

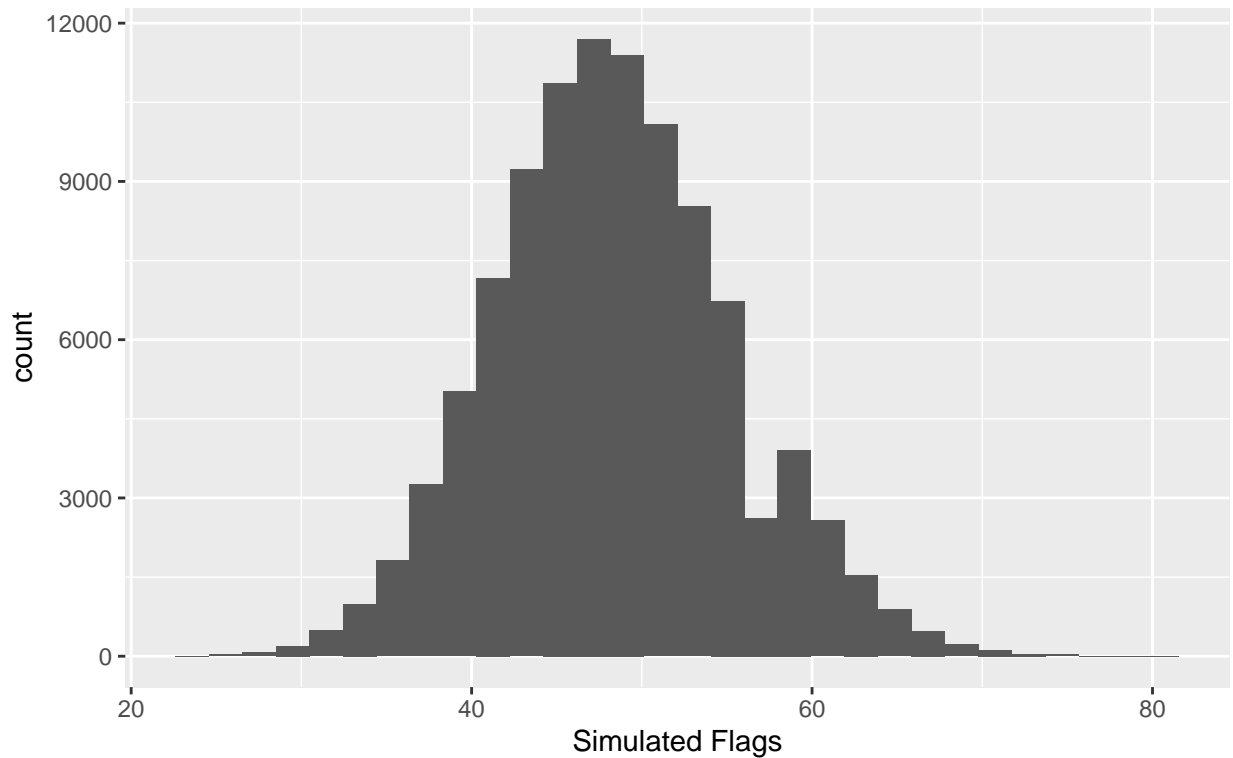
HW4: https://github.com/Panda-nny/SDS315_HW4

Danny Pan (dp36627)

