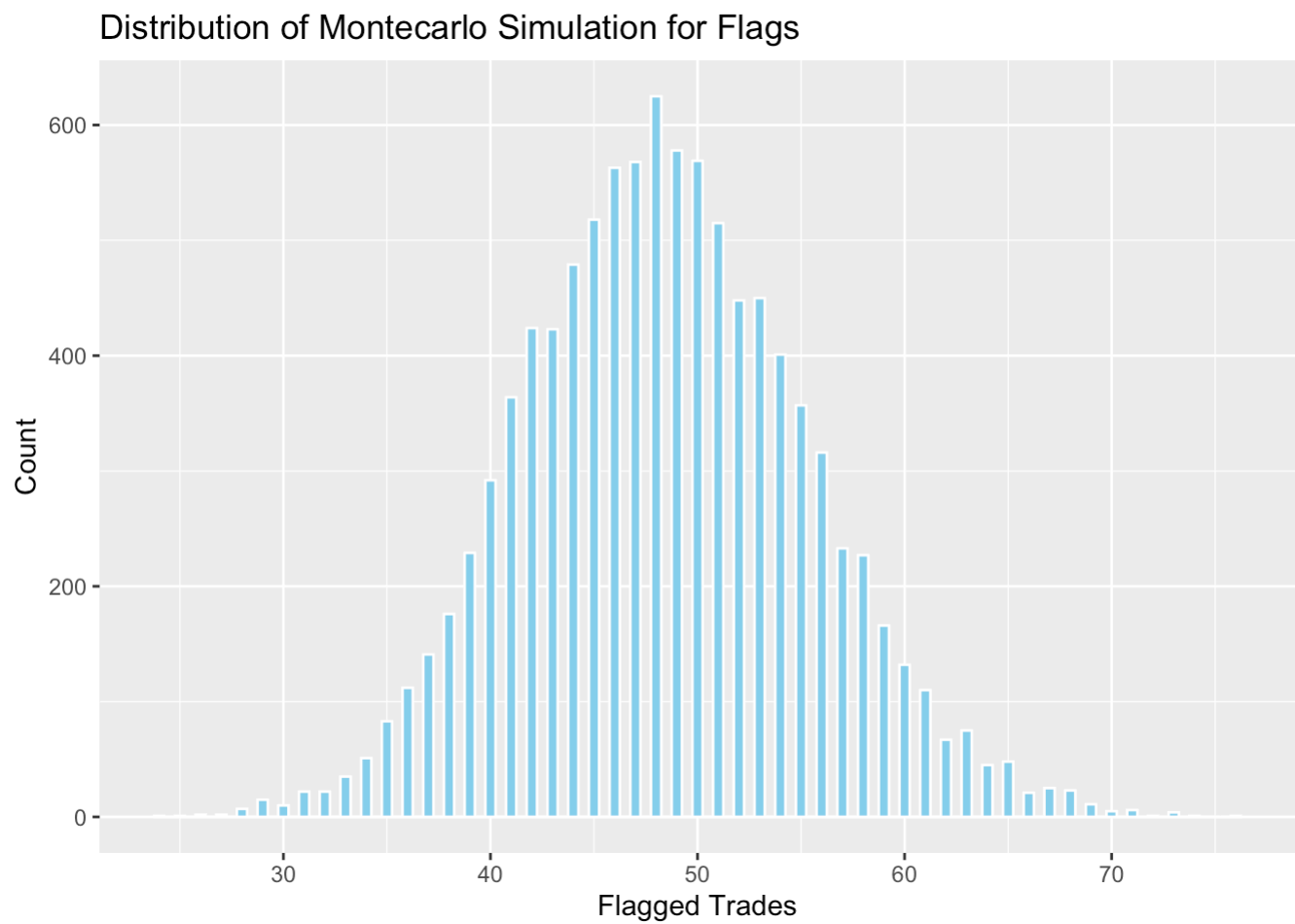
HW5

UT EID: ns36965

[GitHub Link](#)

Problem One - Iron Bank

Use Monte Carlo simulation (with at least 100000 simulations) to calculate a p-value under this null hypothesis.



