Sports Betting Strategy Simulation Using R

Lei Zhao

Introduction

This project analyzes a betting strategy on a baseball series between the Boston Red Sox and New York Yankees. The analysis evaluates three scenarios:

1. **Part 1**: A best-of-three series with games played in Boston, New York, and Boston (if needed).
2. **Part 2**: A best-of-three series with games played in New York, Boston, and New York (if needed).
3. **Part 3**: A best-of-five series alternating between Boston and New York.

The goal is to determine whether the betting strategy (win 500 per Red Sox victory, lose 520 per loss) is favorable.

Methedology

Data and Assumptions

- Red Sox home game win probability: 0.6.
- Yankees home game win probability: 0.57.
  - Red Sox away game win probability: $1-0.57=0.43$
- Payouts: + 500 per win,− 520 per loss.
- Simulations: 10,000 iterations for each series.

Analytical Tools

- **Theoretical Probability Distributions**: Calculated for net winnings, expected value (E(X)), and standard deviation (σ).
- **Monte Carlo Simulation**: Estimated E(X) and 95% confidence intervals.

- **Chi-squared Goodness-of-Fit Test**: Compared simulated frequencies to theoretical distributions.

Results

Part 1: Best-of-Three Series (Boston, New York, Boston)

(i) Probability of Red Sox Winning the Series

Steps:

1. Define win probabilities:
    - Game 1 (Boston): 0.6.
    - Game 2 (New York): 0.43.
    - Game 3 (Boston): 0.6.
2. Calculate winning scenarios:
    - **Scenario 1**: Win both games (Boston, New York). Probability = $0.6 \times 0.43 = 0.258$
    - **Scenario 2**: Win first, lose second, win third. Probability = $0.6 \times 0.57 \times 0.6 = 0.2052$
    - **Scenario 3**: Lose first, win next two. Probability = $0.4 \times 0.43 \times 0.6 = 0.1032$
3. Total probability:$0.258 + 0.2052 + 0.1032 = 0.5664$ (56.64%).

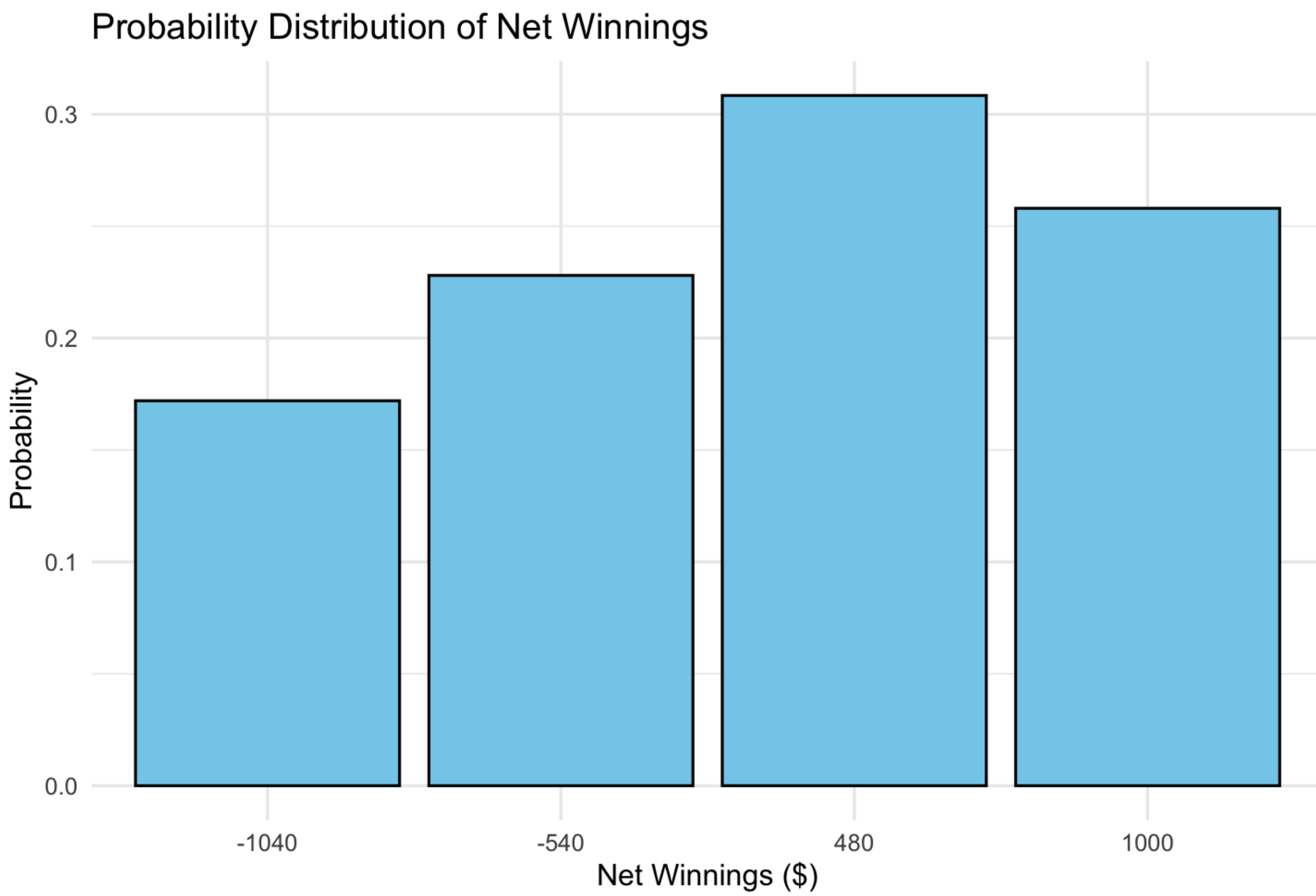
```r
win_first <- 0.6    # the first game is held in Boston
win_second <- 1 - 0.57 # the second game is held in New York
win_third <- 0.6    # the thrid game is held in Boston if necessary

# if Red Sox would win the series, they should win :
# either the firs two games
# or win first, lose second, win third
# or lose first, win second, win therd
senario1 <- win_first * win_second
senario2 <- win_first * (1 - win_second) * win_third
senario3 <- (1 - win_first) * win_second * win_third
win_series <- senario1 + senario2 + senario3
print(win_series)
```

```
## [1] 0.5664
```

