SDS 315 Homework 4

Elizabeth 'Betsy' Sherhart UT EID: eas5778
Click here for link to GitHub repository

February 20, 2025

Problem 1 - Iron Bank

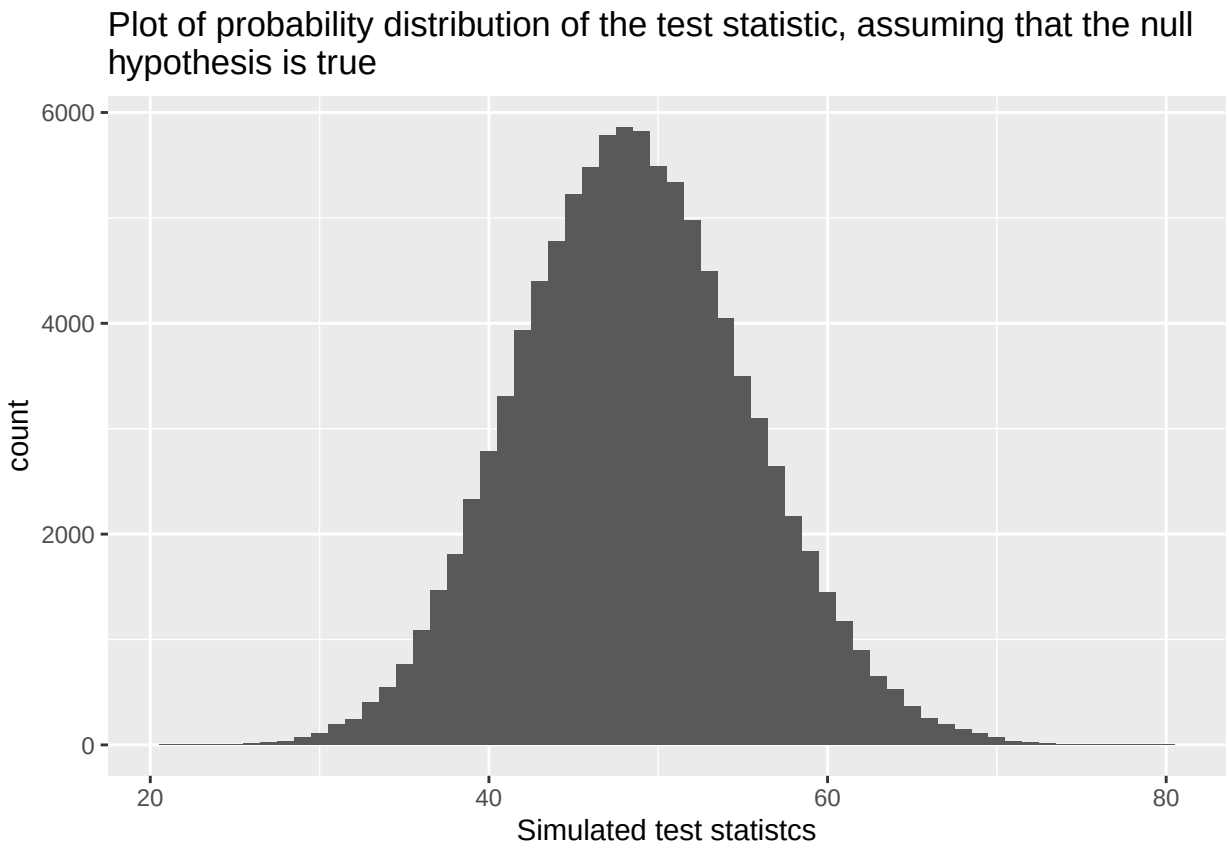
Null hypothesis:

The SEC estimates that the baseline probability that any legal trade will be flagged by their algorithm is 2.4%. The null hypothesis being tested is that over the long run, securities trades from the Iron Bank are flagged at the same 2.4% baseline rate as that of other traders.

Test statistic:

The test statistic that is used to measure the evidence against the null hypothesis was 70 trades flagged out of 2021 trades.

Plot of probability distribution:



p-value:

The p-value for the test statistic of 70 flagged trades in the probability distribution of simulated test statistics is 0.00176.

Conclusion:

Since the p-value is between 10^-6 and 10^-3, the probability that the test statistic of 70 trades out of 2021 trades will happen under the null hypothesis is very small, which means that the null hypothesis being true is highly unlikely, and the SEC should further investigate the Iron Bank.

