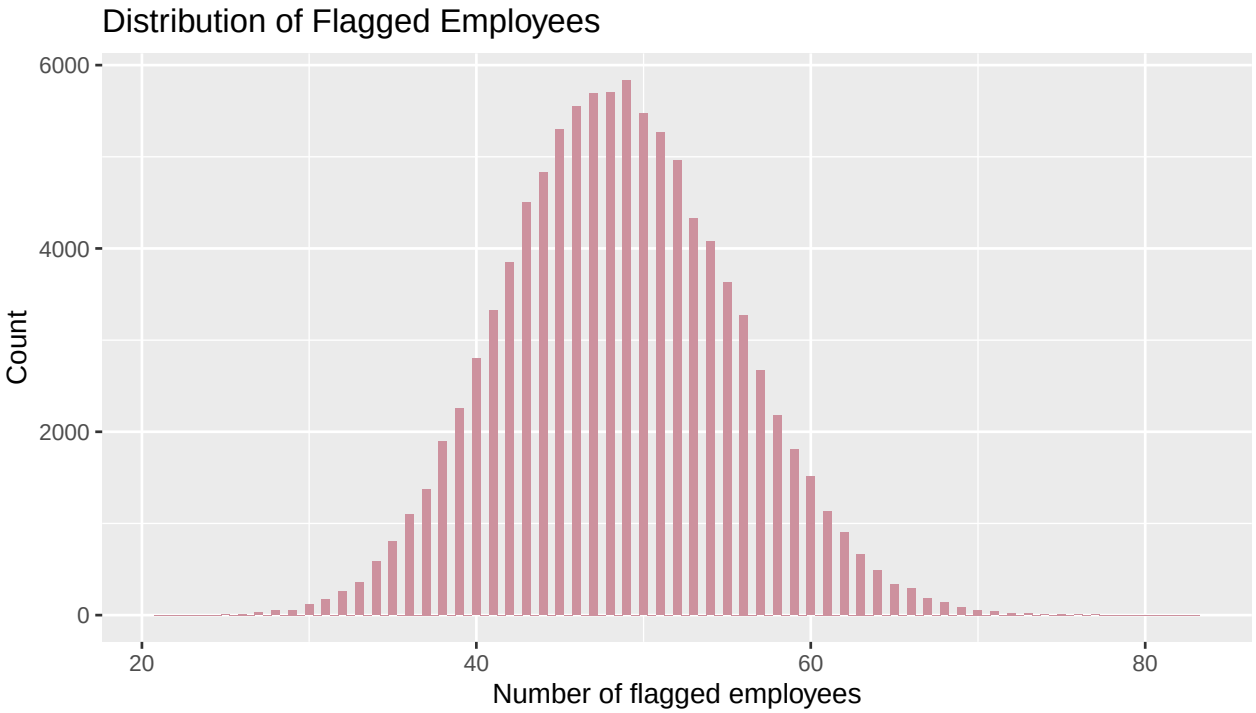
Homework 5

Juhi Malwade

2/27/24

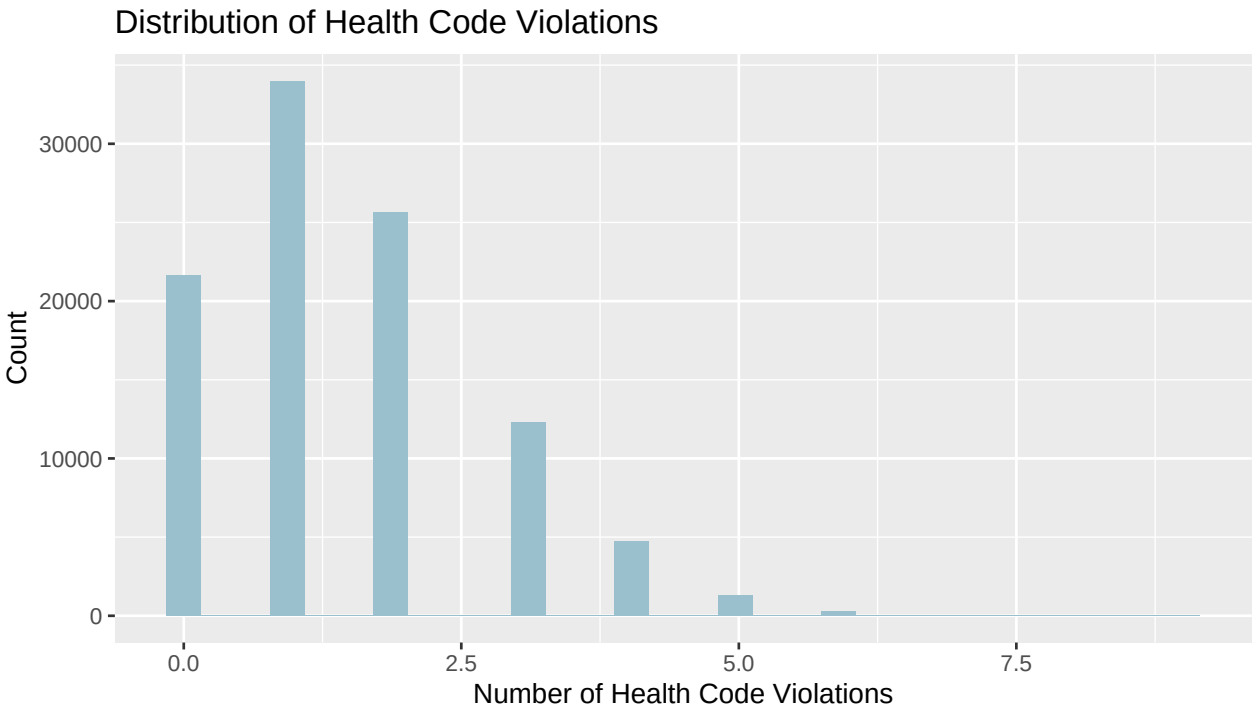
Github Link

Problem 1: Iron Bank



The null hypothesis I am testing is that the baseline probability that any legal trade flagged by SEC's algorithm is 2.4%. The test statistic being used to measure evidence against the null hypothesis is the number of cases that are flagged out of 2021 cases. The probability distribution of the number of cases flagged out of 2021 total cases (assuming the null hypothesis is true) is plotted above. The p-value, or the probability of observing 70 or more flagged cases assuming the null hypothesis is true, is 0.000171. The p-value is extremely low and represents a 0.02% chance of observing 70 or more flagged cases due to blind luck. Thus, I reject the null hypothesis because it does not seem plausible that this observation occurred simply due to chance. There is significant evidence that the employees violated federal insider trading laws.