(ii) Theoretical Probability Distribution

- **Net Winnings Outcomes**: 1,000(2−0),480 (2-1), -540(1−2),−1,040 (0-2).
- **Probabilities**:
  - $1,000: 25.8%.
  - $480: 30.84% (0.2052 + 0.1032).
  - -$540: 22.8%.
  - -$1,040: 17.2%.
- **Expected Value**:
  $E(X) = (1000\times0.258)+(480\times0.3084)+(-540\times0.228)+(-1040\times0.172) = 198.11$dollars.
- **Standard Deviation**:
  σ = sqrt(∑(X−E(X))^2 · P(X))= 782.73 dollars.
- **Visualization**:



(iii) Simulation and Confidence Interval

- Conducted a simulation of 10,000 series to estimate the expected net win and calculate a 95% confidence interval.
- The simulation resulted in an estimated mean net win of $198.112, with a 95% confidence interval of approximately [$192.925, $223.431].

- The theoretical expected value of $198.11 falls within the 95% confidence interval, suggesting that the simulation accurately reflects the theoretical distribution.

```
# 7. Print results
print(paste("Theoretical E(X):", expected_value_corrected))
```

```
## [1] "Theoretical E(X): 198.112"
```

```
print(paste("95% CI: [", conf_interval[1], ",", conf_interval[2], "]"))
```