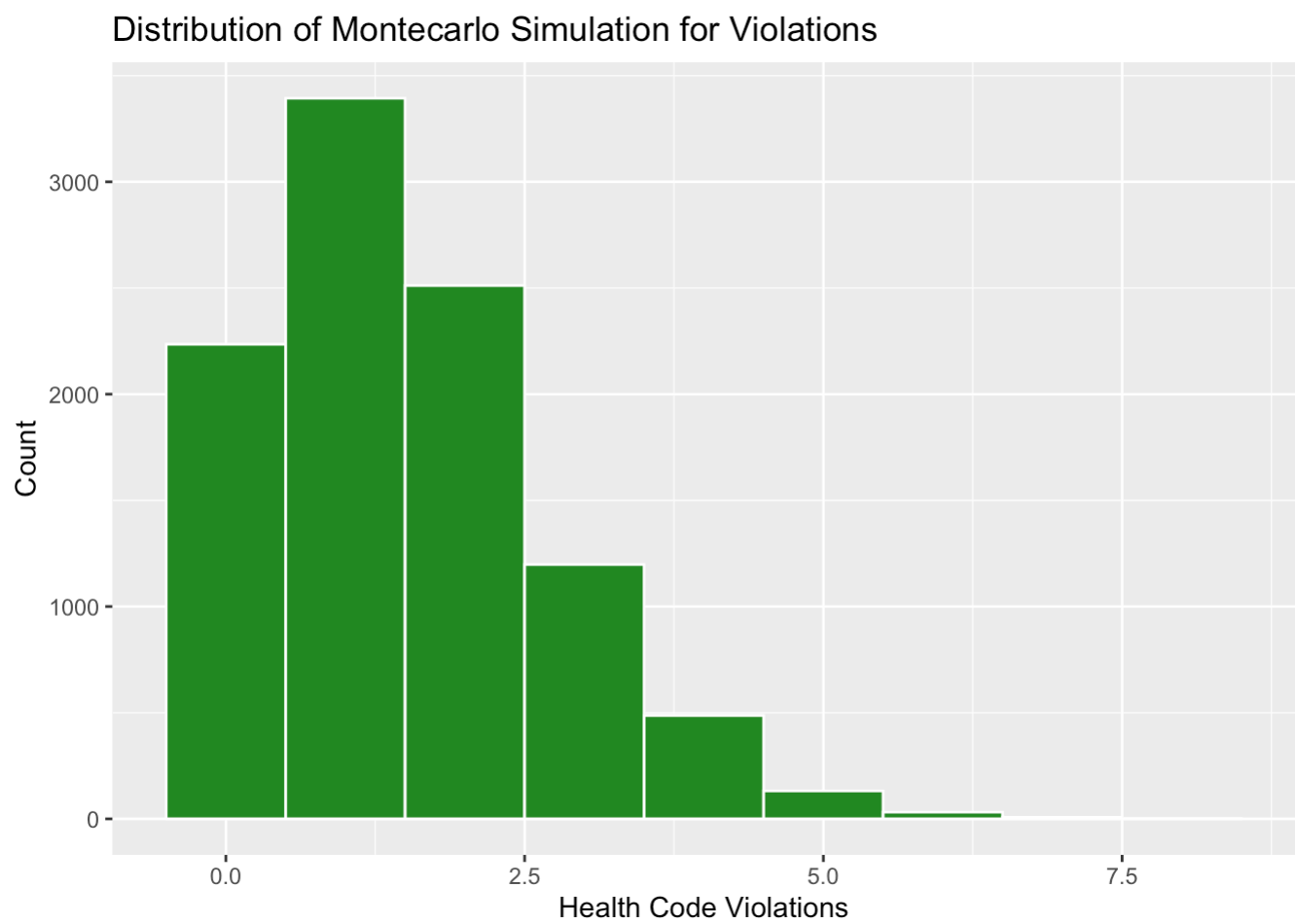
```
## [1] 0.0018
```

Writeup:

Our null hypothesis is that trades from the Iron Bank are flagged at the same 2.4% baseline rate as that of other traders. This means that any legal trade has a 2.4% chance of bein flagged. Our test statistic is the amount of flagged cases, in this case 70 flagged cases out of 2021. The plot shown above is the pobability distribution of the test statistic, assuming that the null hypothesis is true. Our p-value is .0014. This p-value leads me to draw the conclusion that the null hypothesis looks rather implausible due to the p value being around .002 which is a statistically significant value against the null hypothesis. This can be interpreted as at least some of the trades being flagged are illegal trades because the p-value is against the null hypothesis.

Problem Two - Health Inspections

Are the observed data for Gourmet Bites consistent with the Health Department's null hypothesis that, on average, restaurants in the city are cited for health code violations at the same 3% baseline rate?



