Homework 5

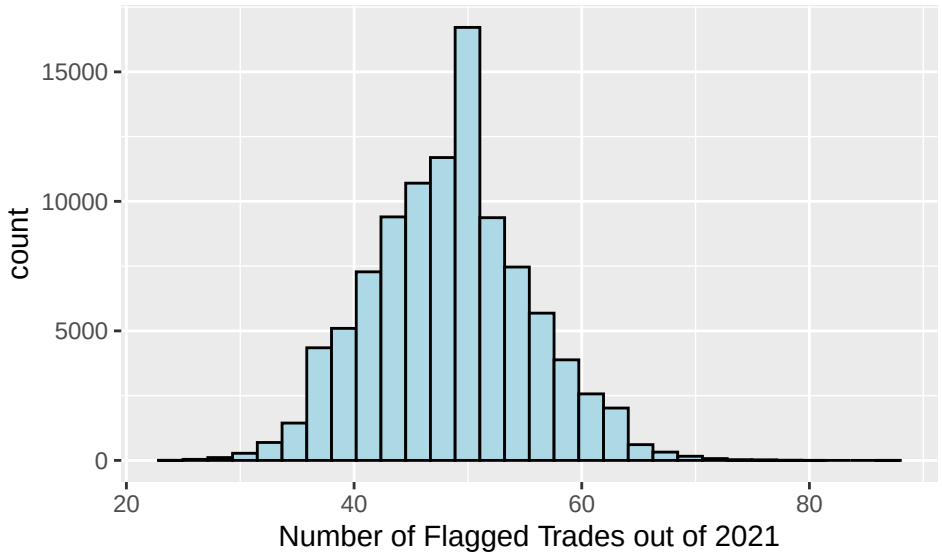
Isaiah Benny EID: ieb357

2024-02-28

Github

Question 1

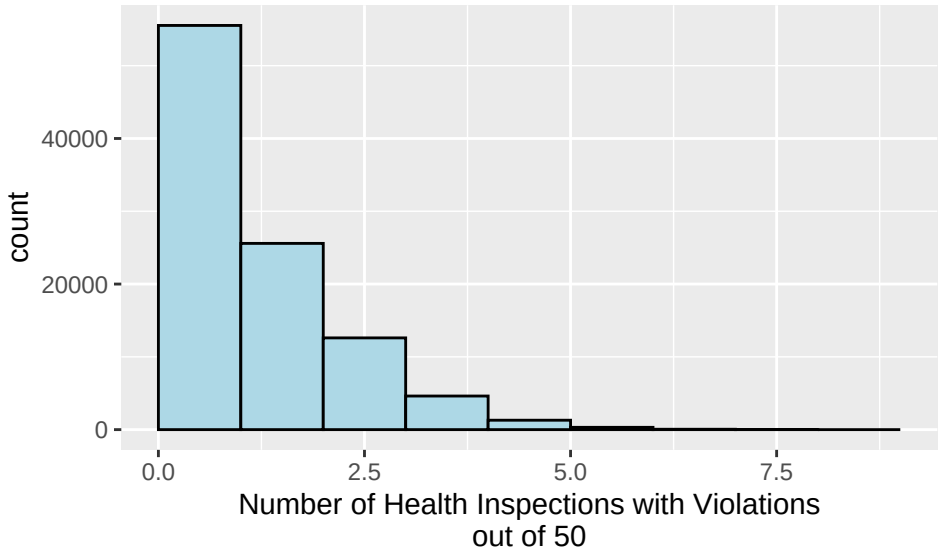
Our null hypothesis is that Iron Bank employees have their trades flagged with a rate of 2.4% on average in the long run. Our test statistic is the number of flagged trades out of 2021 trades.



The histogram above shows the probability distribution of the test statistic, assuming that the null hypothesis is true. Our p-value is 0.00195. In other words, assuming that the null hypothesis is true, the probability of seeing 70 or more flagged trades out of 2021 is 0.00195. Because this p-value is decently below 0.05, it is not very plausible for the null hypothesis to be true; 70 flagged trades out of 2021 would be fairly rare, so it is certainly possible that Iron Bank employees are having their trades flagged at a higher rate.

Question 2

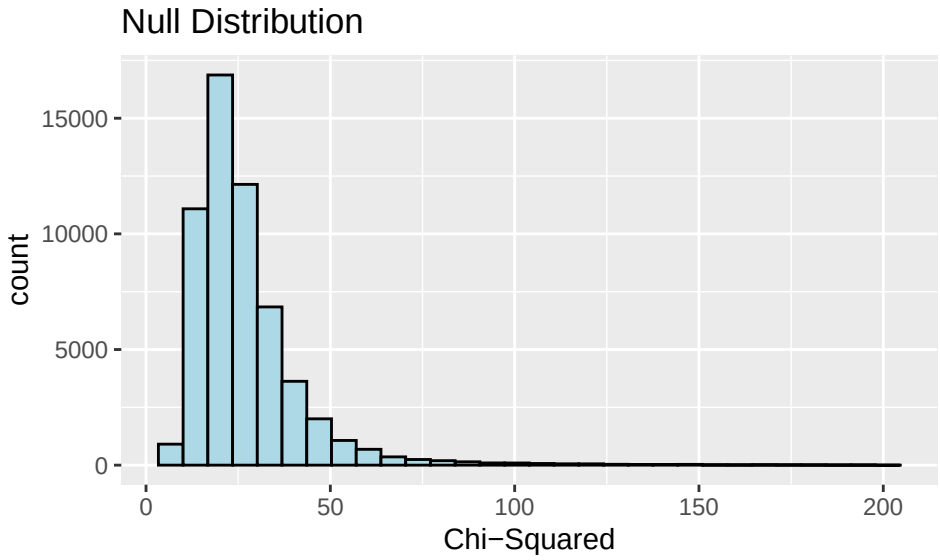
Our null hypothesis is that 3% of Gourmet Bites inspections result in health code violations on average in the long run. Our test statistic will be the number of health inspections that result in a reported health code violation out of 50.



This histogram shows the probability distribution of the test statistic, assuming that the null hypothesis is true. Our p-value is 0.00008. Since this p-value is far below 0.05, it would be very rare to observe a count of 8 health code violations if we assume that Gourmet Bites has health code violations at a rate of 3%, meaning that the null does not seem to be plausible.

Question 3

Part A



This histogram shows the distribution of chi-squared values for normal English sentences based on our defined letter frequencies.

Part B

	P-Values
Sentence 1	0.513
Sentence 2	0.926
Sentence 3	0.076
Sentence 4	0.489
Sentence 5	0.484
Sentence 6	0.009
Sentence 7	0.328
Sentence 8	0.988
Sentence 9	0.084
Sentence 10	0.059

The table above shows the chi-squared statistic based on letter frequencies for each sentence. Based on this table, it appears as if sentence 6 is the sentence that has been watermarked. I know this because the p-value associated with the chi-square statistic of that sentence is the lowest of all 10 by a decent amount, and it is also below 0.05.