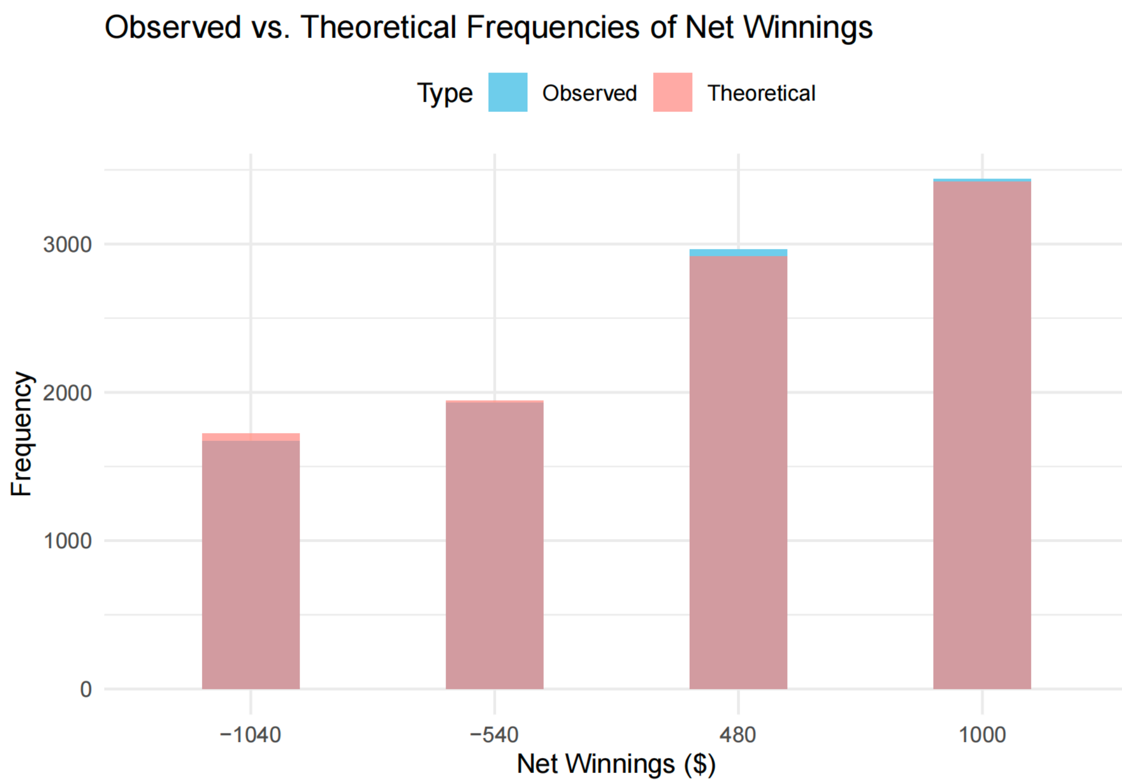
```
## [1] "95% CI: [ 192.924709938202 , 223.431290061798 ]"
```

```
print(paste("Is E(X) within CI?", is_within))
```

```
## [1] "Is E(X) within CI? TRUE"
```

(iv) Chi-squared Goodness-of-Fit Test

- Performed a Chi-squared test to compare the observed frequencies from the simulation with the theoretical probabilities.



- The test resulted in a p-value of 0.4839, indicating that there is no significant difference between the observed and expected frequencies, thus validating the theoretical distribution.

```r
# Perform Chi-squared test with consolidated probabilities
chisq_result <- chisq.test(observed_freq, p = probabilities_unique)
print(chisq_result)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  observed_freq
## X-squared = 2.4525, df = 3, p-value = 0.4839
```

(v) Betting Strategy Analysis

- The positive expected value of $198.11 suggests that the betting strategy is favorable under the conditions of Part 1.
- The risk of loss is present, as indicated by the standard deviation of $782.73, which shows the variability in potential outcomes.
- Overall, the strategy appears profitable but with a significant element of risk.

Part 2: Best-of-Three Series (New York, Boston, New York)

(i) Probability of Red Sox Winning the Series

- Calculated the probability of the Red Sox winning the series under the new game order.
- The total probability of winning the series is approximately 0.479 or 47.9%.

```
## [1] "Probability Red Sox win the series (Part 2): 0.47902"
```

