Probability Distribution
Betting Strategy in Baseball games

Heejae Roh
Northeastern University
ALY 6050: Intro to Enterprise Analytics
Azadeh Mobasher
February 28, 2023

INTRODUCTION

In this assignment, I solve a probability problem with an example of betting in sports, one of the fields where probability is most commonly used. For further learning of Excel, I use Excel. In the process, I learn the process to solve the problem by giving reasons.

ANALYSIS

Understanding

The key metrics is that Boston Red Sox and New York Yankees are scheduled to play three or five games. The probability that the Red Sox win a game in their home stadium is 0.6. The probability that Yankees win their home game is 0.57. Bet on each game, you win $500 if the Red Sox win. You lose $520 if the Red sox lose the game.

PART 1

Probability & Confidence Interval

The possibility of Boston wins the series is 0.57. By Probability Distribution for net win (X), the mean is 57.89 and standard deviation is 795.15. It matches with the 10,000 sampling result whose confidence interval is from 39.34 to 70.57. The expected value of 57.89 is within the CI range.

Chi-squared Test & Strategy

There is not enough evidence to reject null hypothesis which is that -1040, -540, 480 and 1000 is distributed as 22.8%, 20.6%, 30.8%, and 25.8% since 0.87 (p-value) > 0.05. Boston's betting strategy is favorable. Because I have an expected value of 57.89.

PART 2

Probability & Confidence Interval

The possibility of Boston wins the series is 0.48. By Probability Distribution for net win (X), the mean is -31.24 and standard deviation is 799.99. It matches with sampling result whose confidence interval is from -55.81 to -24.47.

Chi-squared Test & Strategy

There is not enough evidence to reject null hypothesis since 0.48 (p-value) > 0.05. Boston's betting strategy is not favorable. Because I have an expected value of -31.24.

PART 3

Probability & Confidence Interval

The possibility of Boston wins the series is 0.56. By Probability Distribution for net win (X), the mean is 76.35 and standard deviation is 1024.73. It matches with sampling result whose confidence interval is from 63.76 to 104.14.

Chi-squared Test & Strategy

There is not enough evidence to reject null hypothesis since 0.20 (p-value) > 0.05. Boston's betting strategy is favorable. Because I have an expected value of 76.35.

CONCLUSION

I learn the way to make decision. And I also get used to the method that make simulation in Excel. By using Excel, I can do many experiments that are difficult to do in reality. Based on this, I can have a basis for making rational decisions.

PART 0. INTRODUCTION

In this module I study probability distributions. Probability distributions can be divided into two which are discrete and continuous. The main other distributions are Uniform/ Triangular/ Exponential/ Normal / Standard Beta/ Gamma/ Log-Normal/ Weibull/ Chi-Squared. In this assignment, I solve a probability problem with an example of betting in sports, one of the fields where probability is most commonly used. For further learning of Excel, I solve the problems with Excel. In the process, I learn the process to solve the problem by giving reasons.

PART 1. ANALYSIS

Understanding Key metrics

1. This is Boston Red Sox and New York Yankees are scheduled to play three games
2. The winner of the series will be the first team that wins two of the three games
3. The probability that the Red Sox win a game in their home stadium is 0.6. The probability that the Yankees win a game in Boston stadium is 0.4
4. The probability that Yankees win their home game is 0.57. The probability that the Boston win a game in Yankees stadium is 0.43
5. Bet on each game, you win $500 if the Red Sox win. You lose $520 if the Red sox lose the game

Win rate of each team	in Boston	in Yankees
Boston	0.6	0.43
Yankees	0.4	0.57

For each game	Boston win	Boston lose
My account	+$500	-$520

PART 1-1. Calculate the Probability

Key metrics of PART 1: 1st game in Boston, 2nd in Yankees, 3rd in Boston
Boston wins, 3 Possible cases

Winner/ Home	1st Boston (B)	2nd Yankees (Y)	3rd Boston	
① 1st B 2nd B	0.6	0.43		0.258
② 1st B 2nd Y 3rd B	0.6	0.57	0.6	0.2052
③ 1st Y 2nd B 3rd B	0.4	0.43	0.6	0.1032
			SUM	0.5664

Yankees wins, 3 Possible cases

Winner/ Home	1st Boston (B)	2nd Yankees (Y)	3rd Boston	
④ 1st Y 2nd Y	0.4	0.57		0.228
⑤ 1st Y 2nd B 3rd Y	0.4	0.43	0.4	0.0688
⑥ 1st B 2nd Y 3rd Y	0.6	0.57	0.4	0.1368
			SUM	0.4336

