HW4 - SDS315

Ahantya Sharma

UT EID: as236366

Github Link: https://github.com/Ahantya/SDS315/blob/main/HW4/HW4Markdown.Rmd
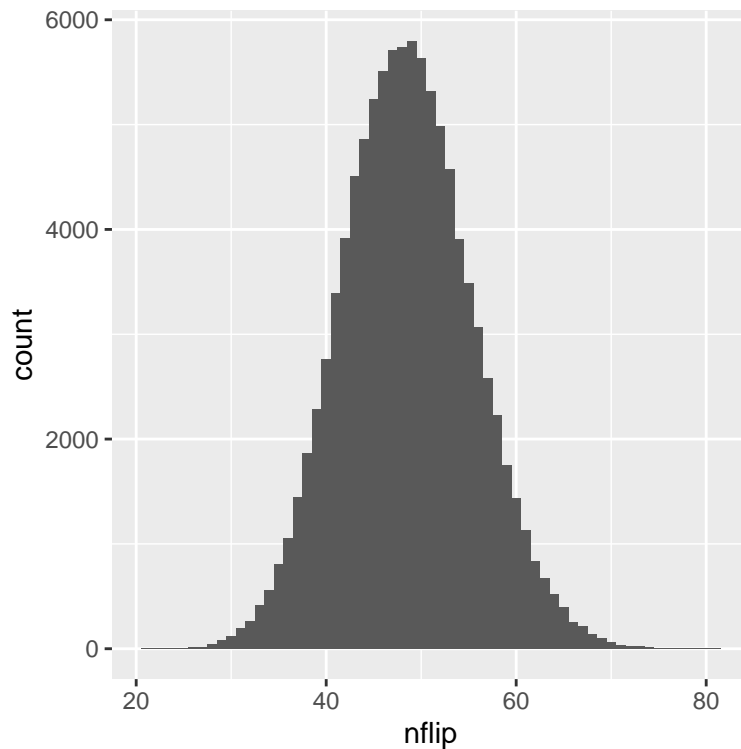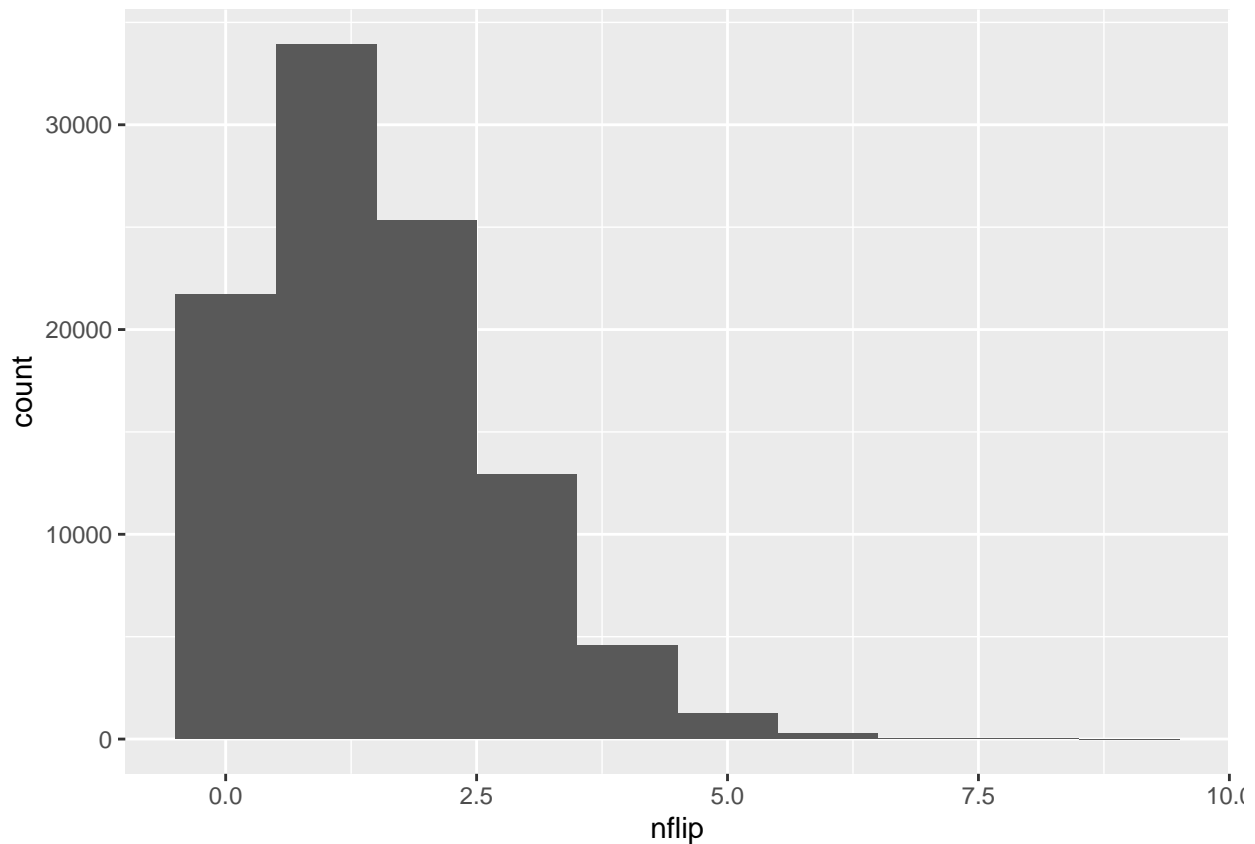
Problem 1 - Iron Bank