2025-02-17

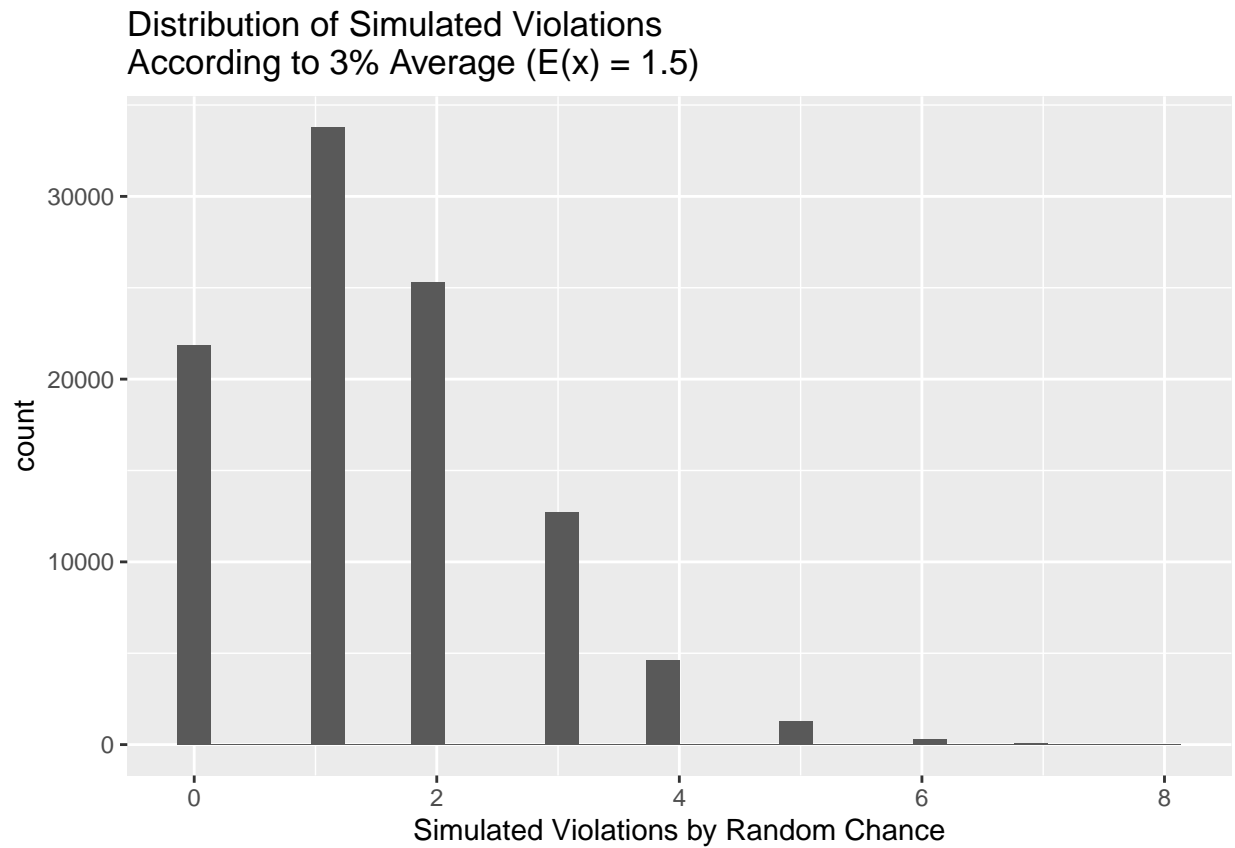
Problem 1 - Iron Bank

Distribution of Simulated Flags of Iron Bank
According to 2.4% Baseline ($E(x) = 48.5$)