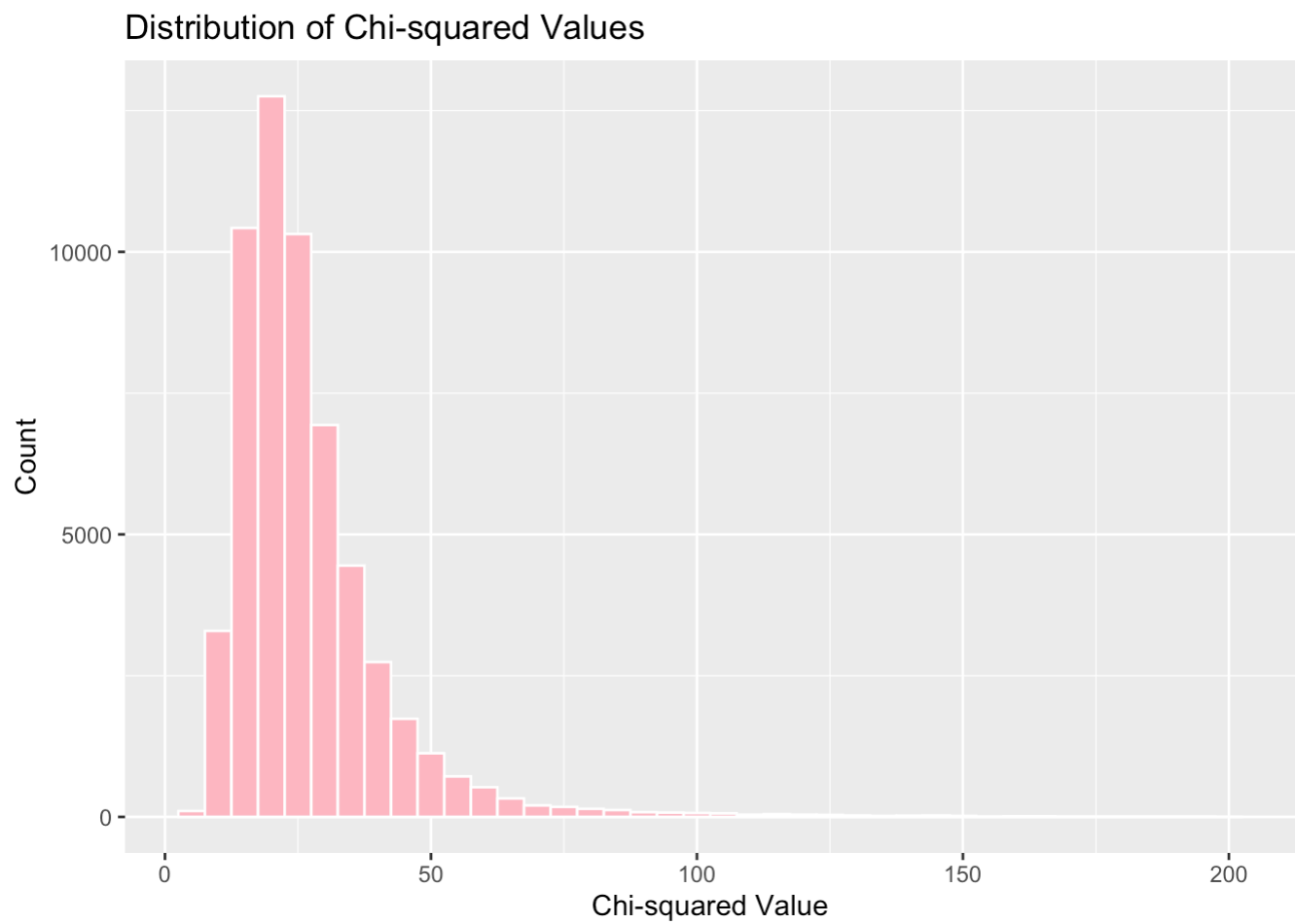
```
## [1] 2e-04
```

Writeup:

Our null hypothesis is that the observed data for Gourmet Bites, 8 health code violations in 50 evaluations, is consistent that on average, restaurants in the city are cited for health code violations at the same 3% baseline rates. This means that any restaurant has a 3% chance of being flagged withouth having any health code violations. Our test statistic is the amount of instances where health code violations were reported, in this case 8 instances of violations occured in 50 evaluations. The plot shown above is the pobability distribution of the test statistic, assuming that the null hypothesis is true. Our p-value is very very small at around 2*10^-4. This p-value leads me to draw the conclusion that the null hypothesis looks rather implausible due to the p value being around .0001 which is a statistically significant value against the null hypothesis. This can be interpreted as the chain Gourmet Bites is being flagged for actual violations rather than the random luck (or in this case unlucky) 3% baseline where violations are reported because the null hypothesis is not being supported by the p-value.

Problem Three - LLM Watermarking

Part A - The null or reference distribution



The graph above represents the distribution of chi-squared values for actual vs expected letter count from the brown sentences file.