Problem 2 - Health Inspections

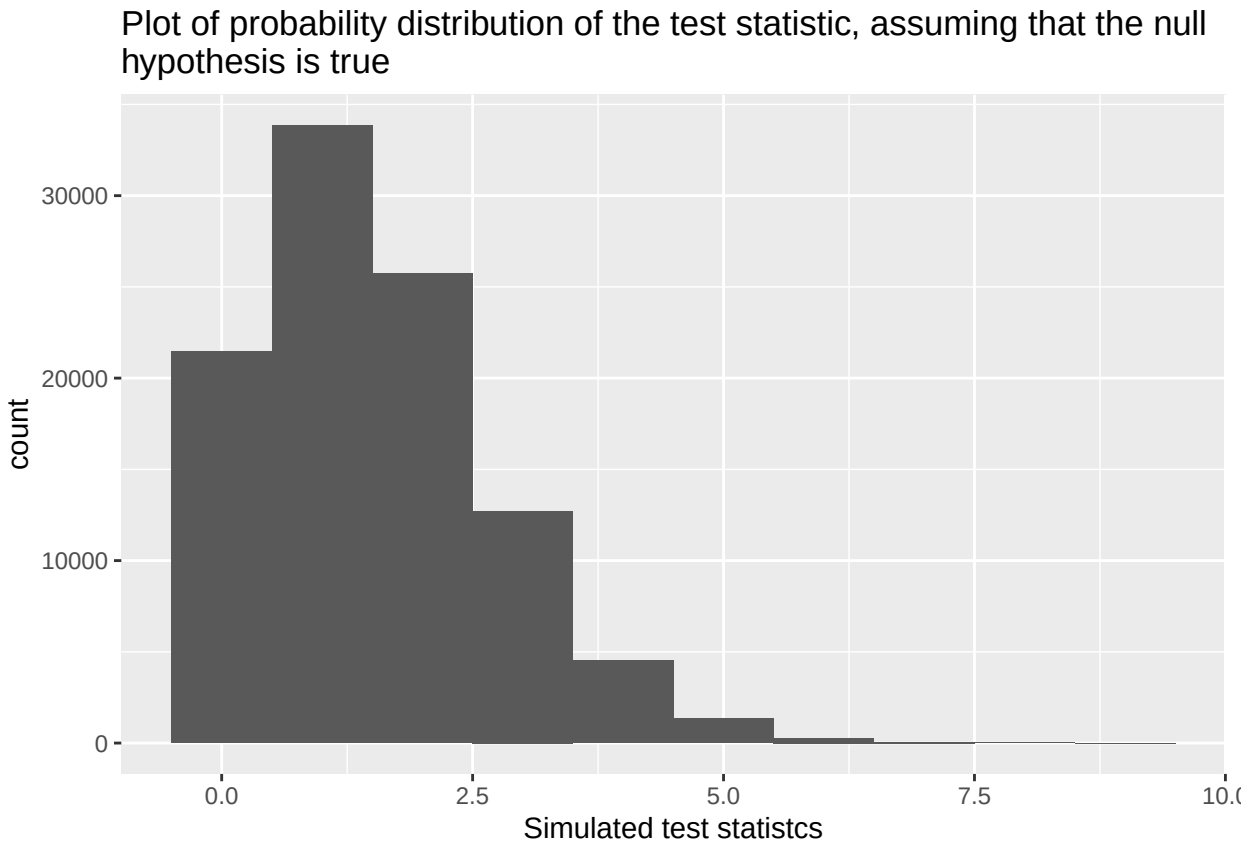
Null hypothesis:

Typically, the Health Department's data shows that, on average, 3% of all restaurant inspections result in health code violations due to random issues that can occur even in well-managed establishments. The null hypothesis being tested is that on average, Gourmet Bites will be cited for health code violations at the same 3% baseline rate.

Test statistic:

The test statistic that is used to measure the evidence against the null hypothesis was of the 50 inspections done at Gourmet Bites, 8 resulted in health code violations being reported.

Plot of probability distribution:



p-value:

The p-value for the test statistic of 8 visits where health code violations were reported out of 50 inspections in the probability distribution of simulated test statistics is 1.4×10^{-4}.

Conclusion:

Since the p-value is between 10^-6 and 10^-3, the probability that the test statistic of 8 visits where health code violations were reported out of 50 inspections will happen under the null hypothesis is very small, which means that the null hypothesis being true is highly unlikely, and the Health Department should further investigate Gourmet Bites.

