Syed Faizan

Analysis of a Betting Strategy in Sports

Analysis of a Betting Strategy in Sports: Red Sox vs. Yankees

Introduction

This report analyzes the betting strategy involved in a best-of-three series between the Boston Red Sox and the New York Yankees. The primary objective is to evaluate the probability of winning the series, the expected net win, and the effectiveness of the betting strategy. We will also examine the simulation results and perform a goodness-of-fit test to validate the theoretical distribution.

Part 1: Best-of-Three Series (First Game in Boston, second in New York, Third in Boston if Needed)

(i) Probability of the Red Sox Winning the Series

The probabilities for each game outcome, given the probabilities of winning at home for each team (0.6 for Boston and 0.57 for New York, we use the complement rule to find probability of a Boston loss at home 0.4 and win at New York 0.43), are calculated using the basic multiplication rule of mutually exclusive event probabilities as follows:

- WW (Boston wins both games): P(WW) = 0.6 × 0.43 = 0.258
- WLW (Boston wins, loses, then wins): P(WLW) = 0.6 × 0.57 × 0.6 = 0.2052
- LWW (Boston loses, wins, then wins): P(LWW) = 0.4 × 0.43 × 0.6 = 0.1032
- WLL (Boston wins, then loses both): P(WLL) = 0.6 × 0.57 × 0.4 = 0.1368
- LWL (Boston loses, wins, then loses): P(LWL) = 0.4 × 0.43 × 0.4 = 0.0688
- LL (Boston loses both games): P(LL) = 0.4 × 0.57 = 0.228

The probability that Boston wins the series is the sum of the probabilities of all outcomes in which Boston wins:

P(Boston wins) = P(WW) + P(WLW) + P(LWW) = 0.258 + 0.2052 + 0.1032 = 0.5664

The first game is played in Boston, second in New York and the Third is in Boston if needed:			
Games' Outcomes	**Probability**	**Series Winner**	**Net Win**
WW	0.2580	Boston	$1,000.00
WLW	0.2052	Boston	$480.00
LWW	0.1032	Boston	$480.00
WLL	0.1368	NY	-$540.00
LWL	0.0688	NY	-$540.00
LL	0.2280	NY	-$1,040.00
SUM:	1.0000		
Probability that Boston wins the series:		**0.5664**	

Figure 1 Probability of outcomes and Net wins tabulated.

(ii) Distribution of Net Win

We construct the probability distribution for the net win, X:

- For WW: Net Win = $1000, Probability = 0.258
- For WLW: Net Win = $480, Probability = 0.2052
- For LWW: Net Win = $480, Probability = 0.1032 (we add these two for $480)
- For WLL: Net Loss = -$540, Probability = 0.1368
- For LWL: Net Loss = -$540, Probability = 0.0688(we add these two for $540)
- For LL: Net Loss = -$1040, Probability = 0.228

Using these, we calculate the expected net win E(X), variance VAR(X), and standard deviation SD(X):

$E(X) = \sum (x \cdot P(x)) = 1000 \cdot 0.258 + 480 \cdot 0.3084 - 540 \cdot 0.2056 - 1040 \cdot 0.228 = \$57.89$
$VAR(X) = \sum [(x^2 \cdot P(x)] - E(X)^2) = 632262.10$
$SD(X) = \sqrt{VAR(X)} = \$795.15$

Answer	Part1(ii)	Distribution of the net win		
	Cum. Prob.	**x**	**Probability**	**x^2**
	1.0000	$1,000.00	0.2580	1000000
	0.7420	$480.00	0.3084	230400
	0.4336	-$540.00	0.2056	291600
	0.2280	-$1,040.00	0.2280	1081600
		SUM:	1.0000	
		E(X)	**$57.89**	
		VAR(X)	632262.10	
		SD(X)	**$795.15**	

Figure 2 Cumulative Probabilities and expected value and standard deviation tabulated.

(iii) Simulation and Confidence Interval

We simulate 10,000 values for Y which is the rubric assigned name for the simulated random values for X. We use the excel RAND () formula to generate random values from a uniform distribution. We justify the utilization of a uniform distribution on the grounds that the theoretical probabilities are already given, and the theoretical expected value and Standard Deviation do not exactly match any other theoretical distribution explicitly (for example if variance and mean were same or almost same we would suspect a Poisson distribution etc). We compute the mean and standard deviation of the simulated values:

- E(Y) = $59.92
- SD(Y) = $799.50

We construct a 95% confidence interval for the expected net win:

$CI = [E(Y) \pm 1.96 \times (SD(Y) / \sqrt{N})] = [44.25, 75.59]$

The confidence interval contains the theoretical expected value E(X) = $57.89. This is important to model the risk.

Answer to Part 1(iii)			
	Simulation		Theoretical:
E(Y)	$	59.92	$57.89
SD(Y)	$	799.50	$795.15
Confidence Interval Lower Limit:		$	44.25
Confidence Interval Upper Limit:		$	75.59
The confidence interval contains the expected values			

Figure 3 Theoretical and observed expected value after 10000 simulations.

Visualization to demonstrate the law of large numbers leading to near convergence of the mean.

