Module 1 Project — Analysis of a Betting Strategy in Sports

Sourabh D. Khot (ID 002754952)

College of Professional Studies, Northeastern University

ALY 6050: Introduction to Enterprise Analytics CRN 81180

Professor Azadeh Mobasher

May 29, 2022

Table of Contents

Problem Introduction

As a baseball enthusiast, I follow the American League game series. There are upcoming series of games to be played between Boston Red Sox and New York Yankee and I am interested in betting money on my favourite team, Boston Red Sox team so I can make good money.

The games will be played alternatively in each team's home stadium. There will be three series of games played as follows:

1. Best-of-three played in order in Boston, New York, and if necessary, Boston.
2. Best-of-three played in order in New York, Boston, and if necessary, New York
3. Best-of-five, played in order in Boston, New York, Boston and if necessary, New York, Boston

Through historical data given, I know the probabilities of Red Sox and Yankees each winning a game against the other in their home stadium. Also, the payoff per game is announced. Through theoretical probability analysis and simulation through random numbers, I have to determine if my strategy of betting on the Boston Red Sox is favorable to me in each of the three series.

Problem Analysis

The probability of a team winning against the other in each stadium is given in Table 1. My payoff for betting on Team Red Sox for each game is given in Table 2.

Team	Stadium	
	Boston	NY
Boston Red Sox	0.60	0.43
NY Yankees	0.40	0.57
Total	1.00	1.00

Table 1. Probability of winning one game

Game Outcome	Amount
Red Sox wins (W)	500
Red Sox loses (L)	-520

Table 2. Pay off per game

If any team wins both the first and second games, the third game will not be played.

Part 1: Best-of-Three in the order Boston-NY-(Boston)

(i) Calculate the probability that the Red Sox will win the series.

Favourable Outcomes

Event, a	**Prob, P(a)**
WW-	0.26
WLW	0.21
LWW	0.10
Fav. Prob.	**0.57**

Table 3. Probability of Red Sox Winning

The favorable outcomes are when the team Red Sox wins two games. The favorable outcomes from the sample space are given in Table 3. Considering the games are independent of each other, the probability that the Red Sox will win the series is 0.566.

(ii) Construct a probability distribution for your net win (X) in the series. Calculate your expected net win (the mean of X) and the standard deviation of X.

Net Win, x	Prob, P(x)
-1040	0.23
-540	0.21
480	0.31
1000	0.26
Total	1.00

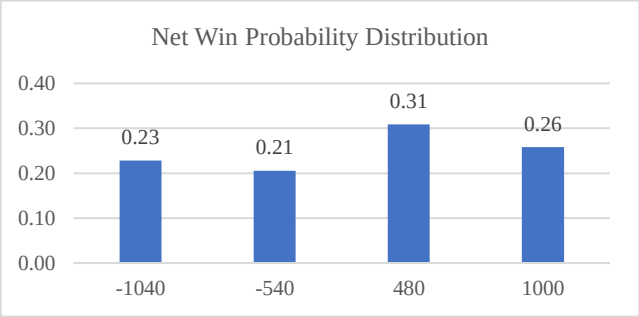

Table & Figure 4. Probability Distribution of Net Win

The net win is computed and shown in Table 4 and Figure 4. The expected net win is 57.89, and its standard deviation is 795.15.

(iii) Use Excel to create 10,000 random values for X. Let these random values be denoted by Y. Use these Y values to estimate your expected net win by using a 95% confidence interval. Does this confidence interval contain E(X)?

The net win is estimated using random numbers and for one such simulation, the observed mean is 52.86 and observed standard deviation is 796.71. Since the simulation is a sample from a larger population, the sampling error exhibits normal distribution, and using the previous two values, the 95% confidence interval comes as 37.24 to 68.48, which contains the theoretically expected value of 57.89.

(iv) Construct a frequency distribution for Y. Next, use the Chi-squared goodness of fit test to verify how closely the distribution of Y has estimated the distribution of X.