(ii) Theoretical Probability Distribution

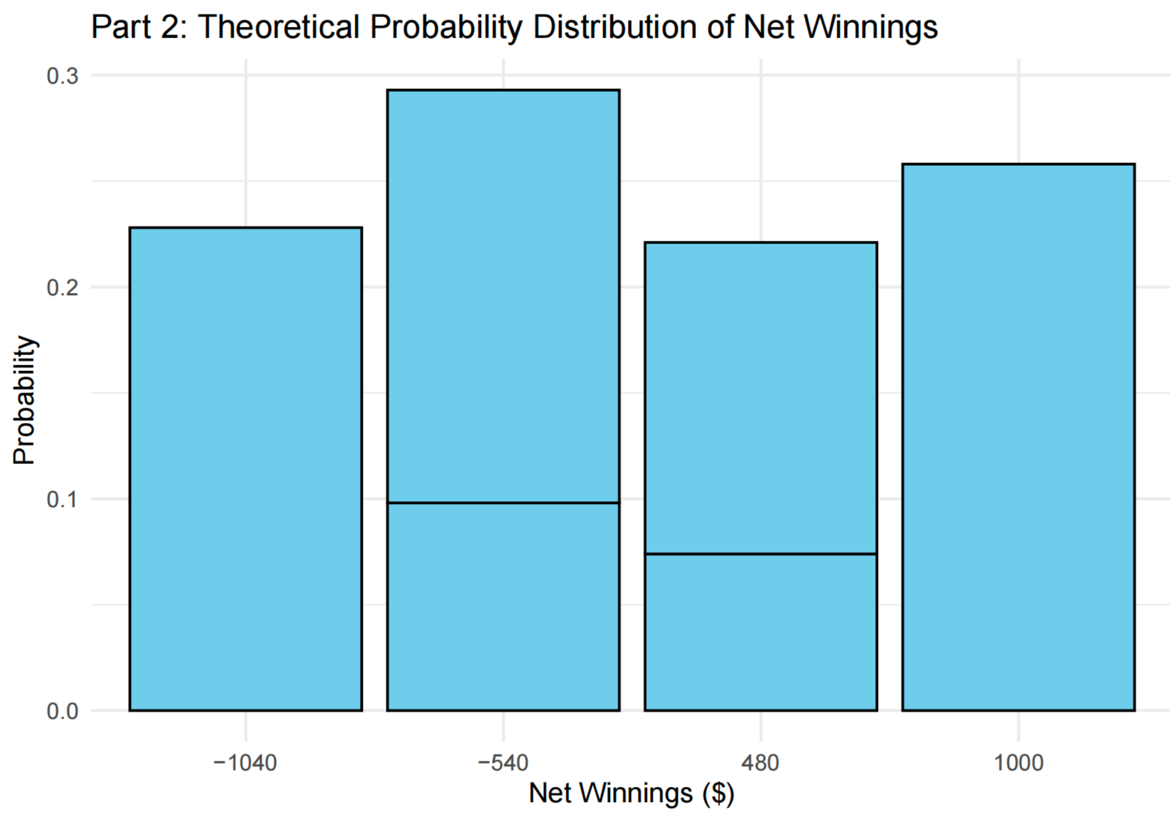
- Constructed a new probability distribution for net winnings based on the updated game order.
- The expected value of net winnings is negative (-$31.24), indicating an unfavorable betting scenario under these conditions.
- The standard deviation is $799.99, showing a high variability in potential outcomes.

```
print(paste("Expected Value (Part 2):", expected_value_part2))
```

```
## [1] "Expected Value (Part 2): -31.2396"
```

```
print(paste("Standard Deviation (Part 2):", std_dev_part2))
```

```
## [1] "Standard Deviation (Part 2): 799.990539563963"
```



(iii) Monte Carlo Simulation

- Simulated 10,000 series outcomes to estimate the expected net win and calculate a 95% confidence interval.
- The simulation resulted in an estimated mean net win of approximately -$31.24, with a 95% confidence interval of approximately [-$58.205, -$26.755].
- The theoretical expected value falls within the confidence interval, confirming the simulation's accuracy.

```r
print(paste("95% CI for Part 2:", ci_part2[1], "to", ci_part2[2]))
```

