**Module One Project**

**Analysis of a Betting Strategy in Sports**

Mohit Ravindra Kamble

College of Professional Studies – Northeastern University

ALY - 6050: Introduction to Enterprise Analytics

Prof. Zhi (Richard) He

February 28, 2024



1. **Overview:**

In this assignment, this analysis examines the betting strategy for the Boston Celtics and Golden State Warriors NBA series in different scenarios. Initially, in a best-of-three format, a $150 bet on the Celtics for each game was considered. The scenario then shifted to a best-of-five format, with alternating games at TD Garden and Chase Center. The analysis further explores a best-of-seven NBA Finals scenario with a unique 2-3-2 format where the Celtics have home-court advantage. This comprehensive study aims to provide insights into the betting dynamics for varying playoff scenarios.

1. **Analysis:**

**Part - 1:**

I investigated a betting strategy for the Boston Celtics and Golden State Warriors NBA series. The series follows a best-of-three format, where the first game is at Chase Center, the second at TD Garden, and the third (if needed) back at Chase Center. I placed a bet of $150 on the Celtics to win each game, with a potential loss of $160 if they lose.

**i) Probability of Winning the Series:**

* The probability that the Celtics will win the series was calculated to be approximately 39.82%. This was determined by considering all possible outcomes of the three-game series.

**O/P:**

|  |
| --- |
| > cat("Probability that Celtics win the series:", celtics\_win\_series)  Probability that Celtics win the series: 0.39816 |

**ii) Net Win Distribution:**

* The expected net win for the series was calculated to be $120.91, with a standard deviation of $115.36. This means that, on average, I could expect to win $120.91 per bet.

**O/P:**

|  |
| --- |
| > cat("Expected net win:", expected\_win, "\n")  Expected net win: 120.9096  > cat("Standard deviation of net win:", standard\_deviation)  Standard deviation of net win: 115.3623 |

**iii) Simulation and Confidence Interval:**

* Simulating 5,000 random scenarios, the estimated expected net win was $303.63. *However, the 95% confidence interval for the simulation was $294.20 to $313.05, which did not contain the expected net win of $120.91.* This indicates a potential deviation from the expected outcome.

**O/P:**

|  |
| --- |
| > cat("Estimated expected net win (simulation):", estimated\_expected\_win, "\n")  Estimated expected net win (simulation): 303.628  > cat("95% confidence interval:", lower\_bound, "-", upper\_bound, "\n")  95% confidence interval: 294.2034 - 313.0526  > if (expected\_win >= lower\_bound & expected\_win <= upper\_bound) {  + cat("Yes, E(X) is within the 95% confidence interval.")  + } else {  + cat("No, E(X) is not within the 95% confidence interval.")  + }  No, the E(X) is not within the 95% confidence interval. |

**iv) Chi-square Goodness of Fit Test:**

* The chi-square goodness of fit test was employed to assess how well the simulated distribution matched the expected distribution. *The test resulted in rejecting the null hypothesis*, suggesting that the simulated data did not precisely match the expected distribution.

**O/P:**

|  |
| --- |
| cat("Chi-square test statistic:", chisq\_test$statistic, "\n")  Chi-square test statistic: 376.4768  > cat("p-value:", chisq\_test$p.value, "\n")  p-value: 0.000000000000000000000000000000000000000000000000000000000000000000000000000000000007278243  > if (chisq\_test$p.value > 0.05) {  + cat("We fail to reject the null hypothesis that the simulated data (Y) follows the expected distribution (X).")  + } else {  + cat("We reject the null hypothesis that the simulated data (Y) follows the expected distribution (X).")  + }  We reject the null hypothesis that the simulated data (Y) follows the expected distribution (X). |

**v) Betting Strategy Analysis Summary:**

* In summary, the expected net win per bet was $120.91, but the simulated average win was $303.63, outside the 95% confidence range. This inconsistency suggests that further analysis is needed to determine if the betting strategy is genuinely favorable.

**Part - 2:**

In this second part of the analysis, I revisited the betting strategy for the Boston Celtics and Golden State Warriors NBA series. The scenario changed, with the first game now played at TD Garden, the second at Chase Center, and the third (if needed) back at TD Garden. The betting amounts and probabilities remained consistent with the first part.

**i) Probability of Winning the Series:**

* The probability that the Celtics will win the series in this new scenario was found to be approximately 35.28%, considering all possible outcomes of the three-game series.