Understanding Key metrics

1. Above table shows the cases of Boston win the series. The possibility is 0.57
2. By calculating 3 possible cases of Yankees win the series. The possibility is 0.43
3. By adding two possibilities, I can confirm that 6 cases contain all possible cases, since the sum of possibility is 1

PART 1-2. Probability Distribution for net win (X)

Suppose B= wins of Boston, Y= wins of Yankees
X= net win

	X	P(X)	X*P(X)	$(X-\mu)^2$	$(X-\mu)^2*P(X)$
B=0 Y=2 ④	-1040	0.228	-237.12	1205358.1	274821.6
B=1 Y=2 ⑤, ⑥	-540	0.2056	-111.024	357470.1	73495.8
B=2 Y=1 ②, ③	480	0.3084	148.032	178178.5	54950.3
B=2 Y=0 ①	1000	0.258	258	887575.0	228994.4
	SUM	1	Mean 57.888		VAR 632262.1
					SD 795.15

Interpretation

1. Converge the number of X to 4 in all cases which means lose $1040, lose $540, earn $480, earn $1000
2. Each probability is the probability of adding the number of each case as shown in the first column of the table
3. The mean is obtained as the sum of the values of X*P(X). Mean is 57.88 which means if I bet on Boston in every game, the expected earning is 57.89
4. The variance is the sum of (X-μ)2*P(X). Root-squared VAR is SD which is 795.15
5. These are the expected values; the population mean and the variance of population

PART 1-3. 10,000 random values for X & Confidence interval

10 Head of 10,000 random values		Table for VLOOKUP	
RAND()	X	CUMULATIVE	X
0.065173	-1040	0	-1040
0.667324	480	0.228	-540
0.023064	-1040	0.4336	480
0.887411	1000	0.742	1000
0.885267	1000	1	
0.402153	-540		
0.62466	480		
0.078169	-1040		
0.77384	1000		
0.828824	1000		

10,000 Random Values

1. (Left table, r) Using RAND(), Create 10,000 random values
2. (Left table, x) Using VLOOKUP(target, cumulative table) assign specific X
3. (Right table) Table for VLOOKUP which contains the X and cumulative values
4. Since I conduct simulation by creating random values between 0 and 1, I perform VLOOKUP with cumulative values

OBSERVED		CONFIDENCE	95%
average	54.95	15.6146724	
Variance	634701.57	CI	
SD	796.68	39.34	70.57

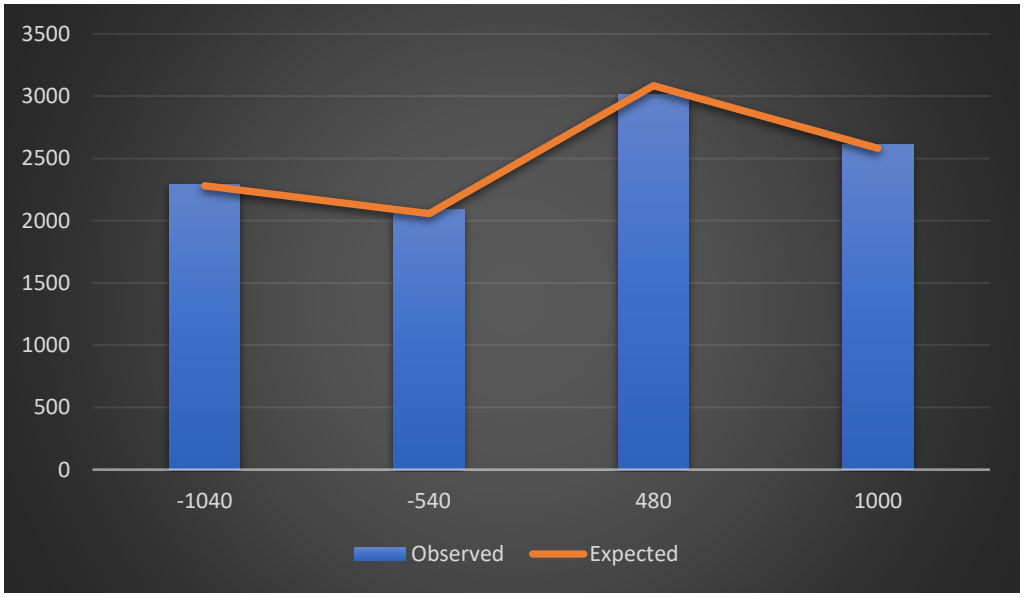
EXPECTED		Df = M - m - 1	
average	57.89	M	4
variance	632262.10	m	0
SD	795.15	Df	3

