Logo

Description automatically generated

**ALY 6050: Enterprise Analytics**

**Project: Analysis of Betting Strategy in Sports**

**Submitted to: Prof. Roy, Wada**

**Submitted by: Abhinav Jain**

**Date: 05/29/2022**

**INTRODUCTION**

**Background:**

In this report, the software used to analyze the outcome of a betting strategy in Sports was Excel, which will help in getting the outcome of the probability of the teams winning and losing on their home or away grounds.

In this analysis, two American league baseball teams are Red Sox which are from Boston, whereas the second team Yankees which are from New York. We have divided this report into three parts:

1. Find the probability and other factors if the first match played in Boston by Red Sox teams in 3 match series?
2. Find the Probability and other factors if the first match played in New York by Red Sox in 3 match series?
3. Find the probability and other factors for the 5-match series in which the first match is played in Boston with the best 3-match winning?

While doing the analysis we have calculated the probability of the Red Sox winning in the series, expected net winning which means the betting amount, and moreover the standard deviation of the betting. In this problem, the data is given that if the probability of the Red Sox winning in their home stadium is given 0.59, on the other hand, the Yankees winning in the home stadium is 0.55.

However, the betting amount for the game played in the Red Sox win will be $505 and Red Sox lose the game was $525 which we have taken while finding the probability of the Red Sox in the home stadium or Yankees in the home stadium.

During the analysis used the chi-Square goodness of fit test to check the close distribution for Y by using random values of the confidence interval @95% to estimate the distribution of X. To understand the favorable condition of the betting to predict the outcome of the match.

**ANALYSIS AND INTERPRETATION**

**Part – I**

1: Find the probability of whether the Red Sox wins the series if the first match is played in Boston between the Red Sox and Yankees within 3 match series?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Baseball Match Between Red Sox (Boston) and Yankee(New York) #3 Match Series Two American Baseball League (PART I-First Match in Boston) | | | | |
|  |
|  |
|  |
| **Match Schedule** | **Boston** | | **New York** | |  |
| **Teams** | **Win** | **Lose** | **Win** | **Lose** |  |
|  | | | | |  |
| **Red Sox** | **0.59** | **0.41** | **0.35** | **0.65** |  |
| **Yankee** | **0.4** | **0.6** | **0.55** | **0.45** |  |
|  |  |  |  |  |  |

**Chart 1: Probability of the match**

**Chart 2: Betting Amount Analysis**

Table: 1 Chance of winning in the home stadium.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
| **Probability of Red Sox Chances of Winning** | | | | |
| **Team Name** | **1st Game Boston** | **2nd Game New York** | **3rd Game Boston** | **Probability** |
| Red Sox | 0.59 | 0.35 |  | 0.2065 |
| Red Sox | 0.59 | 0.65 | 0.59 | 0.226265 |
| Red Sox | 0.41 | 0.35 | 0.41 | 0.058835 |
| Red Sox | 0.41 | 0.35 | 0.59 | 0.084665 |
| Red Sox | 0.59 | 0.65 | 0.41 | 0.157235 |
| Red Sox | 0.41 | 0.65 |  | 0.2665 |
|  |  |  |  | **1** |
|  |  | Prob of Red Sox win/lose Series | Prob of Winning | 0.51743 |
|  |  | Prob of Losing | 0.48257 |

**Analysis:**

The probability of winning the match by the Red Sox is 51%, whereas the probability of losing is 48% in the home stadium. In the below graph we can clearly see the below bar graph left bar is showing the probability of winning the Red Sox team in the home stadium, on the other hand, the left bar shows the low probability of losing the match.

**Chart 3: Probability of RED SOX winning in Home Stadium**

2: Findings on Probability Distribution for the net win in the series.

The average winning of the Red Sox is $-99 and the variance of the expected winning of the betting amount is 631982.601. whereas the standard deviation of the outcome is 794.97

**Table 2: Expected Value of Winning Team**

|  |  |
| --- | --- |
| Values | Expected Winning of X |
| **-99.00** | Avg Winning of Red fox |
| **631982.60** | Variance |
| **794.97** | Standard Deviation |