**O/P:**

|  |
| --- |
| > cat("Probability that Celtics win the series in Part 2:", celtics\_win\_series\_part2, "\n")  Probability that Celtics win the series in Part 2: 0.3528 |

**ii) Net Win Distribution:**

* The expected net win for the series in this scenario was calculated to be $108.36, with a standard deviation of $118.10. This means that, on average, I could expect to win $108.36 per bet.

**O/P:**

|  |
| --- |
| > cat("Expected net win in Part 2:", expected\_win\_part2, "\n")  Expected net win in Part 2: 108.36  > cat("Standard deviation of net win in Part 2:", standard\_deviation\_part2, "\n")  Standard deviation of net win in Part 2: 118.1015 |

**iii) Simulation and Confidence Interval:**

* Simulating 5,000 random scenarios, the estimated expected net win was $307.20. *However, the 95% confidence interval for the simulation was $298.18 to $316.00, and again, it did not contain the expected net win of $108.36.* This indicates a potential deviation from the expected outcome.

**O/P:**

|  |
| --- |
| > cat("Estimated expected net win in Part 2:", estimated\_expected\_win\_part2, "\n")  Estimated expected net win in Part 2: 307.196  > cat("95% confidence interval in Part 2:", lower\_bound\_part2, "-", upper\_bound\_part2, "\n")  95% confidence interval in Part 2: 298.3911 - 316.0009  > # Checking if E(X) is within the confidence interval for Part 2  > if (expected\_win\_part2 >= lower\_bound\_part2 & expected\_win\_part2 <= upper\_bound\_part2) {  + cat("Yes, E(X) in Part 2 is within the 95% confidence interval.")  + } else {  + cat("No, E(X) in Part 2 is not within the 95% confidence interval.")  + }  No, the E(X) in Part 2 is not within the 95% confidence interval. |

**iv) Chi-square Goodness of Fit Test:**

* The chi-square goodness of fit test was employed to assess how well the simulated distribution matched the expected distribution. *The test resulted in rejecting the null hypothesis*, suggesting that the simulated data did not precisely match the expected distribution in this scenario as well.

**O/P:**

|  |
| --- |
| > cat("Chi-square test statistic in Part 2:", chisq\_test\_part2$statistic, "\n")  Chi-square test statistic in Part 2: 964.4832  > cat("p-value in Part 2:", chisq\_test\_part2$p.value, "\n")  p-value in Part 2: 9.429264e-212  > # Interpreting the results for Part 2  > if (chisq\_test\_part2$p.value > 0.05) {  + cat("We fail to reject the null hypothesis that the simulated data (Y) in Part 2 follows the expected distribution (X).")  + } else {  + cat("We reject the null hypothesis that the simulated data (Y) in Part 2 follows the expected distribution (X).")  + }  We reject the null hypothesis that the simulated data (Y) in Part 2 follows the expected distribution (X). |

**v) Betting Strategy Analysis Summary:**

* In summary, the probability of the Celtics winning the series decreased to 35.28%, and the expected net win per bet decreased to $108.36 in this new scenario. The simulated average win was $307.20, outside the 95% confidence range of the expected win. The chi-square test again indicated potential variability in the betting strategy outcome for this scenario.

**Part - 3:**

In this third part of the analysis, the betting strategy for the Boston Celtics and Golden State Warriors NBA series was reexamined under a new scenario. The series format changed to a best-of-five, where the first team to win three games would secure victory. The games alternated between TD Garden and Chase Center, with the initial match played at TD Garden.

**i) Probability of Winning the Series (Best of Five):**

* The probability that the Celtics would win the best-of-five series was determined to be approximately 20.88%, considering all possible outcomes of the series.

**O/P:**

|  |
| --- |
| > cat("Probability that Celtics win the best of five series in Part 3:", celtics\_win\_series\_part3, "\n")  Probability that Celtics win the best of five series in Part 3: 0.2088 |

**ii) Net Win Distribution (Best of Five):**

* The expected net win for the best-of-five series was calculated to be $93.96, with a standard deviation of $162.69. On average, I could expect to win $93.96 per bet, considering the new series format.