With the assumption of the null hypothesis being true, this histogram shows the distribution (of 100000 simulations) of the amount of flagged trades out of 2021 trades.

In this question, 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.

The test statistic used to measure evidence against the null hypothesis is the number of flagged trades out of 2021 trades. Specifically, we are observing the amount of simulations where there are 70 or more flagged trades in 2021 trades.

The p-value is approximately 0.0019, which means that, there is only a 0.19% probability of observing 70 or more flagged trades in 2021 trades out of chance if the true flagging rate was 2.4%. This means that we have strong evidence against any plausibility of the null hypothesis because the Iron Bank is most likely engaging often in illegal trade (more than the baseline rate).

Problem 2 - Health Inspections

With the assumption of the null hypothesis being true, this histogram shows the distribution (of 100000 simulations) of the amount of health code inspections (out of 50) from Gourmet Bites.

In this question, the null hypothesis that is being tested is that the rate of health code violations for Gourmet Bites is the same as the citywide average, which is 3%.

The test statistic used to measure evidence against the null hypothesis is the number of health code violations from Gourmet Bites out of 50. Specifically, we are observing the amount of simulations where there are 8 or more health code violations from Gourmet Bites out of 50.

The p-value is approximately 0.0001, which means that there is only a 0.01% probability of observing 8 or more health code violations out of 50 by chance if the true rate of health code violations for Gourmet Bites was the same as the citywide average (3%). This means that we have strong evidence against any plausibility of the null hypothesis because Gourmet Bites is most likely violating health code inspections at a higher rate than the citywide average at 3%.