**Table 3: Observed Value of Winning Team**

|  |  |  |  |
| --- | --- | --- | --- |
| **Observed ValueObserved Value** | | | |
| Average | | | -38.29099307 |
| Variance | | | 641313.615 |
| Standard Deviation | | | 800.8205885 |
| Confidence interval | | | 15.69579512 |
| CI Lower Limit | | | -53.98678819 |
| CI Upper Limit | | | -22.59519796 |
| **Table 4: Chi-Square Test** | | |  |
| Expected outcomes vs Observed Outcomes | | | |
| **E(X)** | **O(X)** | **Chi-Square = ((O-E) ^2)/E** | |
| 2065 | 2093 | 0.38 | |
| 3109.3 | 3051 | 1.09 | |
| 2665 | 2650 | 0.08 | |
| 2160.7 | 2207 | 0.99 | |
|  | **CHI-SQUARE** | **2.55** | |
|  | **P-value** | **0.18** | |

The Expected value for the chi-square test score we have found is 2.55, whereas the p-value is 0.18 of the chi-squares is dependent on the random value of the data form for the 10k observation of the data during the analysis. In this hypothesis testing is significant which shows the high correlation between the observed value and the expected values. Therefore, the random population of the expected outcome of the good fit.

**Chart 4: Expected and Observed Outcome**

**PART II**

Find the probability of whether the Red Sox wins the series if the first match is played in New York between the Red Sox and Yankees within 3 match series?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Baseball Match Between Red Sox (Boston) and Yankee(New York) #3 Match Series Two American Baseball League (PART II-First Match in New York) | | | | |
|  |
|  |
|  |
| **Match Schedule** | **New York** | | **Boston** | |  |
| **Teams** | **Win** | **Lose** | **Win** | **Lose** |  |
|  | | | | |  |
| **Red Sox** | **0.35** | **0.65** | **0.59** | **0.41** |  |
| **Yankee** | **0.55** | **0.45** | **0.4** | **0.6** |  |

**Chart 5: Probability of Win in Home and Away**

**Chart 6: Betting Amount Analysis**

**Analysis:**

**Table 5: Probability of Red Sox Chances of Winning**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Probability of Red Sox Chances of Winning** | | | | |
| **Team Name** | **1st Game Boston** | **2nd Game New York** | **3rd Game Boston** | **Probability** |
| Red Sox | 0.35 | 0.59 |  | 0.2065 |
| Red Sox | 0.35 | 0.41 | 0.35 | 0.050225 |
| Red Sox | 0.65 | 0.59 | 0.65 | 0.249275 |
| Red Sox | 0.65 | 0.59 | 0.35 | 0.134225 |
| Red Sox | 0.35 | 0.41 | 0.65 | 0.093275 |
| Red Sox | 0.65 | 0.41 |  | 0.2665 |
|  |  |  |  | **1** |
|  |  | Prob of Red Sox win/lose Series | Prob of Winning | 0.39095 |
|  |  | Prob of Losing | 0.60905 |

The probability of winning the match by the Red Sox is 40%, whereas the probability of losing is 60% in the away stadium. In the below graph we can clearly see the below bar graph left bar is showing the probability of winning the Red Sox team in the away stadium, on the other hand, the left bar shows the high probability of losing the match.

**Chart 7: Probability of RED SOX winning in Away Stadium**

**Table 6: Probability Distribution**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
| Probability Distribution of Winning Red Fox | | | | |
| **Probability Range** | Betting Amount (X) | Prob Red Sox P(X) | X^2 | Cumulative P(X) |
| 0 | $ 1,010.00 | 0.21 | $ 1,020,100.00 | 0.2065 |
| 0.2065 | $ 485.00 | 0.18 | $ 235,225.00 | 0.39095 |
| 0.39095 | $ (1,050.00) | 0.27 | $ 1,102,500.00 | 0.65745 |
| 0.65745 | $ (545.00) | 0.34 | $ 297,025.00 | 1 |

**Table 7: Expected Values in Away Team**

|  |  |
| --- | --- |
| Values | Expected Winning of X |
| **-99** | Avg Winning of Red fox |
| **639799.065** | Variance |
| **799.8744058** | Standard Deviation |

**Table 8: Observed Values**

|  |  |
| --- | --- |
| **Observed Value** | |
| Average | -147.7678482 |
| Variance | 623468.034 |
| Standard Deviation | 789.5999202 |
| Confidence interval | 15.47587406 |
| CI Lower Limit | -163.2437222 |
| CI Upper Limit | -132.2919741 |