**O/P:**

|  |
| --- |
| > cat("Expected net win in Part 3 (Best of Five):", expected\_win\_part3, "\n")  Expected net win in Part 3 (Best of Five): 93.96  > cat("Standard deviation of net win in Part 3 (Best of Five):", standard\_deviation\_part3, "\n")  Standard deviation of net win in Part 3 (Best of Five): 162.6912 |

**iii) Simulation and Confidence Interval (Best of Five):**

* Simulating 5,000 random scenarios, the estimated expected net win was $97.47. The 95% confidence interval for the simulation ranged from approximately -$265.88 to $460.82. Surprisingly, *this interval contained the expected net win of $93.96, suggesting a potential for significant variability in the outcomes.*

**O/P:**

|  |
| --- |
| > cat("Estimated expected net win in Part 3 (Best of Five):", estimated\_expected\_win\_part3, "\n")  Estimated expected net win in Part 3 (Best of Five): 97.47  > cat("95% confidence interval in Part 3 (Best of Five):", lower\_bound\_part3, "-", upper\_bound\_part3, "\n")  95% confidence interval in Part 3 (Best of Five): -265.8799 - 460.8199  > # Checking if E(X) is within the confidence interval for Part 3  > if (expected\_win\_part3 >= lower\_bound\_part3 & expected\_win\_part3 <= upper\_bound\_part3) {  + cat("Yes, E(X) in Part 3 (Best of Five) is within the 95% confidence interval.")  + } else {  + cat("No, E(X) in Part 3 (Best of Five) is not within the 95% confidence interval.")  + }  Yes, the E(X) in Part 3 (Best of Five) is within the 95% confidence interval. |

**iv) Chi-square Goodness of Fit Test (Best of Five):**

* The chi-square goodness of fit test was employed to assess how well the simulated distribution matched the expected distribution. *The test resulted in rejecting the null hypothesis*, indicating that the simulated data did not precisely match the expected distribution for this best-of-five scenario.

**O/P:**

|  |
| --- |
| > chisq\_test\_part3  Chi-squared test for given probabilities  data: simulated\_data\_counts\_part3  X-squared = 1606.3, df = 1, p-value < 0.00000000000000022  > if (chisq\_test\_part3$p.value > 0.05) {  + cat("We fail to reject the null hypothesis that the simulated data (Y) in Part 3 (Best of Five) follows the expected distribution (X).")  + } else {  + cat("We reject the null hypothesis that the simulated data (Y) in Part 3 (Best of Five) follows the expected distribution (X).")  + }  We reject the null hypothesis that the simulated data (Y) in Part 3 (Best of Five) follows the expected distribution (X). |

**v) Betting Strategy Analysis Summary:**

* Comparing observed and expected probabilities, the analysis revealed a significant discrepancy. While the expected probability of winning was 20.88%, the observed probability based on simulations was 78.34%. This further emphasizes the unpredictability and potential risks associated with the betting strategy in the best-of-five series.

**O/P:**

**#Note: *Please excuse the output mentioned below, which is in picture format, just to maintain the indentation.***

|  |
| --- |
|  |

**Part - 4:**

In this final part of the analysis, we explore a scenario where both the Boston Celtics and Golden State Warriors reach the NBA Finals. The series is a best-of-seven format, with the first team to win four games securing victory. The unique 2-3-2 format designates the team with home-court advantage to host games 1, 2, 6, and 7, while the opponent hosts games 3, 4, and 5. The Celtics have the home-court advantage in this scenario.

**i) Probability of Winning the NBA Finals (Best of Seven):**

* The probability that the Celtics would win the NBA Finals was found to be approximately 45.6%, taking into account various scenarios in the best-of-seven series.

**O/P:**

|  |
| --- |
| > cat("Probability that Celtics win the series in Part 4:", celtics\_win\_series\_part4, "\n")  Probability that Celtics win the series in Part 4: 0.456 |

**ii) Net Win Distribution (Best of Seven):**

* The expected net win for the best-of-seven series was calculated to be $259.2, with a standard deviation of $212.85. This suggests that, on average, a bettor could expect to win $259.2 per bet under this NBA Finals scenario.