Number of Simulations N	Observed Average Win	Theoretical Expected Win
100	$56.80	$57.89
200	$41.20	$57.89
300	$73.53	$57.89
400	$74.40	$57.89
500	$65.68	$57.89
600	$85.43	$57.89
700	$86.46	$57.89
800	$97.33	$57.89
900	$98.93	$57.89
1000	$99.82	$57.89
2000	$86.82	$57.89
3000	$65.08	$57.89
4000	$57.98	$57.89
5000	$57.02	$57.89
6000	$55.24	$57.89
7000	$59.69	$57.89
8000	$56.70	$57.89
9000	$54.68	$57.89
10000	$59.92	$57.89

Figure 4 The observed mean of the outcome gets closer to the theoretical expected value as the simulation progresses.

The above table depicts the mean of the simulations approaching the expected value in confirmation of the law of large numbers.

Below is a graphical representation of the means of the 10,000 simulations towards the theoretical expected value in the form of two, line graphs.

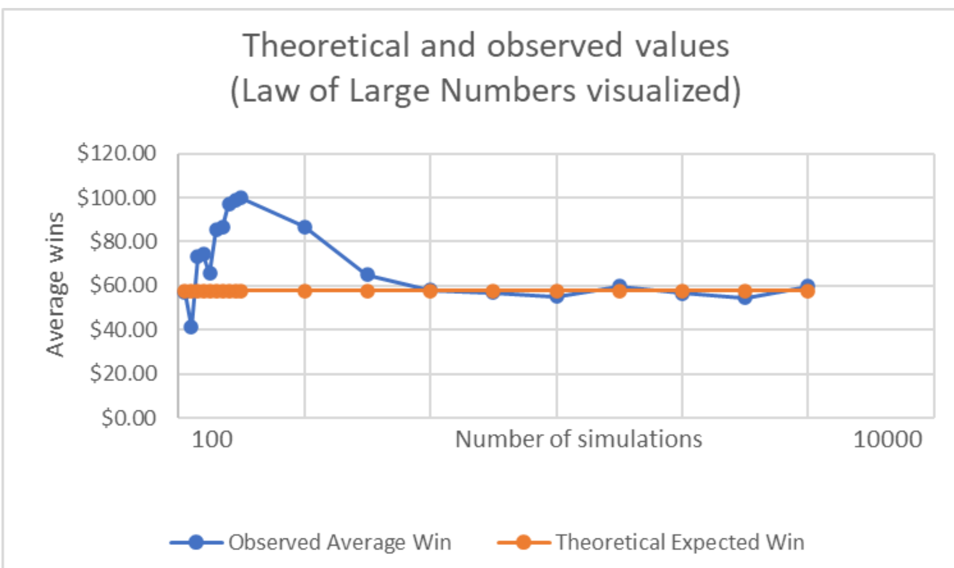


Figure 5 Line Graph demonstrating the Law of Large Numbers for our simulation.

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

The chi-squared goodness-of-fit test is used to determine how well the observed frequency distribution of net wins aligns with the theoretical distribution. Here are the detailed steps and hypotheses for the test:

Hypotheses:

Null Hypothesis (H0): The observed distribution of net wins is the same as the theoretical distribution.
Alternative Hypothesis (H1): The observed distribution of net wins is different from the theoretical distribution.

Test Statistic:

The chi-squared test statistic is calculated using the formula:
$\chi^2 = \sum ((Oi - Ei)^2 / Ei)$
Where Oi is the observed frequency and Ei is the expected frequency for each category of net win.

Test Statistic Value:
$\chi^2 = 3.472$

Win	Observed Frequency	Theoretical Probability	Expected Frequency	(Observed - Expected)^2/Expected
1,000.00	2638	0.2580	2580	1.3039
480.00	3033	0.3084	3084	0.8434
-540.00	2015	0.2056	2056	0.8176
-1,040.00	2314	0.2280	2280	0.5070
SUM:	10000	1	10000	**3.472**
			Chi-squared Metric:	**3.472**
			P-value	**0.3244**
			Decision:	Do not reject Ho: There is no significant difference between the distributions of X and Y.

Figure 6 Frequency Distribution of the outcomes and the Chi-Square Goodness of fit test.

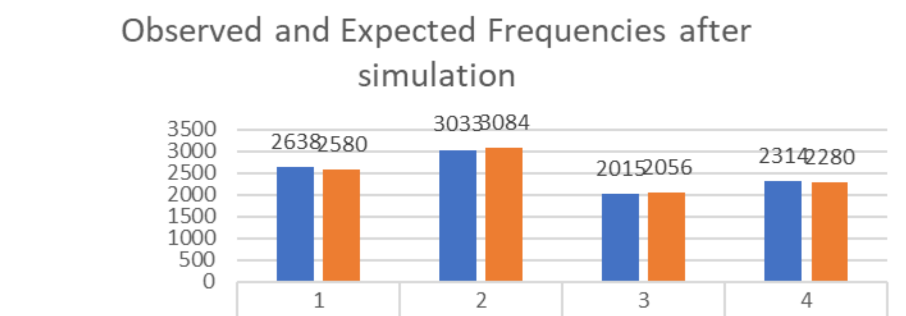
Above is a frequency distribution of the observed and expected values in fulfillment of the assignment rubric and in order to calculate the test statistic and the p-value.