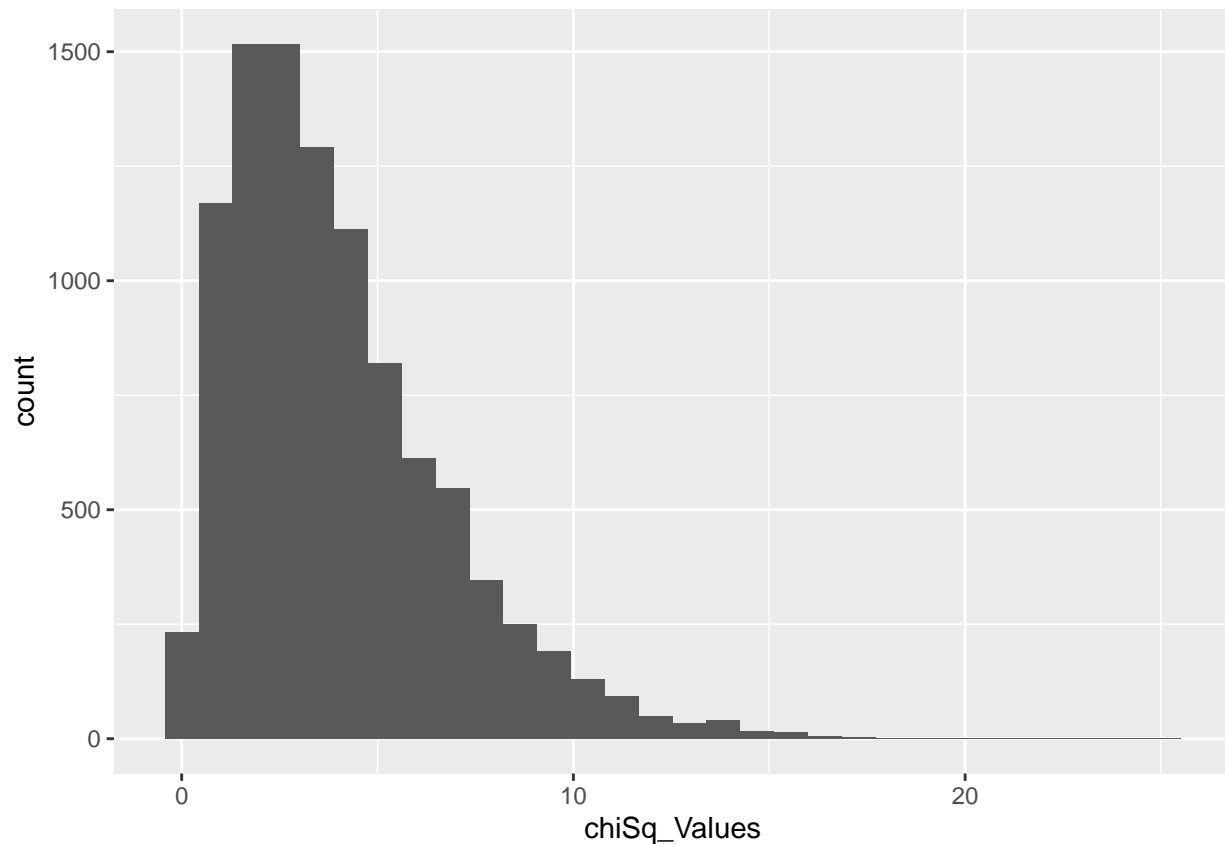
To test the null hypothesis that the SEC detection algorithm flags securities trades from the Iron Bank at the same 2.4% baseline, 100 thousand simulations of a similar number of 2021 trades were run. According to the baseline, around 48.5 flags are expected to occur on average. That being said, only 202 (0.202%) of these simulations resulted in 70 or more flagged trades, and only 58 simulations (0.058%) resulted in 27 or less flags, which details the other end since a two-tailed test was performed. This results in a p-value of .0026, which is pretty convincing evidence (beyond reasonable doubt) to suggest that the null hypothesis is implausible.

Problem 2 - Health Inspections



To test the null hypothesis that the observed data for Gourmet Bites is consistent with the Health Department's 3% baseline, 100 thousand simulations of a similar number of 50 inspections were run. According to the baseline, around 1.5 are expected to occur on average. That being said, only 11 (0.011%) of these simulations resulted in 8 or more healthcode violations. This results in a p-value of .00011, which is pretty convincing evidence (beyond reasonable doubt) to suggest that the observed data does not align with the 3% baseline of health code violations due to random chance.