**O/P:**

|  |
| --- |
| > cat("Expected net win in Part 4 (Best of Seven):", expected\_win\_part4, "\n")  Expected net win in Part 4 (Best of Seven): 259.2  > cat("Standard deviation of net win in Part 4 (Best of Seven):", standard\_deviation\_part4, "\n")  Standard deviation of net win in Part 4 (Best of Seven): 212.8542 |

**iii) Simulation and Confidence Interval (Best of Seven):**

* Simulating 5,000 random scenarios, the estimated expected net win was $254.73. The 95% confidence interval for the simulation ranged from approximately -$306.01 to $815.47. *Interestingly, this interval still contained the expected net win of $259.2, indicating considerable variability in the outcomes.*

**O/P:**

|  |
| --- |
| > cat("Estimated expected net win in Part 4 (Best of Seven):", estimated\_expected\_win\_part4, "\n")  Estimated expected net win in Part 4 (Best of Seven): 254.73  > cat("95% confidence interval in Part 4 (Best of Seven):", lower\_bound\_part4, "-", upper\_bound\_part4, "\n")  95% confidence interval in Part 4 (Best of Seven): -306.0098 - 815.4698  > # Check if E(X) is within the confidence interval for Part 4  > if (expected\_win\_part4 >= lower\_bound\_part4 & expected\_win\_part4 <= upper\_bound\_part4) {  + cat("Yes, E(X) in Part 4 (Best of Seven) is within the 95% confidence interval.")  + } else {  + cat("No, E(X) in Part 4 (Best of Seven) is not within the 95% confidence interval.")  + }  Yes, the E(X) in Part 4 (Best of Seven) is within the 95% confidence interval. |

**iv) Chi-square Goodness of Fit Test (Best of Seven):**

* The chi-square goodness of fit test was employed to assess the match between the simulated and expected distributions. *The test resulted in rejecting the null hypothesis*, implying that the simulated data did not precisely match the expected distribution for this NBA Finals scenario.

**O/P:**

|  |
| --- |
| > cat("Chi-square test statistic in Part 4 (Best of Seven):", chisq\_test\_part4$statistic, "\n")  Chi-square test statistic in Part 4 (Best of Seven): 55.7568  > cat("p-value in Part 4 (Best of Seven):", chisq\_test\_part4$p.value, "\n")  p-value in Part 4 (Best of Seven): 0.0000000000000820141  > # Interpreting the results for Part 4  > if (chisq\_test\_part4$p.value > 0.05) {  + cat("We fail to reject the null hypothesis that the simulated data (Y) in Part 4 (Best of Seven) follows the expected distribution (X).")  + } else {  + cat("We reject the null hypothesis that the simulated data (Y) in Part 4 (Best of Seven) follows the expected distribution (X).")  + }  We reject the null hypothesis that the simulated data (Y) in Part 4 (Best of Seven) follows the expected distribution (X). |

**v) Comparison of Observed and Expected Probabilities:**

* Comparing observed and expected probabilities, the analysis revealed a significant discrepancy. While the expected probability of winning was 36%, the observed probability based on simulations was 35.66%. The observed probabilities further underscore the unpredictability of NBA Finals outcomes and highlight potential risks associated with the betting strategy.

**O/P:**

**#Note: *Please excuse the output mentioned below, which is in picture format, just to maintain the indentation.***

|  |
| --- |
|  |

|  |
| --- |
| > # Interpreting the final results for Part 4  > if (chisq\_test\_part4$p.value > 0.05) {  + cat("The betting strategy appears to be favorable.\n")  + } else {  + cat("The betting strategy may not be favorable. Further analysis is recommended.\n")  + }  The betting strategy may not be favorable. Further analysis is recommended. |

1. **Conclusion:**

In conclusion, the initial analysis of the betting strategy for the Boston Celtics and Golden State Warriors NBA series raised concerns about its consistency, as simulations and chi-square tests indicated potential deviations from expected outcomes. Adjustments in subsequent scenarios led to changes in probabilities and expected net wins, emphasizing the importance of adapting the strategy to different game sequences. The shift to a best-of-five series revealed significant variability, necessitating further refinements to enhance reliability. While the NBA Finals betting strategy showed promise, the simulations indicated substantial variability and potential risks, highlighting the importance of careful consideration and adjustments to optimize the strategy's effectiveness across diverse scenarios. Overall, this comprehensive analysis underscores the need for ongoing refinement and adaptation of the betting strategy to ensure its reliability in varying playoff formats.

1. **Citations:**

* Probability calculations: [source](https://youtu.be/cx5EZXJIDjo?si=uceL5CP1UjUty2b0).
* Simulation methodology: [source](https://youtu.be/HRC-POXak0M?si=nGd2TxpdA78-HbmD).
* Chi-square goodness of fit testing: [source](https://youtu.be/n5c11B5FJ24?si=yhRyn99R0xaJIKJj).