Interpretation

1. Mean is 57.89 and SD is 795.15 which is expected value calculated through probability
2. This means that the average expected amount is 57.89 when betting on Boston's victory in all games in the above home & away condition
3. As a result of repeating this 10,000 times, the average was 54.95 and the SD was 796.68
4. The Confidence interval is from 39.34 to 70.57, so the expected value of 57.89 is within the CI range

PART 1-4. A Frequency Distribution for Y & chi-squared

Y	Expected. Freq.	Observed. Freq.	Chi-Squared	
				Chi-Squared
-1040	2280	2293	0.074122807	Metric
-540	2056	2071	0.109435798	0.72
480	3084	3045	0.493190661	P-value
1000	2580	2591	0.046899225	0.87



Chi-squared goodness of Fit Test

Step 0 Assumption

Distribution about categorical variable which are -1040, -540, 480 and 1000. Sample was randomly selected, and minimum of five observations expected in each group (Shaun, 2022)

Step 1 Hypothesis.

Null: The population follows the specified distribution which is that -1040, -540, 480 and 1000 is distributed as 22.8%, 20.6%, 30.8%, and 25.8%

Alternative: The population does not follow the specified distribution

Step 2 Find the critical value.

- The p-value is $\alpha = 0.05$

Step 3 Compute the test value.

- The p-value with 1-CHISQ.DIST() in excel is 0.87
- Because CHISQ.DIST() calculate the area on the left. Chi-squared test is always right area test, so I have to minus chi-squared from the one

Step 4 Make the decision.

- There is not enough evidence to reject null hypothesis
  since 0.87 (p-value) > 0.05

Step 5 Summarize the results.

- There is not enough evidence to reject the claim that the population follows the specified distribution, which is that -1040, -540, 480 and 1000 is distributed as 22.8%, 20.6%, 30.8%, and 25.8%

Interpretation

1. Theoretically, applying the probabilities, when X= -1040, -540, 480, and 1000, the numbers are 2280, 2056, 3084, and 2580 out of 10,000
2. According to our randomly generated simulations, the Observed Frequencies are 2293, 2071, 3045, and 2591
3. The metric calculated by chi-squared is 0.72 and the p-value is 0.87
4. The chi-square statistic compares the observed values to the expected values. This test statistic is used to determine whether the difference between the observed and expected values is statistically significant (Diana, n.d). When Df is 3 in 0.05 alpha, the critical value of chi-squared table is 7.815. In this case the chi-squared metric is 0.72, I cannot reject the null hypothesis and also p-value is greater than 0.05

PART 1-5. My betting Strategy

Strategy Analysis

1. In common sense, if I had to choose whether or not to bet, I would of course choose to bet only on home matches and not bet on away matches
2. This means betting on the first match, not betting on the second match, and betting on the third match
3. If I could bet on just one team in the series not in each game, I would have to bet on Boston and the resulting expected value would be 54.95. In sampling, when simulating 10,000 times, the mean is 57.8, which justifies my strategy because the expected value in the between confidence interval [39.34, 70.57].

PART 2-1. Calculate the Probability

Key metrics of PART 1: 1st game in Yankees, 2nd in Boston, 3rd in Yankees

Boston wins, 3 Possible cases

Winner/ Home	1st Yankees (Y)	2nd Boston (B)	3rd Yankees	
①1st B 2nd B	0.43	0.6		0.258
②1st B 2nd Y 3rd B	0.43	0.4	0.43	0.0740
③1st Y 2nd B 3rd B	0.57	0.6	0.43	0.1471
			SUM	0.4790

Yankees wins, 3 Possible cases

Winner/ Home	1st Yankees (Y)	2nd Boston (B)	3rd Yankees	
④1st Y 2nd Y	0.57	0.4		0.228
⑤1st Y 2nd B 3rd Y	0.57	0.6	0.57	0.1950
⑥1st B 2nd Y 3rd Y	0.43	0.4	0.57	0.0980
			SUM	0.5210

Understanding Key metrics

1. Above table shows the cases of Boston win the series. The possibility is 0.48
2. By calculating 3 possible cases of Yankees win the series. The possibility is 0.52
3. By adding two possibilities, I can confirm that 6 cases contain all possible cases, since the sum of possibility is 1

PART 2-2. Probability Distribution for net win (X)