Problem 3 - Evaluating Jury Selection for Bias



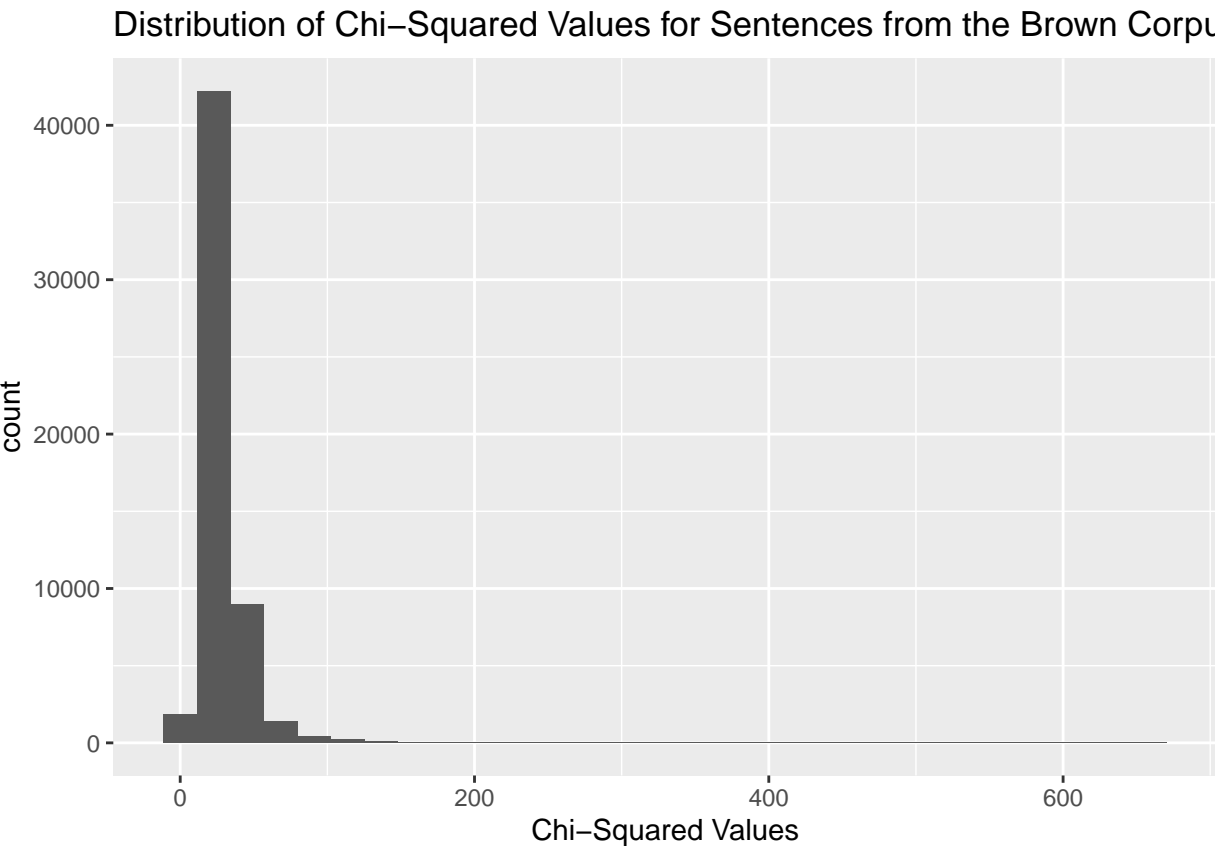
H₀: There is no significant difference between the distribution of jurors respective to their racial/ethnic group and the jurors selected by a certain judge.

T: The chi-squared statistic comparing the observed counts of jurors in each group to the expected amount according to the demographic breakdown of the county's eligible jury pool results in a value of $\chi^2 = 12.426$

P(T | H₀): According to 10,000 simulations of normal variance, only 131 simulated times did the simulated χ^2 value equate to or exceed 12.426. This results in a p value of 0.0131.

Conclusion: A p-value of 0.0131 suggests that there is likely some sort of bias (intentional or not) influencing the judge's jury selection. This may be explained by the fact that potentially, certain groups are more likely to be removed "for cause" or due to non-racial peremptory challenges. Further investigation is needed to conclude if this is a possibility though. Additionally, if the state's juror pool does not reflect a similar demographic breakdown as the country's population, random selection of potential jurors are unlikely to reflect the same demographic proportions. To evaluate this, the state's public demographic statistics can be compared to the country's.

Problem 4 - LLM watermarking



The above histogram displays the distribution of Chi Squared values of the difference in expected letter frequencies for each sentence in the Brown Corpus. This was used as a baseline to evaluate the extremity of character frequencies in 10 sentences, one being watermarked and having irregular letter frequencies. The table below lists each sentence, their total Chi-Squared value, and the p-value of the Chi-Squared value relating to the distribution of irregularity of letters compared to the Brown Corpus. Compared to the null hypothesis that the sentences have a typical letter distribution, since the 6th sentence contains a unique low p-value of 0.011, we determine that the 6th sentence is likely the watermarked entry.

Sentence	chiSquared	pVal
She opened the book and started to read the first chapter, eagerly anticipating what might come next.	22.931	0.522
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.	13.051	0.928
The museum's new exhibit features ancient artifacts from various civilizations around the world.	46.286	0.080
He carefully examined the document, looking for any clues that might help solve the mystery.	23.546	0.499
The students gathered in the auditorium to listen to the guest speaker's inspiring lecture.	23.676	0.494
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.	96.453	0.011
The chef demonstrated how to prepare a delicious meal using only locally sourced ingredients, focusing mainly on some excellent dinner recipes from Spain.	28.271	0.334
They watched the sunset from the hilltop, marveling at the beautiful array of colors in the sky.	9.635	0.988
The committee reviewed the proposal and provided many points of useful feedback to improve the project's effectiveness.	44.929	0.088
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.	49.961	0.063