P-Value:
Using a chi-squared distribution table with 3 degrees of freedom (k-p-1, since p or parameters is zero as we are simulating a uniform distribution, whose theoretical parameters were given in the problem. Hence degrees of freedom is-k or number of categories - 1), the p-value corresponding to $\chi^2 = 3.472$ is approximately 0.3244.

Alpha (α):
The significance level is set at α = 0.05.

Decision:
Since the p-value (0.3244) is greater than the significance level (α = 0.05), we fail to reject the null hypothesis.

Conclusion:
There is no significant difference between the observed and theoretical distributions of net wins. The chi-squared goodness-of-fit test confirms that the simulation aligns well with the theoretical distribution. In addition, as the bar chart below makes clear and as we have used random values from a uniform distribution, we may posit reasonably that both the values are drawn from a uniform distribution.



	1	2	3	4
Observed Frequency	2638	3033	2015	2314
Expected Frequency	2580	3084	2056	2280

Figure 7 A Bar Graph of the expected and theoretical frequencies indicates a uniform distribution.

(v) Summary

The betting strategy appears favorable based on the expected net win of $57.89. The simulation results corroborate the theoretical analysis, and the confidence interval includes the expected value. The goodness-of-fit test confirms that the simulation aligns well with the theoretical distribution, supporting the reliability of the model. However, the large value of the standard deviation ($795.15) must give caution as to the risk in the betting. In business analytics standard deviation is a crucial measure that reflects risk.

Analysis of a Betting Strategy in Sports: Red Sox vs. Yankees (Part 2)

Introduction

This report continues the analysis of the betting strategy for a best-of-three series between the Boston Red Sox and the New York Yankees. In this part, we consider the scenario where the first game is played in New York, the second game in Boston, and the third game, if necessary, is in New York. The primary objectives are to calculate the probability of Boston winning the series, determine the expected net win, and evaluate the effectiveness of the betting strategy. We will also perform a simulation to estimate the expected net win and conduct a chi-squared goodness-of-fit test.

Part 2: Best-of-Three Series (First Game in New York, Second in Boston, third in New York if Needed)

(i) Probability of the Red Sox Winning the Series

The probabilities for each game outcome, given the probabilities of winning at home for each team (0.6 for Boston and 0.57 for New York and so on as described in Part 1), are calculated using the basic multiplication rule of mutually exclusive event probabilities as follows:

- WW (Boston wins both games): P(WW) = 0.258
- WLW (Boston wins, loses, then wins): P(WLW) = 0.074
- LWW (Boston loses, wins, then wins): P(LWW) = 0.1471
- WLL (Boston wins, then loses both): P(WLL) = 0.0980
- LWL (Boston loses, wins, then loses): P(LWL) = 0.1949
- LL (Boston loses both games): P(LL) = 0.2280

The probability that Boston wins the series is the sum of the probabilities of all outcomes in which Boston wins:

P (Boston wins series) = P(WW) + P(WLW) + P(LWW) = 0.258 + 0.0740 + 0.1471 = 0.4790

The first game is played in NY, the second in Boston, and the third game (if it becomes necessary) is in New York			
Games' Outcomes	Probability	Series Winner	Net Win
WW	0.2580	Boston	$1,000.00
WLW	0.0740	Boston	$480.00
LWW	0.1471	Boston	$480.00
WLL	0.0980	NY	-$540.00
LWL	0.1949	NY	-$540.00
LL	0.2280	NY	-$1,040.00
SUM:	1.0000		
Probability that Boston wins the series:		0.4790	

Figure 8 Probability of outcomes and Net wins tabulated.

(ii) Distribution of Net Win

We construct the probability distribution for the net win, X:

- For WW: Net Win = \$1000, Probability = 0.258
- For WLW: Net Win = \$480, Probability = 0.0740
- For LWW: Net Win = \$480, Probability = 0.1471
- For WLL: Net Loss = -\$540, Probability = 0.0980
- For LWL: Net Loss = -\$540, Probability = 0.1949
- For LL: Net Loss = -\$1040, Probability = 0.228

Using these, we calculate the expected net win E(X), variance VAR(X), and standard deviation SD(X):

$E(X) = \sum (x \cdot P(x)) = 1000 \cdot 0.258 + 480 \cdot 0.2210 - 540 \cdot 0.2930 - 1040 \cdot 0.228 = -\$31.24$
$VAR(X) = \sum [(x^2 \cdot P(x)] - E(X)^2) = 639984.86$
$SD(X) = \sqrt{VAR(X)} = \$799.99$

	Distribution of the net win		
Cum. Prob.	x	Probability	x^2
1	\$1,000.00	0.2580	1000000
0.7420	\$480.00	0.2210	230400
0.5210	-\$540.00	0.2930	291600
0.2280	-\$1,040.00	0.2280	1081600
	SUM:	1.0000	
	E(X)	-\$31.24	
	VAR(X)	\$639,984.86	
	SD(X)	\$799.99	

Figure 9 Cumulative Probabilities and expected value and standard deviation tabulated.

(iii) Simulation and Confidence Interval

We simulate 10,000 values for Y and compute the mean and standard deviation:

- E(Y) = -\$24.11
- SD(Y) = \$803.27

We construct a 95% confidence interval for the expected net win:

$CI = [E(Y) \pm 1.96 \times (SD(Y) / \sqrt{N})] = [-39.85, -8.36]$

The confidence interval does not contain the theoretical expected value E(X) = -\$31.24.