Net Win	Prob	Cum Prob	Theoretical	Observed	Chi-Sq
x	P(x)	P(X ≤ x)	Freq Dist	Freq Dist Y	
-1040	0.23	0.23	2280	2310	0.39
-540	0.21	0.43	2056	2056	0.00
480	0.31	0.74	3084	3063	0.14
1000	0.26	1.00	2580	2571	0.03
Total	1.00		10000	10000	

Table 5. Frequency Distribution of Theoretical and Observed Values

The hypothesis is defined here as:

H0: Dist. Y is consistent with dist. X

H1: Dist. Y is not consistent with dist. X

The sum of chi-square metric is 0.57 and degrees of freedom are 3. The p-value is 0.90 (much greater than alpha of 0.05); hence, there is not sufficient evidence to reject H0. Hence, dist. Y closely estimates dist. X

(v) Use your observations of parts (ii) and (iii) above to describe whether your betting strategy is favorable to you.

Expected Value of Net Win

Theoretical Expected Value	57.89
Simulated Expected Value	52.86

Risk of Net Win

Theoretical SD	795.15
Simulated SD	796.71

Table 6: Expected value and Risks

Looking at the expected values, the betting strategy is favorable since the expected values are a good positive number. That means playing this over huge number of iterations, I will earn around $57.89. However, the risk is too high (SD is around 14 times of the theoretical mean).

Part 2: Best-of-Three in the order NY-Boston-(NY)

(i) Calculate the probability that the Red Sox will win the series.

Favourable Outcomes

Event, a	Prob, P(a)
WW-	0.26
WLW	0.07
LWW	0.15
Fav. Prob.	**0.48**

Table 7. Probability of Red Sox Winning

The favorable outcomes from the sample space are given in Table 7. Considering the games are independent of each other, the probability that the Red Sox will win the series is 0.48.

(ii) Construct a probability distribution for your net win (X) in the series. Calculate your expected net win (the mean of X) and the standard deviation of X.

Net Win, x	Prob, P(x)
-1040	0.23
-540	0.29
480	0.22
1000	0.26
Total	1.00

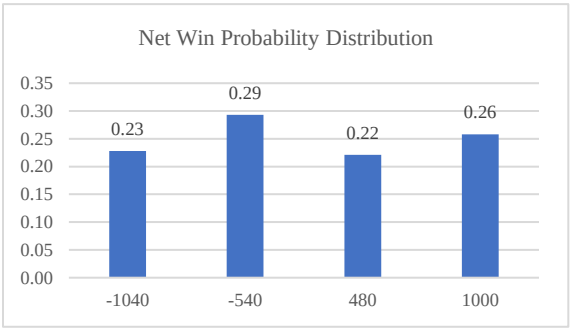

Table & Figure 8. Prob. Dist. of Net Win

The net win is computed and shown in Table 8 and Figure 8. The expected net win is negative 31.24 and its standard deviation is 799.99.

(iii) Use Excel to create 10,000 random values for X. Let these random values be denoted by Y. Use these Y values to estimate your expected net win by using a 95% confidence interval. Does this confidence interval contain E(X)?

The net win is estimated using random numbers and for one such simulation, the observed mean is -36.47 and observed standard deviation is 799.68. Since the simulation is a sample from a larger population, the sampling error exhibits normal distribution, and using the previous two values, the 95% confidence interval comes as -52.14 to -20.79, which contains the theoretically expected value of -31.24.

(iv) Construct a frequency distribution for Y. Next, use the Chi-squared goodness of fit test to verify how closely the distribution of Y has estimated the distribution of X.