Problem 3 - Evaluating Jury Selection for Bias

Histogram of Simulated Chi–Squared Values
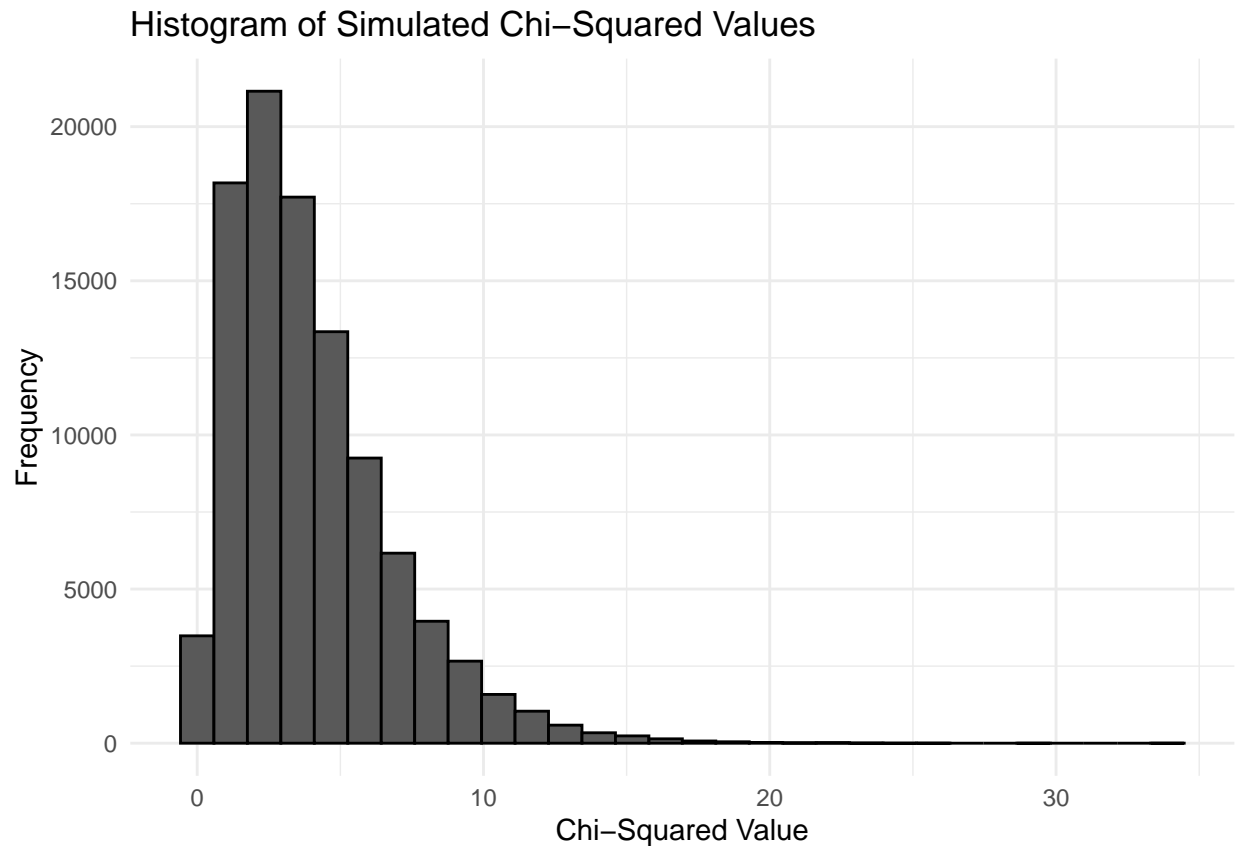
This histogram displays the distribution of the chi-squared values from the 100,000 simulated jury selections under the assumption that the jury selection process is unbiased.

In this question, we are testing whether the jury selection process is biased. Specifically, the null hypothesis being tested is that the distribution of jurors across different groups follows the same expected distribution, meaning that there is no bias in the jury selection process.

The test statistic used to measure evidence against the null hypothesis is the chi-squared statistic. Since we're using the chi-squared statistic within 20 trials, there are 240 jurors.

The p-value is approximately 0.0143, meaning that there is a 1.43% probability of observing a chi-squared value as extreme as or more extreme than the observed one, purely due to random chance, assuming the null hypothesis is true. This means that the null hypothesis should be rejected, as the distribution of jurors empaneled by this judge is significantly different from the county's population proportions. Most likely, there is systematic bias in jury selection by this judge, but it's not 100% definite. Other explanations that may explain the bias could be the difference of work schedules and even socioeconomic factors. To investigate this further and come to an even clearer conclusion, we could explore the pool of jurors, and specifically particular demographic groups are less likely to be called for jury duty based on factors like availability or location in the initial selection pool.