	Simulation	Theoretical:
E(Y)	-24.11	-31.24
SD(Y)	803.27	799.99

Confidence Interval Lower Limit:	-39.85
Confidence Interval Upper Limit:	-8.36
The confidence interval contains the expected values	

Figure 10 Theoretical and observed expected value after 10000 simulations.

Visualization to demonstrate the law of large numbers leading to near convergence of the mean.

Number of Simulations N	Observed Average Win	Theoretical Expected Win
$100.00	-$21.20	-$31.24
$200.00	-$56.60	-$31.24
$300.00	-$40.93	-$31.24
$400.00	-$50.90	-$31.24
$500.00	-$36.32	-$31.24
$600.00	-$37.00	-$31.24
$700.00	-$42.57	-$31.24
$800.00	-$24.53	-$31.24
$900.00	-$28.00	-$31.24
$1,000.00	-$32.26	-$31.24
$2,000.00	-$19.24	-$31.24
$3,000.00	-$18.94	-$31.24
$4,000.00	-$24.20	-$31.24
$5,000.00	-$23.83	-$31.24
$6,000.00	-$28.62	-$31.24
$7,000.00	-$27.29	-$31.24
$8,000.00	-$23.42	-$31.24
$9,000.00	-$21.49	-$31.24
$10,000.00	-$24.11	-$31.24

Figure 11 The observed mean of the outcome gets closer to the theoretical expected value as the simulation progresses.

The above table depicts the mean of the simulations approaching the expected value in confirmation of the law of large numbers.

Below is a graphical representation of the means of the 10,000 simulations towards the theoretical expected value in the form of two, line graphs.

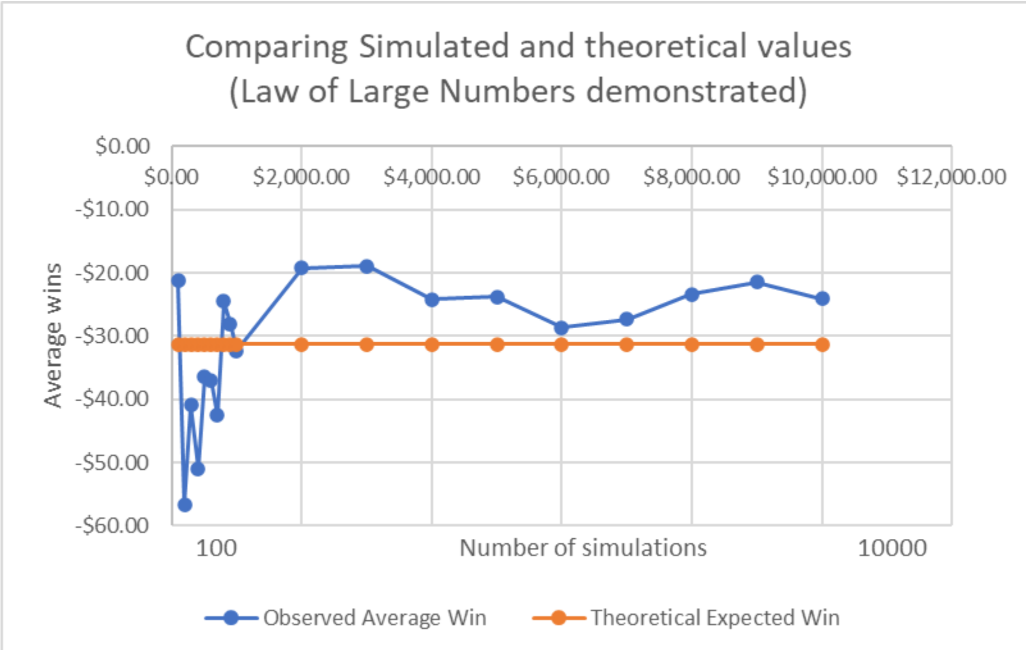


Figure 12 Line Graph demonstrating the Law of Large Numbers for our simulation.

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

The chi-squared goodness-of-fit test is used to determine how well the observed frequency distribution of net wins aligns with the theoretical distribution. Here are the detailed steps and hypotheses for the test:

Hypotheses:

Null Hypothesis (H0): The observed distribution of net wins is the same as the theoretical distribution.
Alternative Hypothesis (H1): The observed distribution of net wins is different from the theoretical distribution.

Test Statistic:

The chi-squared test statistic is calculated using the formula:
$\chi^2 = \sum ((Oi - Ei)^2 / Ei)$
Where Oi is the observed frequency and Ei is the expected frequency for each category of net win.

Test Statistic Value:
$\chi^2 = 2.728$

P-Value:
Using a chi-squared distribution table with 3 degrees of freedom (k-p-1, since p or parameters is zero as we are simulating a uniform distribution, whose theoretical

parameters were given in the problem. Hence degrees of freedom is-k or number of categories - 1), the p-value corresponding to $\chi^2 = 2.728$ is approximately 0.4354.