Net Win	Prob	Cum Prob	Theoretical	Observed	Chi-Sq
x	P(x)	P(X ≤ x)	Freq Dist	Freq Dist Y	
-1040	0.23	0.23	2280	2329	1.05
-540	0.29	0.52	2930	2881	0.81
480	0.22	0.74	2210	2263	1.26
1000	0.26	1.00	2580	2527	1.09
Total	1.00		10000	10000	

Table 9. Frequency Distribution of Theoretical and Observed Values

The hypothesis is defined here as:

H0: Dist. Y is consistent with dist. X

H1: Dist. Y is not consistent with dist. X

The sum of chi-square metric is 4.22 and degrees of freedom are 3. The p-value is 0.24 (greater than alpha of 0.05); hence, there is not sufficient evidence to reject H0. Hence, dist. Y closely estimates dist. X

(v) Use your observations of parts (ii) and (iii) above to describe whether your betting strategy is favorable to you.

Expected Value of Net Win

Theoretical Expected Value	-31.24
Simulated Expected Value	-36.47

Risk of Net Win

Theoretical SD	799.99
Simulated SD	799.68

Table 10: Expected value and Risks

The expected values are negative, both theoretically and in almost all the simulations. That means playing this series large number of times, I will actually lose around $31.24. Hence, the betting strategy is clearly not favorable. The standard deviation is also high (around 25 times of the theoretical mean), which can go in either direction of the expected value, but the risk is not worth it.

Part 3: Best-of-Five in the order Boston-NY-Boston-(NY-Boston)

(i) Calculate the probability that the Red Sox will win the series.

Favourable Outcomes

Event, a	**Prob, P(a)**
WWW--	0.15
WWLW-	0.04
WLWW-	0.09
LWWW-	0.04
WWLLW	0.04
WLWLW	0.07
LWWLW	0.04
WLLWW	0.04
LWLWW	0.02
LLWWW	0.04
Fav. Prob.	**0.56**

Table 11. Probability of Red Sox Winning

The favorable outcomes from the sample space are given in Table 11. The entire sample space is available in the submitted workbook. Considering the games are independent of each other, the probability that the Red Sox will win the series is 0.56.

(ii) Construct a probability distribution for your net win (X) in the series. Calculate your expected net win (the mean of X) and the standard deviation of X.

Net Win, x	Prob, P(x)
-1560	0.09
-1060	0.20
-560	0.15
460	0.23
980	0.18
1500	0.15
Total	1.00

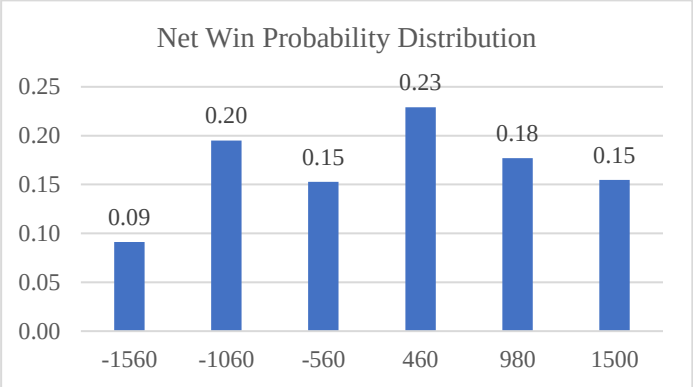

Table & Figure 12. Prob. Dist. of Net Win

The net win is computed and shown in Table 8 and Figure 8. The expected net win is 76.35 and its standard deviation is 1024.73.

(iii) Use Excel to create 10,000 random values for X. Let these random values be denoted by Y. Use these Y values to estimate your expected net win by using a 95% confidence interval. Does this confidence interval contain E(X)?

The net win is estimated using random numbers, and for one such simulation, the observed mean is 72.47 and observed standard deviation is 1024.59. Since the simulation is a sample from a larger population, the sampling error exhibits normal distribution, and using the

previous two values, the 95% confidence interval comes as 52.39 to 92.55, which contains the theoretically expected value of 76.35.

(iv) Construct a frequency distribution for Y. Next, use the Chi-squared goodness of fit test to verify how closely the distribution of Y has estimated the distribution of X.