**Table 9: Chi-Square test**

|  |  |  |
| --- | --- | --- |
|  |  |  |
| **Expected outcomes vs Observed Outcomes** | | |
| **E(X)** | **O(X)** | **Chi-Square = ((O-E)^2)/E** |
| 2065 | 2111 | 1.02 |
| 1844.5 | 1933 | 4.25 |
| 2665 | 2577 | 2.91 |
| 3425.5 | 3381 | 0.58 |
|  | **CHI-SQUARE** | **8.75** |
|  | **P-value** | **0.01** |

**Chart 8: Expected and Observed Outcome**

The expected value for the chi-square test score we have found is 8.75, whereas the p-value is 0.01 of the chi-squares is dependent on the random value of the data form for the 10k observation of the data during the analysis. In this hypothesis testing is significant which shows the high correlation between the observed value and the expected values. Therefore, the random population of the expected outcome of the good fit.

**PART III**

Find the probability and other factors for the 5-match series in which the first match is played in Boston with the best 3-match winning?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Baseball Match Between Red Sox (Boston) and Yankee (New York) # 5 Match Series Two American Baseball League (PART III) | | | | |
|  |
|  |
|  |
| **Match Schedule** | **Boston** | | **New York** | |  |
| **Teams** | **Win** | **Lose** | **Win** | **Lose** |  |
|  | | | | |  |
| **Red Sox** | **0.59** | **0.41** | **0.35** | **0.65** |  |
| **Yankee** | **0.4** | **0.6** | **0.55** | **0.45** |  |

**Chart 9: Probability of win in home and away**

**Analysis:**

**Table 10: Probability of Red Sox Chances of Winning**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| **Probability of Red Sox Chances of Winning** | | | | | | |
| **Team Name** | **1st Game Boston** | **2nd Game New york** | **3rd Game Boston** | **4th Game New York** | **5th Game Boston** | **Probability** |
| Red Sox | 0.59 | 0.35 | 0.59 |  |  | 0.12 |
| Red Sox | 0.59 | 0.35 | 0.41 | 0.35 |  | 0.03 |
| Red Sox | 0.59 | 0.65 | 0.59 | 0.35 |  | 0.08 |
| Red Sox | 0.41 | 0.35 | 0.59 | 0.35 |  | 0.03 |
| Red Sox | 0.59 | 0.65 | 0.59 | 0.65 | 0.59 | 0.09 |
| Red Sox | 0.59 | 0.65 | 0.41 | 0.35 | 0.59 | 0.03 |
| Red Sox | 0.41 | 0.65 | 0.59 | 0.35 | 0.59 | 0.03 |
| Red Sox | 0.41 | 0.35 | 0.41 | 0.35 | 0.59 | 0.01 |
| Red Sox | 0.59 | 0.35 | 0.41 | 0.65 | 0.59 | 0.03 |
| Red Sox | 0.41 | 0.35 | 0.59 | 0.65 | 0.59 | 0.03 |
| Red Sox | 0.41 | 0.65 | 0.41 |  |  | 0.11 |
| Red Sox | 0.41 | 0.65 | 0.59 | 0.65 |  | 0.10 |
| Red Sox | 0.41 | 0.35 | 0.41 | 0.65 |  | 0.04 |
| Red Sox | 0.59 | 0.65 | 0.41 | 0.65 |  | 0.10 |
| Red Sox | 0.41 | 0.35 | 0.41 | 0.35 | 0.41 | 0.01 |
| Red Sox | 0.41 | 0.35 | 0.59 | 0.65 | 0.41 | 0.02 |
| Red Sox | 0.59 | 0.35 | 0.41 | 0.65 | 0.41 | 0.02 |
| Red Sox | 0.59 | 0.65 | 0.59 | 0.65 | 0.41 | 0.06 |
| Red Sox | 0.41 | 0.65 | 0.59 | 0.35 | 0.41 | 0.02 |
| Red Sox | 0.59 | 0.65 | 0.41 | 0.35 | 0.41 | 0.02 |
|  |  |  |  |  |  | **1** |
|  |  |  |  | **Prob of Red Sox win/lose Series** | Prob of Winning | 0.49 |
|  |  |  |  | Prob of Losing | 0.51 |

The probability of winning the match by the Red Sox is 48%, whereas the probability of losing is 52% in the away stadium. In the above graph we can clearly see the below bar graph left bar is showing the probability of winning the Red Sox team is low in the away stadium, on the other hand, the left bar shows the high probability of losing the match.