Win	Observed Frequency	Theoretical Probability	Expected Frequency	(Observed - Expected)^2/Expected
$1,000.00	2649	0.2580	2580	1.8453
$480.00	2174	0.2210	2210.2	0.5929
-$540.00	2901	0.2930	2929.8	0.2831
-$1,040.00	2276	0.2280	2280	0.0070
SUM:	10000	1.0000	10000	**2.728**
			Chi-squared Metric:	2.728
			P-value	0.4354
			Decision:	Do not reject Ho: There is no significant disagreement between the distributions of X and

Figure 13 Frequency Distribution of the outcomes and the Chi-Square Goodness of fit test.

Above is a frequency distribution of the observed and expected values in fulfillment of the assignment rubric and in order to calculate the test statistic and the p-value.

Alpha (α):

The significance level is set at $\alpha = 0.05$.

Decision:

Since the p-value (0.4354) is greater than the significance level ($\alpha = 0.05$), we fail to reject the null hypothesis.

Conclusion:

There is no significant difference between the observed and theoretical distributions of net wins. The chi-squared goodness-of-fit test confirms that the simulation aligns well with the theoretical distribution.

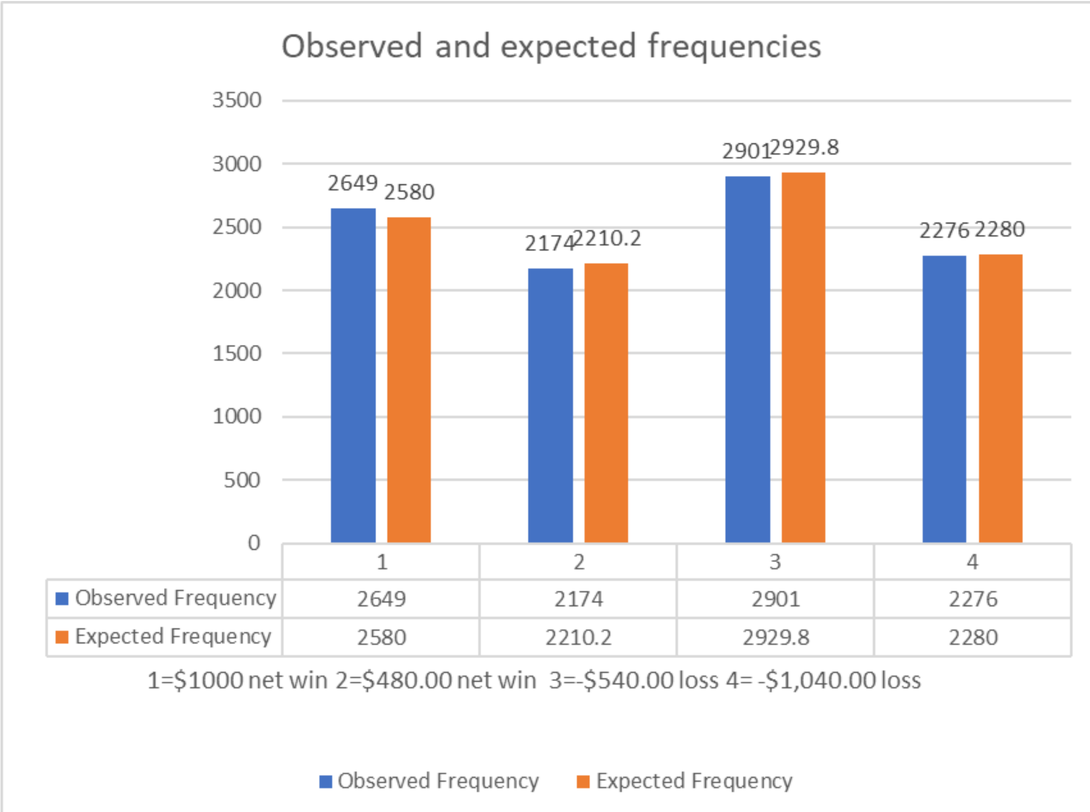


Figure 14 A Bar Graph of the expected and theoretical frequencies indicates a uniform distribution.

We observe that the values of the expected and theoretical frequencies appear to approximate a uniform distribution.

(v) Summary

In this scenario, the betting strategy appears **less favorable** with an expected net loss of -$31.24. The simulation results show a slight difference from the theoretical expected value, as indicated by the confidence interval. However, the chi-squared goodness-of-fit test confirms that the simulation aligns well with the theoretical distribution, supporting the reliability of the model. The high standard deviation of $799.99 also points toward a rather unfavorable betting strategy that would best be avoided. Standard deviation in business analytics reflects the risk of an enterprise.

Analysis of a Betting Strategy in Sports: Red Sox vs. Yankees (Part 3)

Introduction

This report completes the analysis of the betting strategy for a series between the Boston Red Sox and the New York Yankees. In this part, we consider the scenario where the series is a best-of-five series, with games alternating between Boston and New York, starting in Boston. The primary objectives are to calculate the probability of Boston winning the series, determine the expected net win, and evaluate the effectiveness of the betting strategy. We will also perform a simulation to estimate the expected net win and conduct a chi-squared goodness-of-fit test.