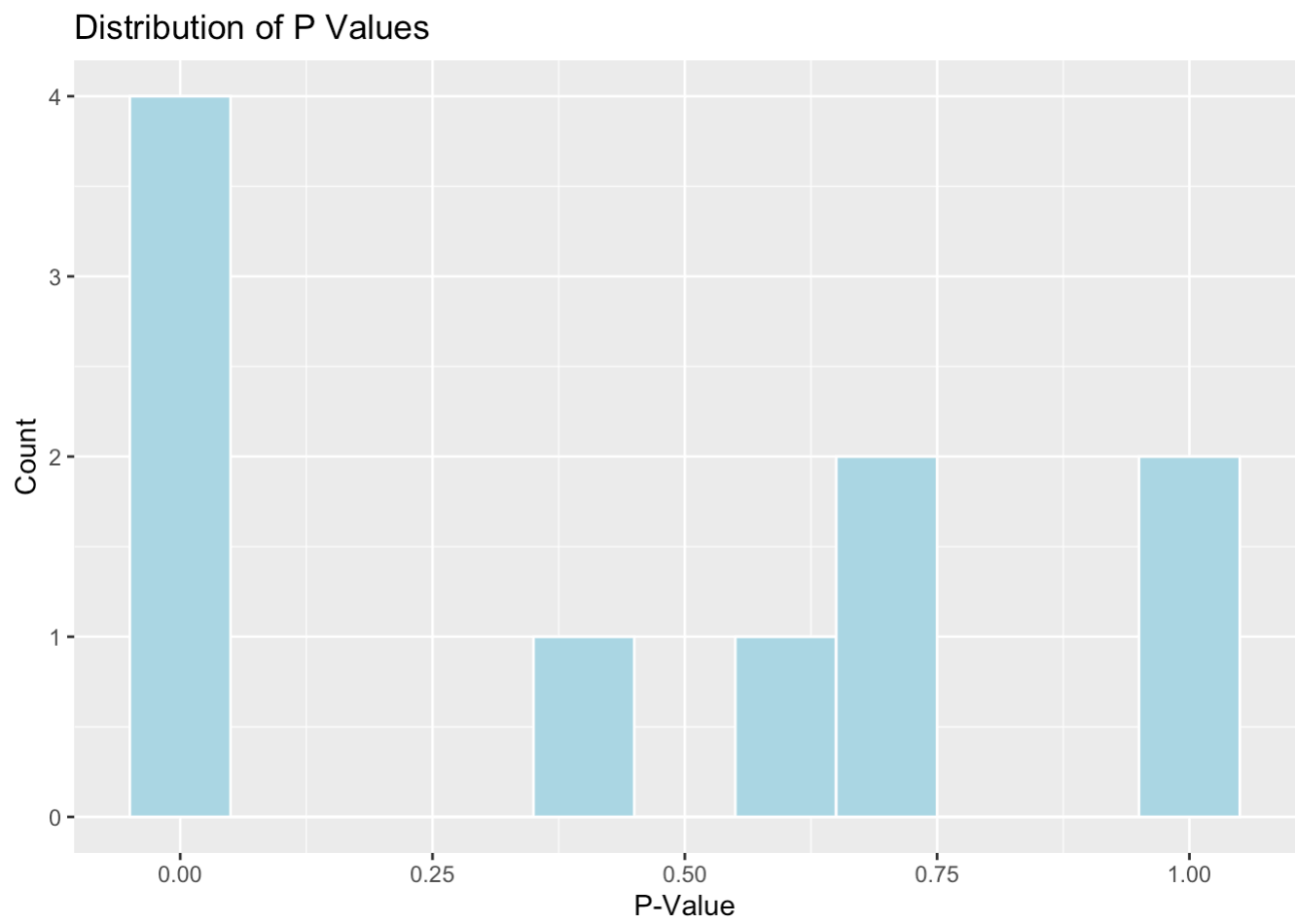
Part B - Checking for A Watermark

One of these sentences has been produced by an LLM, but watermarked by asking the LLM to subtly adjust its frequency distribution over letters. Which sentence is it? How do you know

Ten Sentences Table

Sentence
She opened the book and started to read the first chapter, eagerly anticipating what might come next.
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.
The museum's new exhibit features ancient artifacts from various civilizations around the world.
He carefully examined the document, looking for any clues that might help solve the mystery.
The students gathered in the auditorium to listen to the guest speaker's inspiring lecture.
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 chef demonstrated how to prepare a delicious meal using only locally sourced ingredients, focusing mainly on some excellent dinner recipes from Spain.
They watched the sunset from the hilltop, marveling at the beautiful array of colors in the sky.
The committee reviewed the proposal and provided many points of useful feedback to improve the project's effectiveness.
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.

```
## [1] 6.887220e-01 9.889180e-01 1.185057e-02 6.553473e-01 6.482132e-01
## [6] 9.768617e-10 3.970831e-01 9.991454e-01 1.654464e-02 4.598764e-03
```



The sentence that is likely produced by an LLM 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." It has the lowest p value meaning that it is unsupportive of the null hypothesis. The null hypothesis in this case is the sentence is NOT created by a LLM. Since that p-value for that sentence is so low (less than .001) our null hypothesis is not supported, therefore we conclude that this sentence is likely written by a LLM.