Suppose B= wins of Boston, Y= wins of Yankees
X= net win

	X	P(X)	X*P(X)	$(X-\mu)^2$	$(X-\mu)^2*P(X)$
B=0 Y=2 ④	-1040	0.228	-237.12	1017597.5	232012.2
B=1 Y=2 ⑤, ⑥	-540	0.2930	-158.2092	258837.1	75834.1
B=2 Y=1 ②, ③	480	0.2210	106.0896	261365.9	57767.1
B=2 Y=0 ①	1000	0.258	258	1063455.1	274371.4
	SUM	1	Mean -31.24		VAR 639984.9
					SD 799.99

Interpretation

1. Converge the number of X to 4 in all cases which means lose $1040, lose $540, earn $480, earn $1000
2. Each probability is the probability of adding the number of each case as shown in

the first column of the table

3. The mean is obtained as the sum of the values of X*P(X). Mean is -31.24 which means if I bet on Boston in every game, the expected loss is 31.24
4. The variance is the sum of (X-μ)2*P(X). Root-squared VAR is SD which is 799.99
5. These are the expected values; the population mean and the variance of population

PART 2-3. 10,000 random values for X & Confidence interval

5 Head of 10,000 random values		Table for VLOOKUP	
RAND()	X	CUMULATIVE	X
0.53954	480	0	-1040
0.429574	-540	0.228	-540
0.440712	-540	0.52098	480
0.661805	480	0.742	1000
0.308188	-540	1	

10,000 Random Values

1. (Left table, r) Using RAND(), Create 10,000 random values
2. (Left table, x) Using VLOOKUP(target, cumulative table) assign specific X
3. (Right table) Table for VLOOKUP which contains the X and cumulative values
4. Since I conduct simulation by creating random values between 0 and 1, I perform VLOOKUP with cumulative values

OBSERVED		CONFIDENCE	95%
average	-40.14	15.672645	
Variance	639423.24	CI	
SD	799.64	-55.81	-24.47

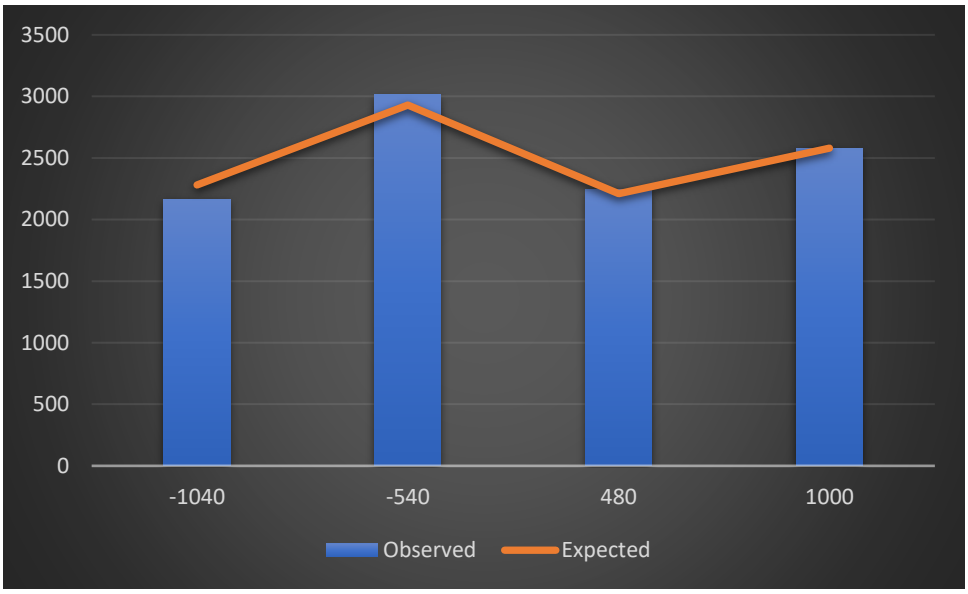
EXPECTED		Df = M - m - 1	
average	-31.24	M	4
variance	639984.86	m	0
SD	799.99	Df	3

Interpretation

1. Mean is -31.24 and SD is 799.99 which is expected value calculated through probability
2. This means that the average expected loss is 31.24 when betting on Boston's victory in all games in the above home & away condition
3. As a result of repeating this 10,000 times, the average was -40.14 and the SD was 799.64
4. The Confidence interval is from -55.81 to -24.47, so the expected value of -31.24 is within the CI range

PART 2-4. A Frequency Distribution for Y & chi-squared

Y	Expected. Freq.	Observed. Freq.	Chi-Squared	
				Chi-Squared
-1040	2280	2330	1.096491228	Metric
-540	2929.8	2915	0.074762782	2.46
480	2210.2	2229	0.15991313	P-value
1000	2580	2526	1.130232558	0.48



Chi-squared goodness of Fit Test

Step 0 Assumption

Distribution about categorical variable which are -1040, -540, 480 and 1000. Sample was randomly selected, and over 2000 observations for each group

Step 1 Hypothesis.