Part 3: Best-of-Five Series (First Game in Boston, Alternating Venues)

(i) Probability of the Red Sox Winning the Series

The probabilities for each game outcome, given the probabilities of winning at home for each team (0.6 for Boston and 0.57 for New York), are calculated using the basic multiplication rule of mutually exclusive event probabilities as follows:

- WWW (Boston wins 3-0): P(WWW) = 0.1548
- WWLW (Boston wins 3-1): P(WWLW) = 0.0444
- WLWW (Boston wins 3-1): P(WLWW) = 0.0882
- LWWW (Boston wins 3-1): P(LWWW) = 0.0444
- WWLLW (Boston wins 3-2): P(WWLLW) = 0.0353
- WLWLW (Boston wins 3-2): P(WLWLW) = 0.0702
- LWWLW (Boston wins 3-2): P(LWWLW) = 0.0353
- WLLWW (Boston wins 3-2): P(WLLWW) = 0.0353
- LWLWW (Boston wins 3-2): P(LWLWW) = 0.0178
- LLWWW (Boston wins 3-2): P(LLWWW) = 0.0353
- LLL (NY wins 3-0): P(LLL) = 0.0912
- LLWL (NY wins 3-1): P(LLWL) = 0.0780
- LWLL (NY wins 3-1): P(LWLL) = 0.0392
- WLLL (NY wins 3-1): P(WLLL) = 0.0780
- LLWWL (NY wins 3-2): P(LLWWL) = 0.0235
- LWWLL (NY wins 3-2): P(LWWLL) = 0.0235
- WWLLL (NY wins 3-2): P(WWLLL) = 0.0235
- LWLWL (NY wins 3-2): P(LWLWL) = 0.0118
- WLLWL (NY wins 3-2): P(WLLWL) = 0.0235
- WLWLL (NY wins 3-2): P(WLWLL) = 0.0468

The probability that Boston wins the series is the sum of the probabilities of all outcomes in which Boston wins:

P (Boston wins) = P(WWW) + P(WWLW) + P(WLWW) + P(LWWW) + P(WWLLW) + P(WLWLW) + P(LWWLW) + P(WLLWW) + P(LWLWW) + P(LLWWW)

P (Boston wins) = 0.1548 + 0.0444 + 0.0882 + 0.0444 + 0.0353 + 0.0702 + 0.0353 + 0.0353 + 0.0178 + 0.0353 = 0.5609

The first game is played in Boston then alternates stadiums till a best of five winner is reached:			
Games' Outcomes	**Probability**	**Series Winner**	**Net Win**
WWW	0.1548	Boston	$1,500.00
WWLW	0.0444	Boston	$980.00
WLWW	0.0882	Boston	$980.00
LWWW	0.0444	Boston	$980.00
WWLLW	0.0353	Boston	$460.00
WLWLW	0.0702	Boston	$460.00
LWWLW	0.0353	Boston	$460.00
WLLWW	0.0353	Boston	$460.00
LWLWW	0.0178	Boston	$460.00
LLWWW	0.0353	Boston	$460.00
LLL	0.0912	NY	-$1,560.00
LLWL	0.0780	NY	-$1,060.00
LWLL	0.0392	NY	-$1,060.00
WLLL	0.0780	NY	-$1,060.00
LLWWL	0.0235	NY	-$560.00
LWWLL	0.0235	NY	-$560.00
WWLLL	0.0235	NY	-$560.00
LWLWL	0.0118	NY	-$560.00
WLLWL	0.0235	NY	-$560.00
WLWLL	0.0468	NY	-$560.00
SUM:	1.000000000		
Probability that Boston wins the series:		**0.5609**	

Figure 15 Probability of outcomes and Net wins tabulated.

(ii) Distribution of Net Win

We construct the probability distribution for the net win, X:

- For WWW: Net Win = $1500, Probability = 0.1548
- For WWLW: Net Win = $980, Probability = 0.0444
- For WLWW: Net Win = $980, Probability = 0.0882
- For LWWW: Net Win = $980, Probability = 0.0444
- For WWLLW: Net Win = $460, Probability = 0.0353
- For WLWLW: Net Win = $460, Probability = 0.0702
- For LWWLW: Net Win = $460, Probability = 0.0353
- For WLLWW: Net Win = $460, Probability = 0.0353
- For LWLWW: Net Win = $460, Probability = 0.0178

- For LLWWW: Net Win = \$460, Probability = 0.0353
- For LLL: Net Loss = -\$1560, Probability = 0.0912
- For LLWL: Net Loss = -\$1060, Probability = 0.0780
- For LWLL: Net Loss = -\$1060, Probability = 0.0392
- For WLLL: Net Loss = -\$1060, Probability = 0.0780
- For LLWWL: Net Loss = -\$560, Probability = 0.0235
- For LWWLL: Net Loss = -\$560, Probability = 0.0235
- For WWLLL: Net Loss = -\$560, Probability = 0.0235
- For LWLWL: Net Loss = -\$560, Probability = 0.0118
- For WLLWL: Net Loss = -\$560, Probability = 0.0235
- For WLWLL: Net Loss = -\$560, Probability = 0.0468

Using these, we calculate the expected net win E(X), variance VAR(X), and standard deviation SD(X):

$E(X) = \sum (x \cdot P(x)) = 1500 \cdot 0.1548 + 980 \cdot 0.1770 + 460 \cdot 0.2291 - 560 \cdot 0.1527 - 1060 \cdot 0.1952 - 1560 \cdot 0.0912 = \$76.35$

