Homework4SDS315

Kapil Taspa

2025-02-19

UT EID: kt27955

GitHub Repo: https://github.com/ktaspa/Homework4SDS315

Problem 1 - Iron Bank

```
## [1] 47
```

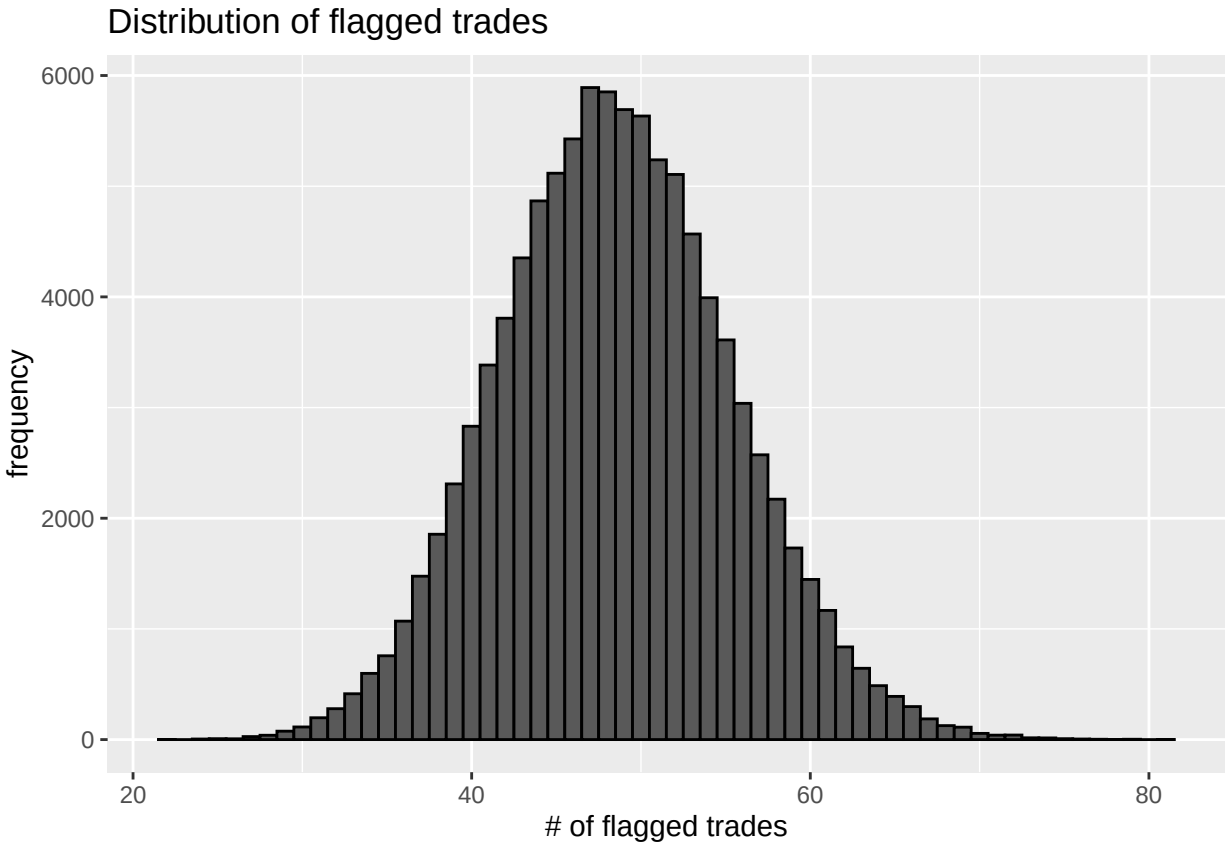
Null Hypothesis

The proportion of flagged trades is 2.4% at the baseline.

Test Statistic

Higher **number of flagged trades in 2021** means stronger evidence for the null hypothesis.

Histogram of Flagged Trades



P-value

```
## [1] 0.00189
```

Conclusion

p-value<0.05 therefore we reject the null hypothesis which means the flagged rate is higher than the baseline for Iron Bank.

