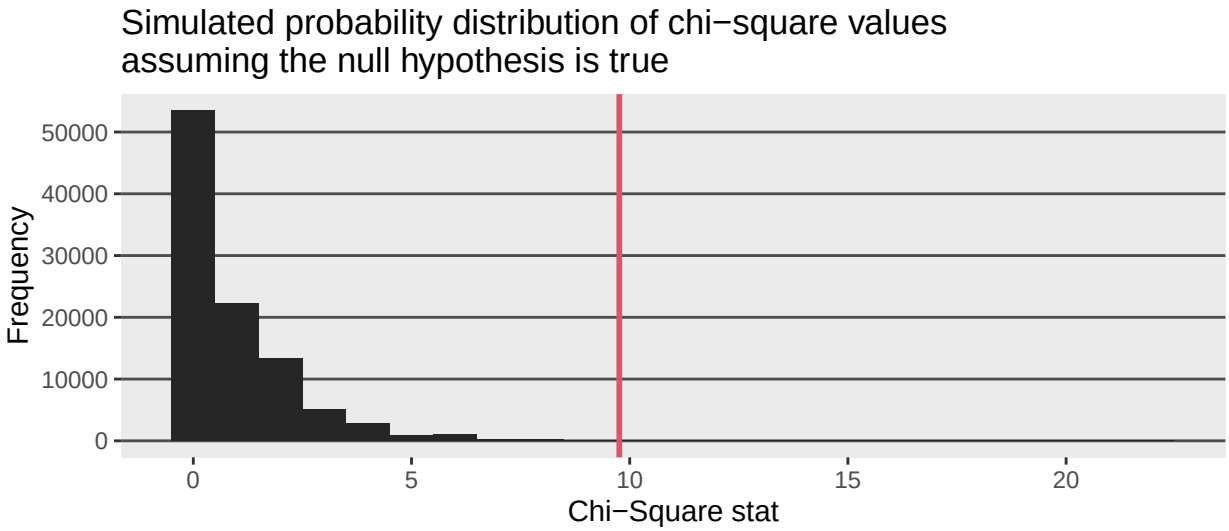
Hw 4

Jaxon Lara

jtl3236

(Github_R_Code)

Problem 1:



Observed Chi−Square value: 9.761

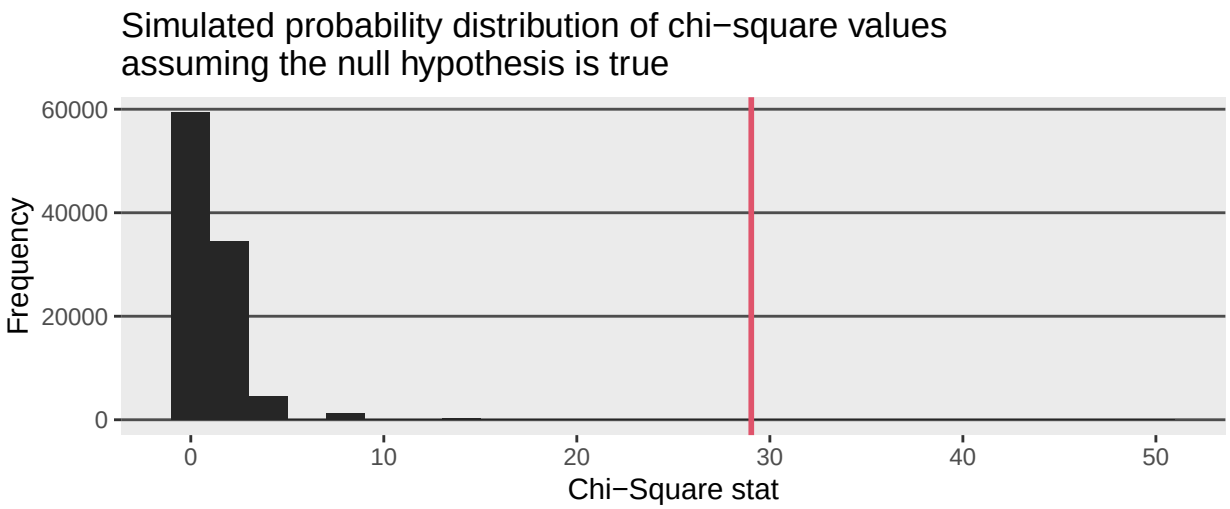
Null hypothesis: The observed data is consistent with the SEC's assertion that trades from the Iron Bank are flagged at the same 2.4% baseline rate as that of other traders.

Test statistic: Chi−Square value using the expected proportion of flagged trades being 2.4% to create a simulated distribution to compare the observed data to.

p−value: 0.002

Conclusion: Using a=0.05, the p−value is statistically significant so the null hypothesis is false, the observed data is not consistent with the SEC's assertion.

Problem 2:



Observed Chi Square value: 29.038

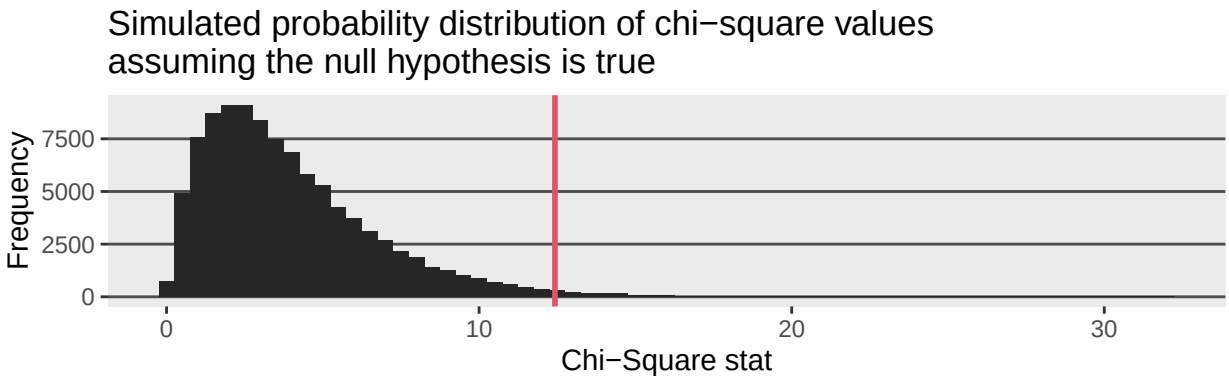
Null hypthesis: Gourmet Bites' rate of health code violations is in line with the average rate 3% of all restaurant inspections result in health code violations.

Test statistic: Chi square value using the expected rate of health code violations to create a simulated distribution to compare Gourney Bites' rate to.

p−value: 0.0001

Conclusion: Using a=0.05, the p−value is statistically significant so the null hypothesis is false. Gourmet Bites' rate of health code violations is not in line with the average rate of health code violations.

Problem 3:



Observed chi square value: 12.426

Null hypothesis: The distribution of jurors empaneled by this judge is not significantly different from the country's population proportions.

Test statistic: Chi square value using the expected demographic proportions of the population to create a simulated distribution to compare the judge's demographic proportion to.

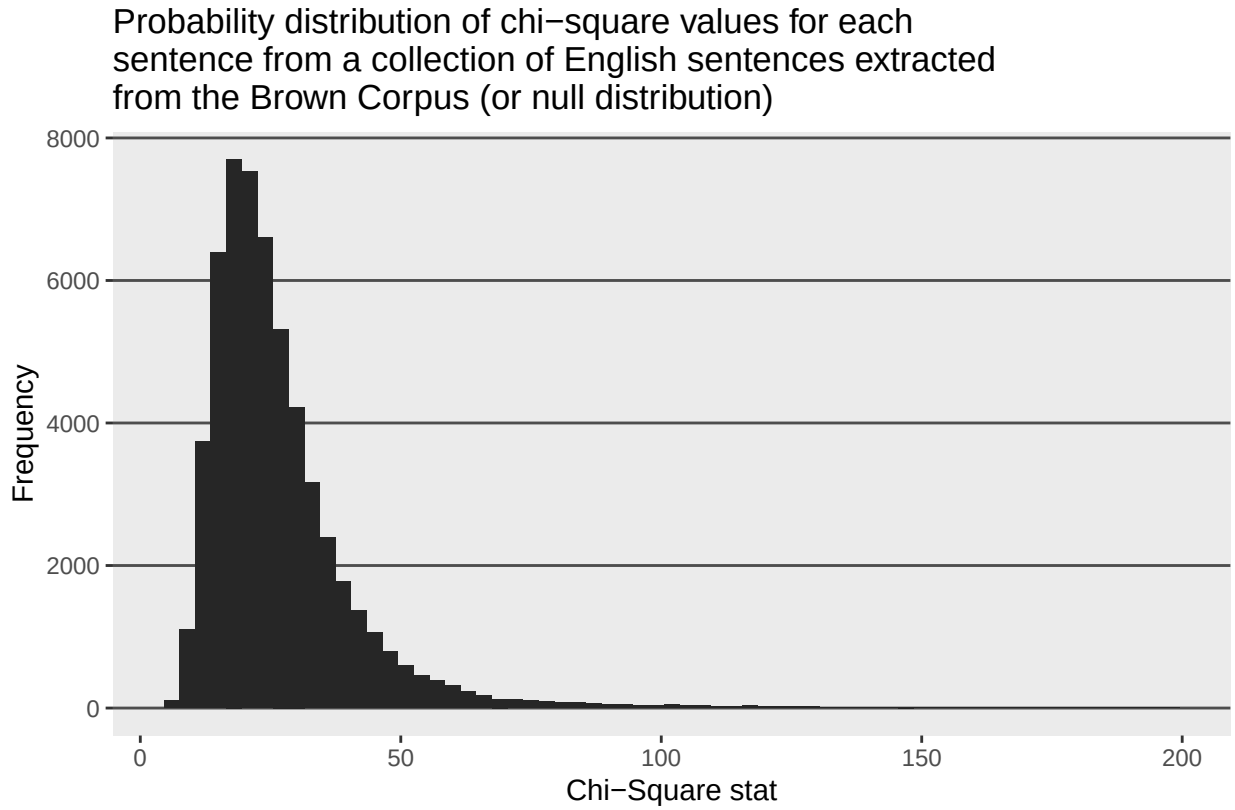
p−value: 0.015

Conclusion: Using a=0.05, the p−value is statistically significant so the null hypothesis is false. The Judge's empaneled jurors demographic proportions is significantly different from the country's population proportions which suggests some bias in jury selection.

Other explanations: It may be possible that the population proportions of where the judge is located is significantly different from the country's population proportions. I would investigate further by comparing the judge's empaneled jurors demographic proportions to the population proportion of the state/province/city that the judge is based out of.

Problem 4:

(A)


Probability distribution of chi−square values for each sentence from a collection of English sentences extracted from the Brown Corpus (or null distribution)
8000
6000
4000
2000
0
Frequency
0
50
100
150
200
Chi−Square stat

(B)

Sentence	PVal
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

It is the 6th 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." because its p-value of 0.009 is less than a=0.05 which is statistically significant. It is also because it has the lowest p-value compared to the other 9 sentences with the next highest being ~10x larger.