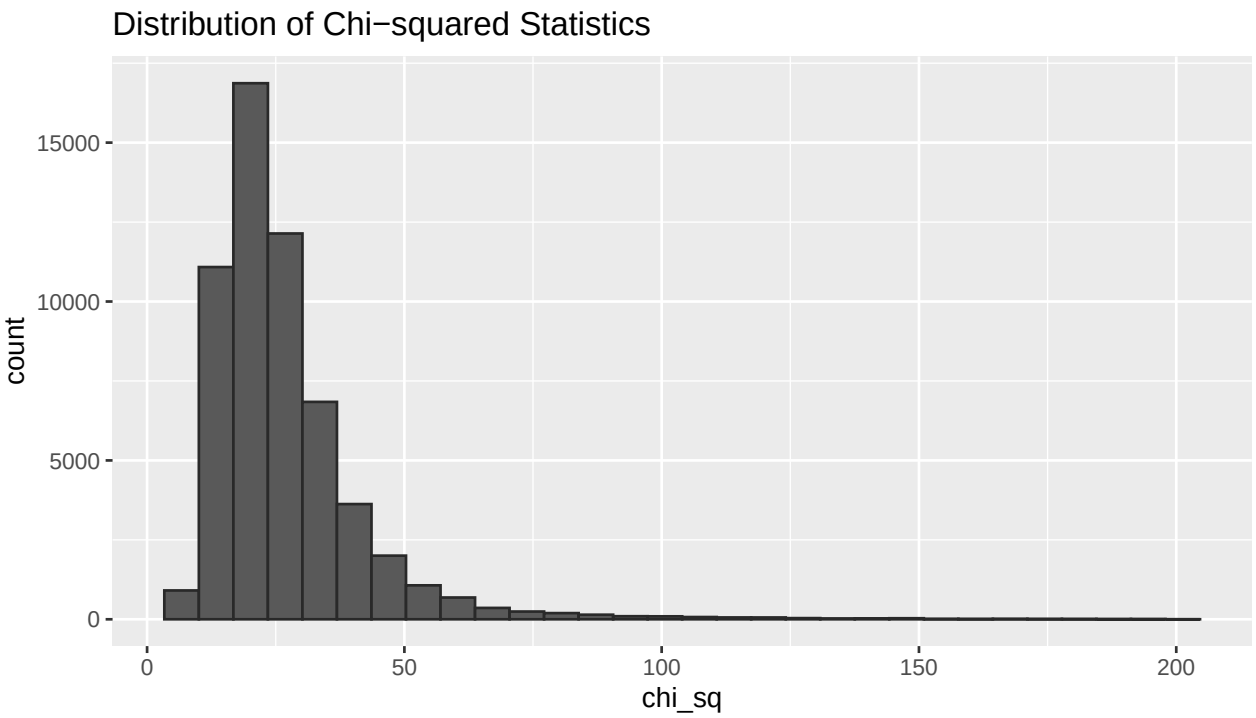
Problem 2: Health Inspections



The null hypothesis I am testing is that, on average, restaurants in the city are cited for health code violations at a 3% baseline rate. The test statistic being used to measure evidence against the null hypothesis is the number of health code violations out of 50 inspections. The probablility distribution of the number of health code violations out of 50 inspections (assuming the null hypothesis is true) is plotted above. The p-value, or the probablility of observing 8 or more violations assuming the null hypothesis is true, is 0.000012. The p-value is extremely low and represents a 0.0012% chance of observing 8 or more violations due to blind luck. Thus, Gourmet Bites' violations likely did not occur simply due to random issues - there is significant evidence their health code violations are indeed much higher than the citywide average of 3% and action should be taken against them.

Problem 3: LLM watermarking

Part A: The null or reference distribution



The graph above shows the null distribution - the range of chi-squared values you might expect to see in normal English sentences based on the predefined letter frequency distribution.

Part B: Checking for a watermark

sentence	p-value
She opened the book and started to read the first chapter, eagerly anticipating what might come next.	0.5128381
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.	0.9262843
The museum's new exhibit features ancient artifacts from various civilizations around the world.	0.0763768
He carefully examined the document, looking for any clues that might help solve the mystery.	0.4890651
The students gathered in the auditorium to listen to the guest speaker's inspiring lecture.	0.4840603
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.	0.0087761
The chef demonstrated how to prepare a delicious meal using only locally sourced ingredients, focusing mainly on some excellent dinner recipes from Spain.	0.3279232
They watched the sunset from the hilltop, marveling at the beautiful array of colors in the sky.	0.9880166
The committee reviewed the proposal and provided many points of useful feedback to improve the project's effectiveness.	0.0840603
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.	0.0590360

Based on the p-values of the sentences above, the sentence “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” was produced by an LLM. This is because its p-value is extremely low at 0.00878, indicating that under the null hypothesis that the sentence follows the “typical” English letter distribution, it is extremely rare to observe a sentence with this particular letter distribution. Thus, there is significant evidence that this sentence was not written by a human but rather an LLM.