**Table 11: Expected Value**

|  |  |
| --- | --- |
| Values | Expected Winning of X |
| **$ (239.00)** | Avg Winning of Red fox |
| **$ 1,007,301.31** | Variance |
| **$ 1,003.64** | Standard Deviation |

**Table 12: Observed Value**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Observed Value** | | |  | | |
| Average | | | $ (375.69) | | |
| Variance | | | $ 1,589,906.95 | | |
| Standard Deviation | | | $ 1,260.92 | | |
| Confidence interval | | | $ 24.71 | | |
| CI Lower Limit | | | $ (400.40) | | |
| CI Upper Limit | | | $ (350.97) | | |
| **Table 13: Probability Distribution of Winning Red Fox** | | |  | | |
|  |  |  | |  |  |
| **Probability Distribution of Winning Red Fox** | | | | | |
| **Probability Range** | Betting Amount (X) | Prob Red Sox P(X) | | X^2 | Cumulative P(X) |
| 0.00 | $ 1,515.00 | 0.12 | | $ 2,295,225.00 | 0.12 |
| 0.12 | $ 990.00 | 0.14 | | $ 980,100.00 | 0.26 |
| 0.26 | $ 465.00 | 0.23 | | $ 216,225.00 | 0.49 |
| 0.49 | $ (1,575.00) | 0.11 | | $ 2,480,625.00 | 0.60 |
| 0.60 | $ (1,070.00) | 0.24 | | $ 1,144,900.00 | 0.84 |
| 0.84 | $ (565.00) | 0.16 | | $ 319,225.00 | 1.00 |

**Table 14: Chi-Square test**

|  |  |  |
| --- | --- | --- |
|  |  |  |
| Expected outcomes vs Observed Outcomes | | |
| **E(X)** | **O(X)** | **Chi-Square = ((O-E)^2)/E** |
| 1218.35 | 1215 | 0.01 |
| 1384.58 | 1413 | 0.58 |
| 2287.98 | 2186 | 4.55 |
| 1092.65 | 5185 | 15327.26 |
| 2426.48 | 0 | 2426.48 |
| 1589.95 | 0 | 1589.95 |
|  | **CHI-SQUARE** | **19348.83** |
|  | **P-value** | **0.00000** |

The expected value for the chi-square test score we have found is 19348.83, whereas the p-value is 0.00 of the chi-squares is dependent on the random value of the data form for the 10k observation of the data during the analysis. In this hypothesis testing is significant which shows the high correlation between the observed value and the expected values. Therefore, the random population of the expected outcome of the good fit.

**Chart 10: Expected and Observed Outcome**

**Conclusion:** The result of this report's analysis is that the team's victory depends on the stadium's condition and the crowd's condition, as well as the practice. The team that trains on its home ground has a good chance of beating the rival in a game. While performing the analysis in Part I, the results show that the probability of winning is high because the first game was played in the home stadium, whereas the observed and expected values give a significant result rejecting the null hypothesis because the results are not significant, even though the results are dependent on the random values of the testing random test's outcome. Furthermore, when the first match is held in the away team's stadium, the winning probability decreases; yet, when the chi-square test is performed, such significant levels of observed values suggest that the null hypothesis is accepted. In Part III, we may estimate the likelihood will be lower, and the winner's result will be determined by five matches, with the significance level decreasing to zero to determine whether the team will win or lose the game.

**References:**

**[1] Interpreting Chi Square Results in SPSS**

By

[**https://ezspss.com/interpreting-chi-square-results-in-spss/#:~:text=Put%20simply%2C%20the%20more%20these,are%20associated%20with%20each%20other.**](https://ezspss.com/interpreting-chi-square-results-in-spss/#:~:text=Put%20simply%2C%20the%20more%20these,are%20associated%20with%20each%20other.)

[2] **Chi Square P Value Excel: Easy Steps, Video**

[**https://www.statisticshowto.com/probability-and-statistics/excel-statistics/calculate-chi-square-p-value-excel/**](https://www.statisticshowto.com/probability-and-statistics/excel-statistics/calculate-chi-square-p-value-excel/)

[3] **Chi-Square Test in Excel**

Article byWallstreetmojo Editorial TeamReviewed byDheeraj Vaidya et al.

[**https://www.wallstreetmojo.com/chi-square-test-in-excel/**](https://www.wallstreetmojo.com/chi-square-test-in-excel/)