Net Win	**Prob**	**Cum Prob**	**Theoretical**	**Observed**	**Chi-Sq**
x	**P(x)**	**P(X ≤ x)**	**Freq Dist**	**Freq Dist Y**	
-1560	0.09	0.09	912	909	0.01
-1060	0.20	0.29	1952	1968	0.14
-560	0.15	0.44	1527	1540	0.10
460	0.23	0.67	2291	2263	0.34
980	0.18	0.85	1770	1788	0.19
1500	0.15	1.00	1548	1532	0.17
Total	1.00		10000	10000	

Table 13. Frequency Distribution of Theoretical and Observed Values

The hypothesis is defined here as:

H0: Dist. Y is consistent with dist. X

H1: Dist. Y is not consistent with dist. X

The sum of chi-square metric is 0.95 and degrees of freedom are 5. The p-value is 0.97 (much greater than alpha of 0.05); hence, there is not sufficient evidence to reject H0. Hence, dist. Y closely estimates dist. X.

(v) Use your observations of parts (ii) and (iii) above to describe whether your betting strategy is favorable to you.

Expected Value of Net Win

Theoretical Expected Value	76.35
Simulated Expected Value	72.47

Risk of Net Win

Theoretical SD	799.99
Simulated SD	799.68

Table 6: Expected value and Risks

The expected values (both theoretical and simulated) are good positive numbers (around 76), which are greater than those in part 1 (around 57). That means if I play this series for a very long time, I will earn around $76.35. By expected value, this is a favourable case of betting.

However, there is similar amount of risk as in part 1, since the standard deviation is 14 times of the expected value.

Conclusion

Considering probability of individual games, the probability distribution of net win was constructed, the outcomes were simulated 10,000 times, and the expected values and risk were calculated in Mirosoft Excel. The conclusions are summarized below.

In case of part 1 series of games, by analyzing the theoretical and simulated mean, the betting strategy was favorable. But the degree of risk is high, which depends on each individual if they are willing to take the risk. I personally am risk-neutral, but may not bet in this game as the standard deviation is 14 times of the expected value.

In the case of part 2 series of games, the betting strategy was clearly not favorable. Any person must avoid betting on Red Sox team in this series.

In the part 3 series of games, which was best-of-three in five games, the betting on Red Sox team was favourable. The risk was similar to part 1, however, the expected value was higher than part 1 at $76.35. Over multiple simulation iterations, the simulated expected value was also positive. Risk-taking individuals may bet in this game.

References

Albright, S. (2016). *Business Analytics. Sixth Edition.* Boston, MA: Cengage Learning.

APA Style Table: APA.org. (n.d.). Retrieved from https://apastyle.apa.org/style-grammar-guidelines/tables-figures/tables

Canvas Module 1 Project: Analysis of a Betting Strategy in Sports. (n.d.). Retrieved from https://northeastern.instructure.com/courses/110059/assignments/1346025

Chi-Square Goodness of Fit Test: yale.edu. (n.d.). Retrieved from http://www.stat.yale.edu/Courses/1997-98/101/chigf.htm#:~:text=%3D%20(r%20%2D%201)(c,square%20statistic%20may%20be%20calculated.

Concept of Prob: TheFactFactor.com. (n.d.). Retrieved from https://thefactfactor.com/facts/pure_science/mathematics/statistics-and-probability/probability/15095/#:~:text=The%20result%20of%20a%20random,is%20denoted%20by%20%27S%27.

CONFIDENCE Function: CorporateFinanceInstitute.com. (n.d.). Retrieved from https://corporatefinanceinstitute.com/resources/excel/functions/confidence-interval-excel-function/

Evans, J. R. (2013). *Statistics, data analysis, and decision modeling.* New York: Pearson.

Payoff Table: Kaplan Financial Knowledge Bank. (n.d.). Retrieved from https://kfknowledgebank.kaplan.co.uk/payoff-tables-

Playoff Format: Wikipedia. (n.d.). Retrieved from https://en.wikipedia.org/wiki/Playoff_format#Best-of-three_playoff

<End of Report>