Problem 2 - Health Inspections

```
## [1] 5
```

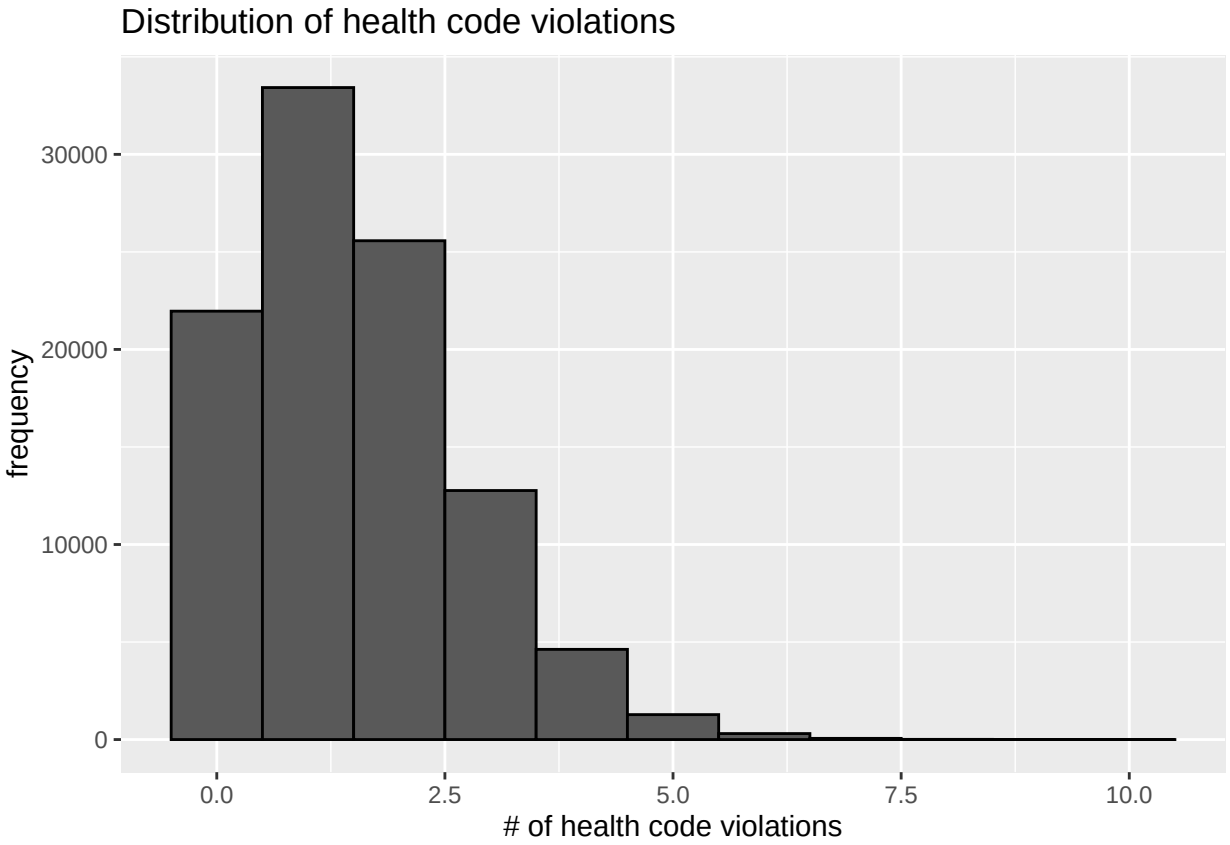
Null Hypothesis

The proportion of health code violations is 3% at the baseline.

Test Statistic

Higher **number of health code violations in 50 inspections** means stronger evidence for the null hypothesis.

Histogram of Flagged Trades



P-value

```
## [1] 0
```

Conclusion

p-value<0.05 therefore we reject the null hypothesis which means the health code violation rate is higher than the baseline for Gourmet Bites.

Problem 3 - Evaluating Jury Selection for Bias

Null Hypothesis

The distribution of empaneled jurors is the same as the county's population proportions.

Alternative Hypothesis

The distribution of empaneled jurors is different than the county's population proportions.

Test Statistic

Chi Square Test.

P-value

```
## [1] 0.0144468
```

Conclusion

p-value(0.01445)<0.05 therefore we reject the null hypothesis which means distribution of empaneled jury is significantly different from the county's population proportions with 95% confidence. This could suggest that systematic bias exist in jury selection, however, there are many other reasons that could have caused this such as bias thats done on purpose so the process is fair. More investigation can be done by comparing multiple different jury's.

Problem 4 - LLM Watermarking

Part A

```
##    Sentence ChiSquared
## 1         1   27.56914
## 2         2   34.16732
## 3         3   50.39108
## 4         4   30.34946
## 5         5   34.26534
## 6         6   36.63042
## 7         7   20.26351
## 8         8   32.49365
## 9         9   29.52727
## 10       10   48.71959
```

Part B

```
##    Sentence ChiSquared     P_Value
## 1         1  22.930848 0.512838136
## 2         2  13.051050 0.926284254
## 3         3  46.285861 0.076376773
## 4         4  23.546278 0.489065116
## 5         5  23.676149 0.484060270
## 6         6  96.452677 0.008776104
## 7         7  28.271419 0.327923165
## 8         8   9.635023 0.988016565
## 9         9  44.928631 0.084060270
## 10       10  49.960559 0.059036038
```

```
## The most anomalous sentence is sentence number: 6
```