$VAR(X) = \sum [(x^2 \cdot P(x)] - E(X)^2) = 1050061.84$

$SD(X) = \sqrt{VAR(X)} = \$1024.73$

	Distribution of the net win		
Cum. Prob.	x	Probability	x^2
1.0000	\$1,500.00	0.1548	\$ 2,250,000.00
0.8452	\$980.00	0.1770	\$ 960,400.00
0.6682	\$460.00	0.2291	\$ 211,600.00
0.4391	-\$560.00	0.1527	\$ 313,600.00
0.2864	-\$1,060.00	0.1952	\$ 1,123,600.00
0.0912	-\$1,560.00	0.0912	\$ 2,433,600.00
	SUM:	1.0000	
	E(X)	**\$76.35**	
	VAR(X)	1050061.84	
	SD(X)	**\$1,024.73**	

Figure 16 Cumulative Probabilities and expected value and standard deviation tabulated.

The above table depicts the probability distribution of the random variable in theoretical terms.

(iii) Simulation and Confidence Interval

We simulate 10,000 values for Y which is the rubric assigned name of the simulated random variable, and we compute the mean and standard deviation:

- E(Y) = \$60.20
- SD(Y) = \$1032.03

We construct a 95% confidence interval for the expected net win:

CI = [E(Y) ± 1.96 × (SD(Y) / √N)] = [39.98, 80.43]

The confidence interval contains the theoretical expected value E(X) = $76.35.

	Simulation	Theoretical:
E(Y)	$60.20	$76.35
SD(Y)	$1,032.03	$1,024.73

Confidence Interval Lower Limit:	$39.98
Confidence Interval Upper Limit:	$80.43

Figure 17 Theoretical and observed expected value after 10000 simulations.

Visualization to demonstrate the law of large numbers leading to near convergence of the mean.

Number of Simulations N	Observed Average Win	Theoretical Expected Win
100	$106.20	$76.35
200	$76.10	$76.35
300	$67.33	$76.35
400	$51.75	$76.35
500	$61.64	$76.35
600	$78.57	$76.35
700	$88.26	$76.35
800	$73.93	$76.35
900	$56.07	$76.35
1000	$50.94	$76.35
2000	$47.61	$76.35
3000	$66.80	$76.35
4000	$77.89	$76.35
5000	$70.99	$76.35
6000	$75.70	$76.35
7000	$72.79	$76.35
8000	$68.20	$76.35
9000	$66.97	$76.35
10000	$60.20	$76.35

Figure 18 The observed mean of the outcome gets closer to the theoretical expected value as the simulation progresses.

As indicated in the tabulation above, larger simulations favor an approximation of the expected value. The below graphical representation of the means of the 10,000 simulations towards the theoretical expected value through line graphs embodies the same visually.

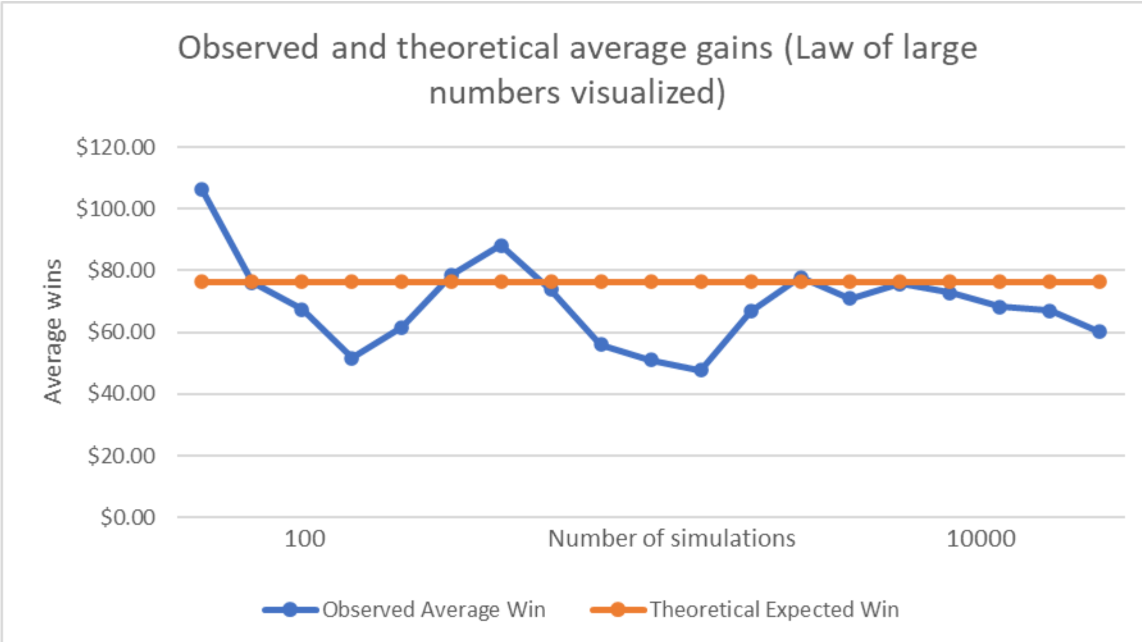


Figure 19 Line Graph demonstrating the Law of Large Numbers for our simulation.