Problem 4 - LLM watermarking

Part A - Null or Reference Distribution
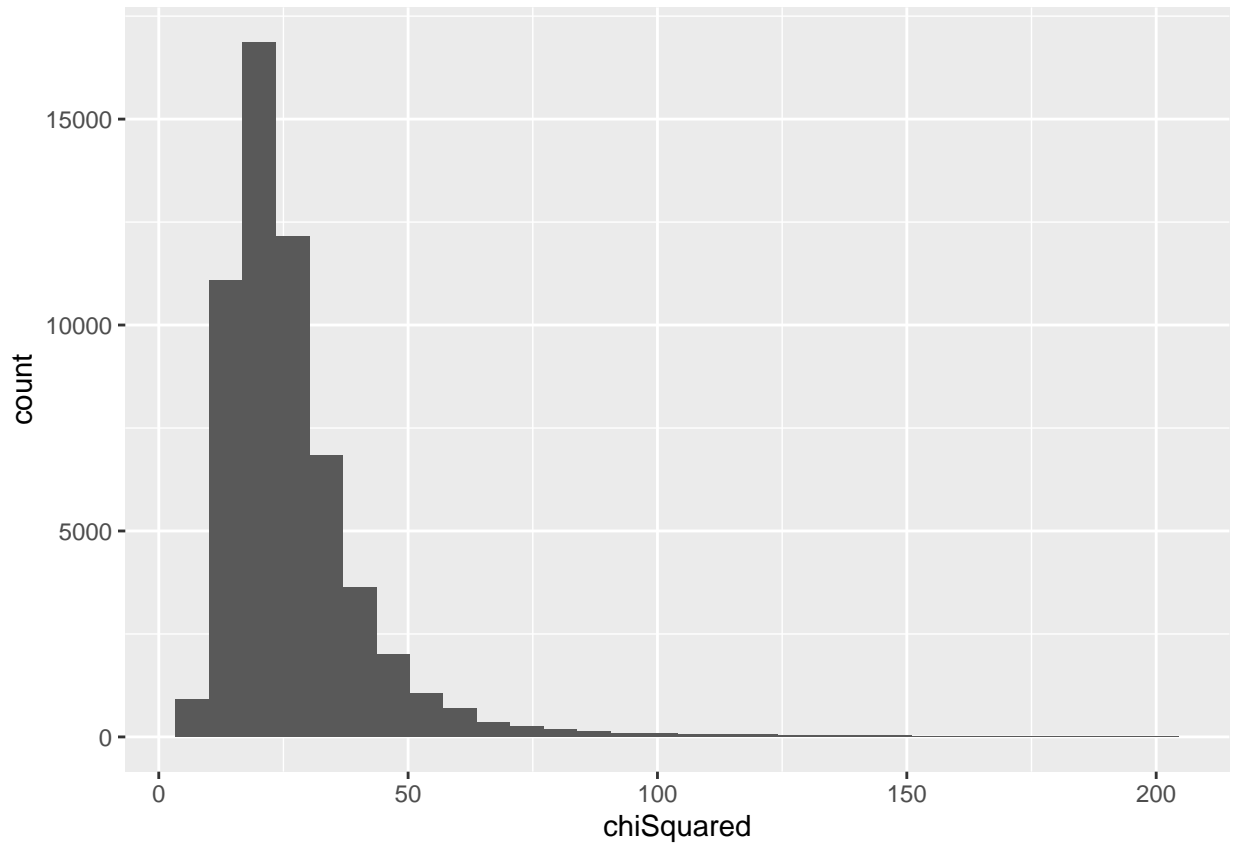
This displays the null distribution of the chi-squared test statistic based on letter frequency by using a collection of sentences from the Brown Corpus.

Problem B - Checking for a Watermark

Sentence	ChiSquared	PValue
She opened the book and	22.931	0.513
Despite the heavy rain, they	13.051	0.926
The museum's new exhibit features	46.286	0.076
He carefully examined the document,	23.546	0.489
The students gathered in the	23.676	0.484
Feeling vexed after an arduous	96.453	0.009
The chef demonstrated how to	28.271	0.328
They watched the sunset from	9.635	0.988
The committee reviewed the proposal	44.929	0.084
Despite the challenges faced during	49.961	0.059

Sentence 6 has the lowest p-value at 0.009 with the highest Chi-Squared value. This means that the observed letter frequencies in that sentence deviate the most from what we would expect in typical English text. As a result, this means it has an unusual or non-typical letter distribution than what we'd expect written from a human, which means that this sentence has been watermarked by an LLM.