Problem 3 - Evaluating Jury Selection for Bias

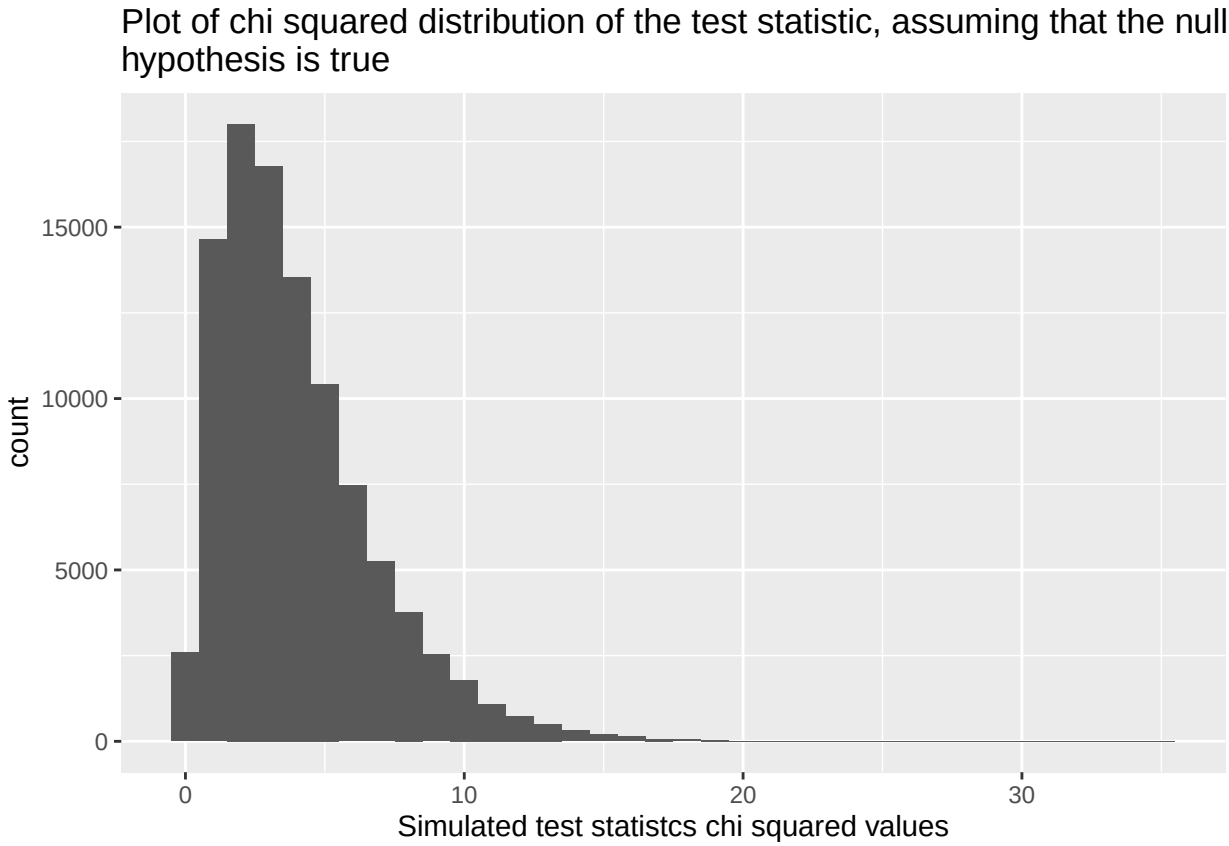
Null hypothesis:

The juries for state court cases are selected through a multi-step process designed to ensure that they represent a fair cross-section of the community. The null hypothesis is that on average, the 20 juries under the judge in question will reflect the counties population meaning their group proportions, with some variation due to variability.

Test statistic:

The test statistic that is used to measure the evidence against the null hypothesis was of the 240 people on the 20 juries under the judge in question, 85 were from group 1, 56 were from group 2, 59 were from group 3, 27 were from group 4, and 13 were from group 5.

Plot of probability distribution:



p-value:

The chi squared for the test statistic of 240 people on the 20 juries under the judge in question, 85 were from group 1, 56 were from group 2, 59 were from group 3, 27 were from group 4, and 13 were from group 5 is 12.4263889. This means that the p-value is 0.01416 where that is the probability that the simulated chi squared value is greater than or equal to the chi squared of the test statistic, 12.4263889.