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

The chi-squared goodness-of-fit test is used to determine how well the observed frequency distribution of net wins aligns with the theoretical distribution. Here are the detailed steps and hypotheses for the test:

Hypotheses:

Null Hypothesis (H_0): The observed distribution of net wins is the same as the theoretical distribution.

Alternative Hypothesis (H_1): The observed distribution of net wins is different from the theoretical distribution.

Test Statistic:

The chi-squared test statistic is calculated using the formula:

$\chi^2 = \sum ((O_i - E_i)^2 / E_i)$

Where O_i is the observed frequency and E_i is the expected frequency for each category of net win.

Test Statistic Value:

$\chi^2 = 7.127$

P-Value:

Using a chi-squared distribution table with 5 degrees of freedom (number of categories - 1), the p-value corresponding to $\chi^2 = 7.127$ is approximately 0.2114.

Alpha (α):

The significance level is set at $\alpha = 0.05$.

Win	Observed Frequency	Theoretical Probability	Expected Frequency	$(\text{Observed - Expected})^2$/Expected
$1,500.00	1530	0.1548	1548	0.2093
$980.00	1766	0.1770	1769.88	0.0085
$460.00	2255	0.2291	2291.064	0.5677
-$560.00	1479	0.1527	1527.376	1.5322
-$1,060.00	2001	0.1952	1951.68	1.2463
-$1,560.00	969	0.0912	912	3.5625
SUM:	10000	1.0000	10000	**7.127**
			Chi-squared Metric:	7.127
			P-value	0.2114
			Decision:	Do not reject Ho: There is no significant disagreement between the distributions of X and Y.

Figure 20 Frequency Distribution of the outcomes and the Chi-Square Goodness of fit test.

Decision:
Since the p-value (0.2114) is greater than the significance level (α = 0.05), we fail to reject the null hypothesis.

Conclusion:
There is no significant difference between the observed and theoretical distributions of net wins. The chi-squared goodness-of-fit test confirms that the simulation aligns well with the theoretical distribution.

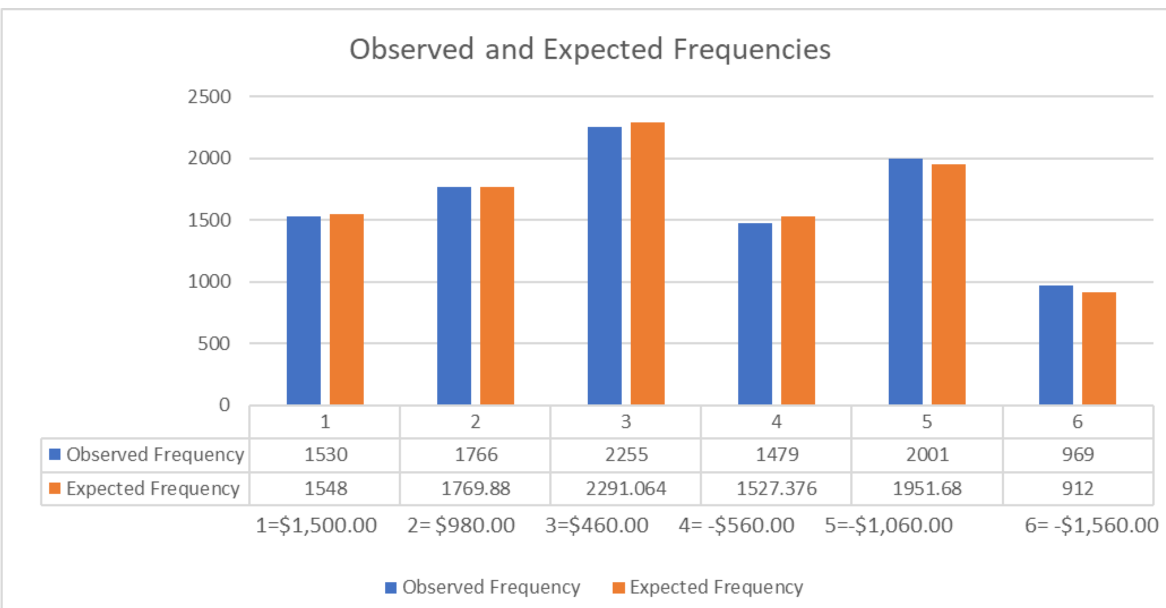


Figure 21 A Bar Graph of the expected and theoretical frequencies indicates a uniform distribution.

(v) Summary

In this scenario, the betting strategy appears favorable with an expected net win of $76.35. The simulation results corroborate the theoretical analysis, and the confidence interval includes the expected value. The chi-squared goodness-of-fit test confirms that the simulation aligns well with the theoretical distribution, supporting the reliability of the model. However, again as seen in the previous betting strategy in Part 1 a very high standard deviation suggests that a measure of caution is necessary before embarking on this betting enterprise.

To summarize all three betting strategies: strategy 1 and 3 seem sound but with substantial variation in the possible net wins. Strategy 2, on the other hand, appears unsound with an expected loss and ought to be revisited.

References

Evans, J. R. (2021). *Statistics, data analysis, and decision modeling* (5th ed.). Pearson.

Mendenhall, W., Beaver, R. J., & Beaver, B. M. (2012). *Introduction to probability and statistics. Cengage Learning.*