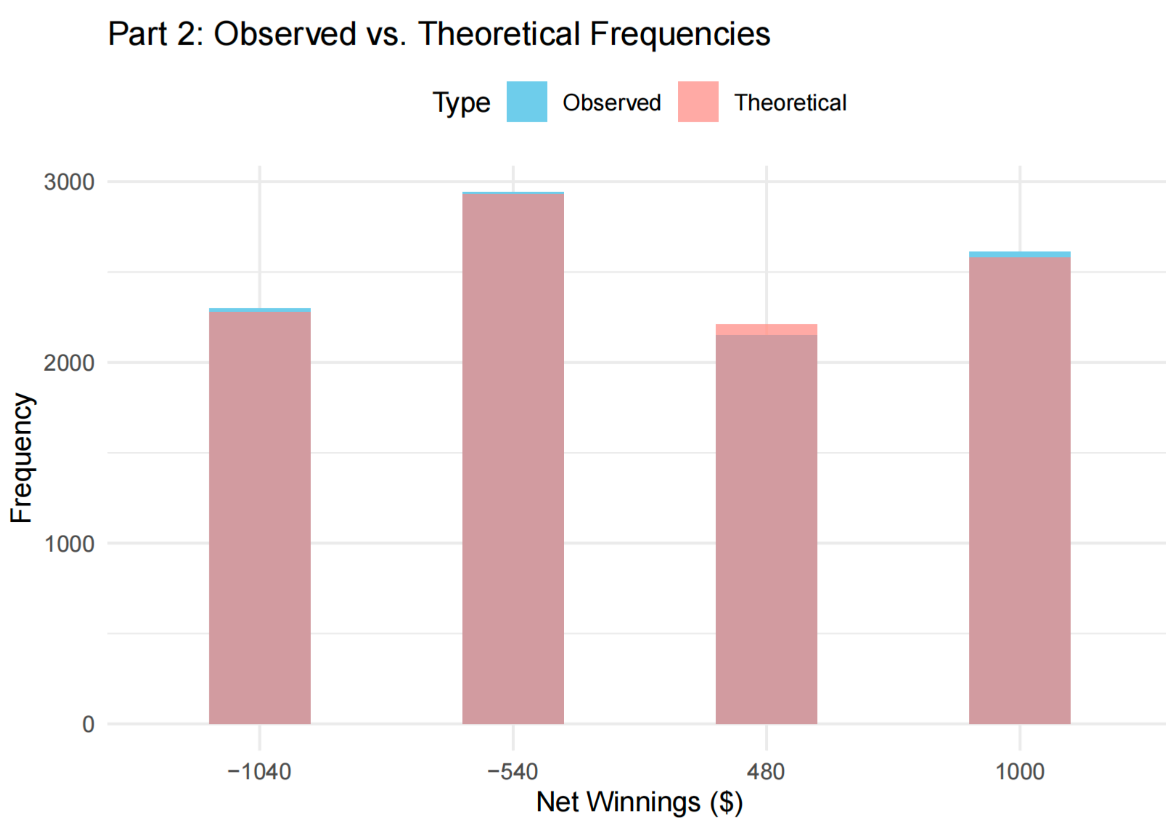
```
## [1] "95% CI for Part 2: -58.2047916359029 to -26.7552083640971"
```

```r
print(paste("Theoretical E(X) in CI?:",
            expected_value_part2 >= ci_part2[1] && expected_value_part2 <= ci_part2[2]))
```

```
## [1] "Theoretical E(X) in CI?: TRUE"
```

(iv) Chi-squared Goodness-of-Fit Test

- Conducted a Chi-squared test to compare the observed frequencies from the simulation with the theoretical probabilities.



- The test resulted in a p-value of 0.5186, indicating no significant difference between the observed and expected frequencies.

```
##
##  Chi-squared test for given probabilities
##
## data:  observed_part2
## X-squared = 2.2682, df = 3, p-value = 0.5186
```

(v) Betting Strategy Analysis

- The negative expected value suggests that the betting strategy is not favorable under the conditions of Part 2.
- The high variability in outcomes, as indicated by the standard deviation, further emphasizes the risk associated with this scenario.

Part 3: Best-of-Five Series (Alternating Boston and New York)

(i) Probability of Red Sox Winning the Series

- Calculated the probability of the Red Sox winning the best-of-five series.
- The total probability of winning the series is approximately 0.5609 or 56.09%.

```
## [1] "Probability Red Sox win best-of-five: 0.5608944"
```