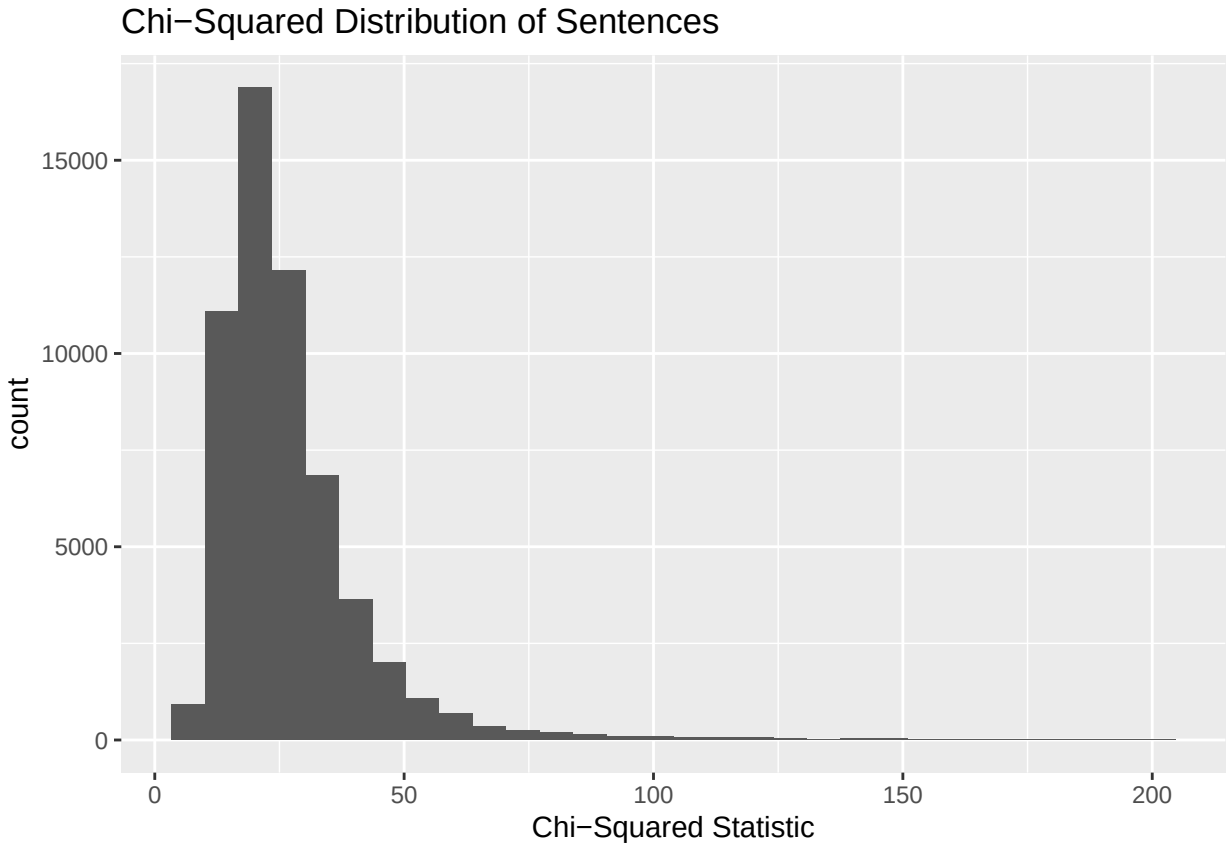
Conclusion:

Since the p-value is less than 0.05 and greater than 0.01, the probability that the test statistic will happen under the null hypothesis is small but not unlikely, which means that the null hypothesis could be true or false, and should be investigated further with more data. The p-value does suggest that the jury selection of this judge is less likely to occur under the null hypothesis, and could point to systematic bias in jury selection. However, possible explanations exist other than systematic bias, such as economic disparities that prevent a group from taking off work for long trials, exclusion of non-citizens or non-English speakers, voter registration differing between groups, and much more. The bias could be further investigated by looking into other judges in the same area to see if a county bias, or looking into economic and social aspects of the groups that could affect likeliness to serve on a jury to name a few.

Problem 4 - LLM watermarking

Part A: the null or reference distribution

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



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.	0.582
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.	0.976
The museum's new exhibit features ancient artifacts from various civilizations around the world.	0.006
He carefully examined the document, looking for any clues that might help solve the mystery.	0.546
The students gathered in the auditorium to listen to the guest speaker's inspiring lecture.	0.538
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.	0.000
The chef demonstrated how to prepare a delicious meal using only locally sourced ingredients, focusing mainly on some excellent dinner recipes from Spain.	0.296
They watched the sunset from the hilltop, marveling at the beautiful array of colors in the sky.	0.998
The committee reviewed the proposal and provided many points of useful feedback to improve the project's effectiveness.	0.009
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.	0.002

Conclusion:

Sentence 6 appears to have the most significant deviation from the expected letter distribution because it has the lowest p-value, which is 0.000. The p-value indicates that it's been generated or altered in a way that

doesn't reflect typical English frequency distribution of letters, meaing it could be the watermarked LLM sentence.