(ii) Theoretical Probability Distribution

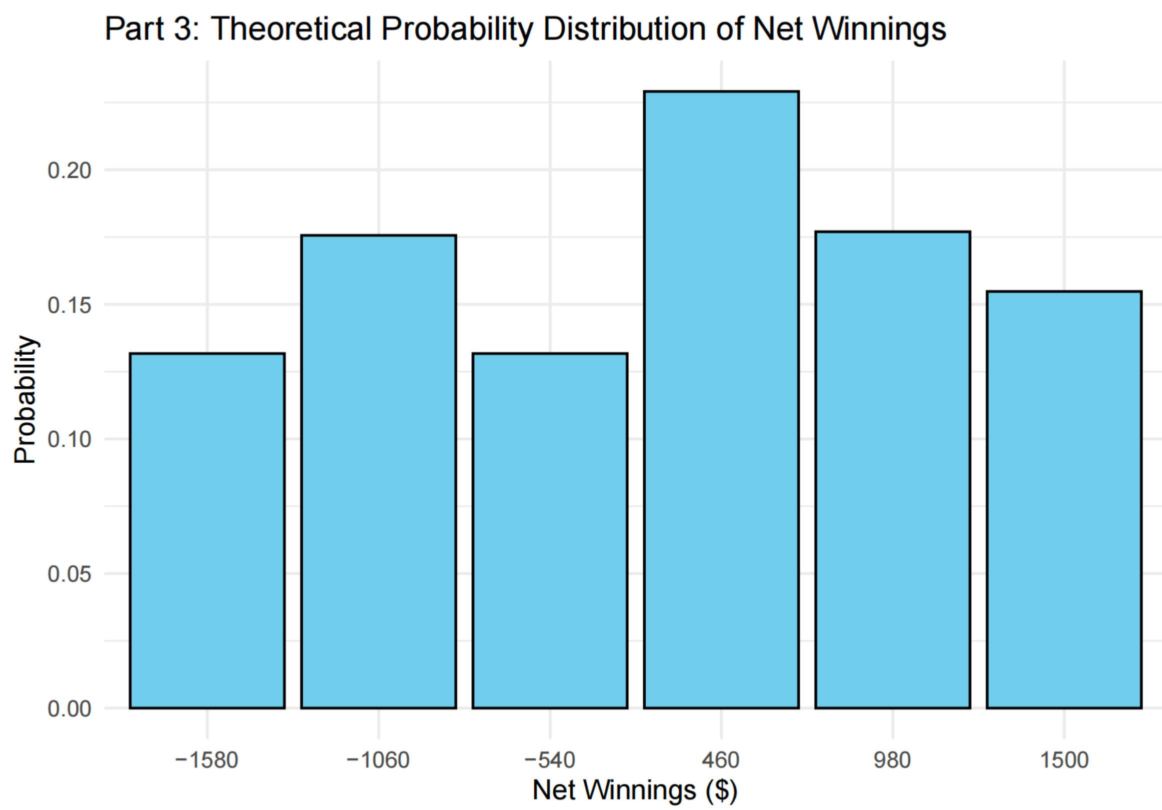
- Constructed a probability distribution for net winnings based on the best-of-five format.
- The expected value of net winnings is positive ($45.59), indicating a potentially favorable betting scenario.
- The standard deviation is $1062.69, showing a high variability in potential outcomes.

```r
print(paste("Expected Value (Part 3):", expected_value_part3))
```

```
## [1] "Expected Value (Part 3): 45.585248"
```

```r
print(paste("Standard Deviation (Part 3):", std_dev_part3))
```

```
## [1] "Standard Deviation (Part 3): 1062.68514589636"
```



(iii) Monte Carlo Simulation

- Simulated 10,000 series outcomes to estimate the expected net win and calculate a 95% confidence interval.
- The simulation resulted in an estimated mean net win of approximately $45.59, with a 95% confidence interval of approximately [$21.191, $63.057].
- The theoretical expected value falls within the confidence interval, confirming the simulation's accuracy.

```r
print(paste("95% CI for Part 3:", ci_part3[1], "to", ci_part3[2]))
```

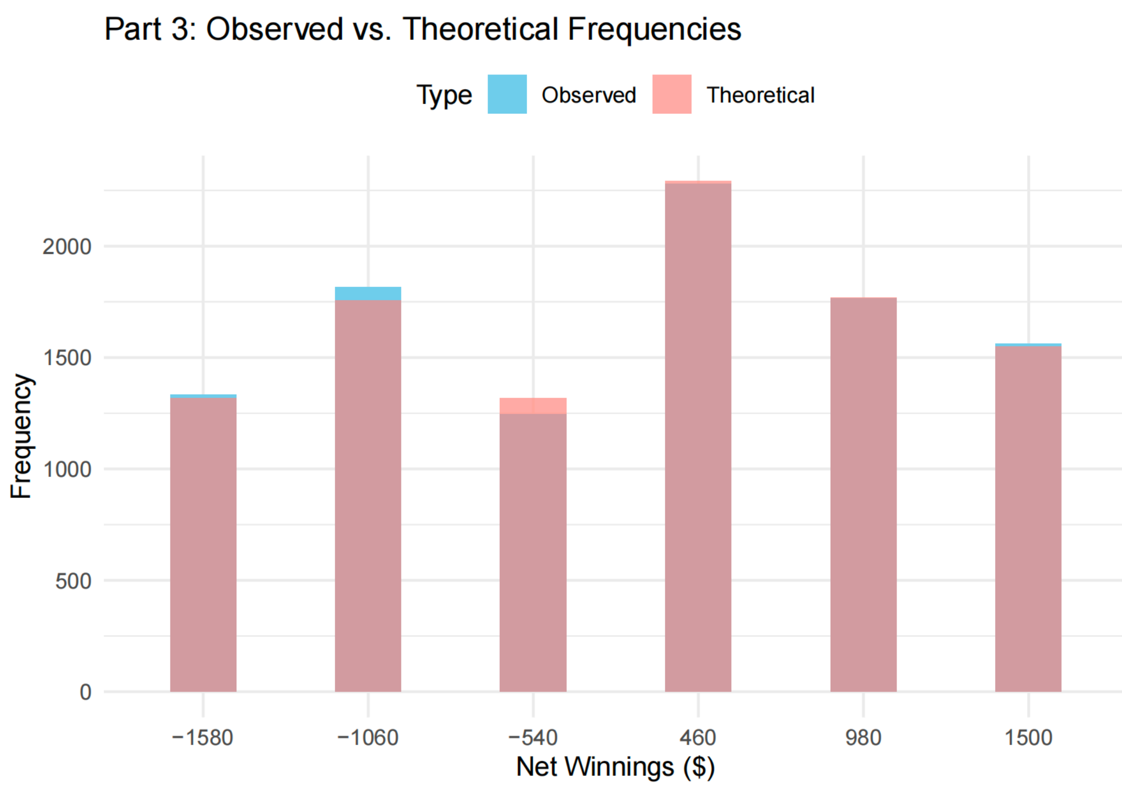
```
## [1] "95% CI for Part 3: 21.1909405275054 to 63.0570594724947"
```

```r
print(paste("Theoretical E(X) in CI?:",
            expected_value_part3 >= ci_part3[1] && expected_value_part3 <= ci_part3[2]))
```

```
## [1] "Theoretical E(X) in CI?: TRUE"
```

(iv) Chi-squared Goodness-of-Fit Test

- Conducted a Chi-squared test to compare the observed frequencies from the simulation with the theoretical probabilities.



- The test resulted in a p-value of 0.2707, indicating no significant difference between the observed and expected frequencies.

```
##
##  Chi-squared test for given probabilities
##
## data:  observed_part3
## X-squared = 6.3829, df = 5, p-value = 0.2707
```

(v) Betting Strategy Analysis

- The positive expected value suggests that the betting strategy is favorable under the conditions of Part 3.
- Despite the high variability in outcomes, the potential for profit makes this scenario more attractive for betting.

Conclusion

This project provides a comprehensive analysis of a betting strategy across three different series formats. The results indicate that the betting strategy is favorable in a best-of-three series with games played in Boston, New York, and Boston, and in a best-of-five series alternating between Boston and New York. However, the strategy is not favorable in a best-of-three series with games played in New York, Boston, and New

York. The analysis highlights the importance of considering the specific conditions of a series when evaluating betting strategies. Future research could explore additional factors that might influence the outcomes, such as team performance variations or changes in betting odds.

References

- Bluman, A. G. (2017). Elementary statistics: A step by step approach (10th ed.). McGraw-Hill Education.
- Kabacoff, R. I. (2015). R in action: Data analysis and graphics with R (2nd ed.). Manning Publications.