Null: The population follows the specified distribution which is that -1040, -540, 480 and 1000 is distributed as 22.8%, 29.3%, 22.1%, and 25.8%

Alternative: The population does not follow the specified distribution

Step 2 Find the critical value.

- The p-value is $\alpha = 0.05$

Step 3 Compute the test value.

- The p-value with 1-CHISQ.DIST() in excel is 0.48
- Because CHISQ.DIST() calculate the area on the left. Chi-squared test is always

right area test, so I have to minus chi-squared from the one

Step 4 Make the decision.

- There is not enough evidence to reject null hypothesis since 0.87 (p-value) > 0.05

Step 5 Summarize the results.

- There is not enough evidence to reject the claim that the population follows the specified distribution, which is that -1040, -540, 480 and 1000 is distributed as 22.8%, 29.3%, 22.1%, and 25.8%

Interpretation

1. Theoretically, applying the probabilities, when X= -1040, -540, 480, and 1000, the numbers are 2280, 2929.8, 2210.2, and 2580 out of 10,000
2. According to our randomly generated simulations, the Observed Frequencies are 2330, 2915, 2229, and 2526
3. The metric calculated by chi-squared is 2.46 and the p-value is 0.48.
4. When Df is 3 in 0.05 alpha, the critical value of chi-squared table is 7.815. In this case the chi-squared metric is 2.46, I cannot reject the null hypothesis and also p-value is greater than 0.05

PART 2-5. My betting Strategy

Strategy Analysis

1. In common sense, if I had to choose whether or not to bet, I would of course choose to bet only on home matches and not bet on away matches
2. This means not betting on the first match, betting on the second match, and not betting on the third match
3. If I could bet on just one team in the series not in each game, I would have to not bet on Boston and the resulting expected value would be -31.24. In sampling, when simulating 10,000 times, the mean is -40.14, which justifies my strategy because the expected value in the between confidence interval [-55.81, -24.47].

PART 3-1. Calculate the Probability

Key metrics of PART 1: 1st game in Boston, 2nd in Yankees, 3rd in Boston, 4th in Yankees, 5th in Boston

Boston wins, 10 Possible cases

Winner/ Home	1st Boston (B)	2nd Yankees (Y)	3rd B	4th Y	5th B	
BBB	0.6	0.43	0.6			0.1548
BBYB	0.6	0.43	0.4	0.43		0.0444
BYBB	0.6	0.57	0.6	0.43		0.0882
YBBB	0.4	0.43	0.6	0.43		0.0444
YYBBB	0.4	0.57	0.6	0.43	0.6	0.0353
YBYBB	0.4	0.43	0.4	0.43	0.6	0.0178
YBBYB	0.4	0.43	0.6	0.57	0.6	0.0353
BYYBB	0.6	0.57	0.4	0.43	0.6	0.0353
BYBYB	0.6	0.57	0.6	0.57	0.6	0.0702
BBYYB	0.6	0.43	0.4	0.57	0.6	0.0353
					SUM	0.5609

Yankees wins, 10 Possible cases

Winner/ Home	1st Boston (B)	2nd Yankees (Y)	3rd B	4th Y	5th B	
YYY	0.4	0.57	0.4			0.0912
YYBY	0.4	0.57	0.6	0.57		0.077976
YBYY	0.4	0.43	0.4	0.57		0.039216
BYYY	0.6	0.57	0.4	0.57		0.077976
BBYYY	0.6	0.43	0.4	0.57	0.4	0.0235296
BYBYY	0.6	0.57	0.6	0.57	0.4	0.0467856
BYYBY	0.6	0.57	0.4	0.43	0.4	0.0235296
YBBYY	0.4	0.43	0.6	0.57	0.4	0.0235296
YBYBY	0.4	0.43	0.4	0.43	0.4	0.0118336
YYBBY	0.4	0.57	0.6	0.43	0.4	0.0235296
					SUM	0.4391

Understanding Key metrics

1. Above table shows the cases of Boston win the series. The possibility is 0.56
2. By calculating 10 possible cases of Yankees win the series. The possibility is 0.44
3. By adding two possibilities, I can confirm that 20 cases contain all possible cases, since the sum of possibility is 1

PART 3-2. Probability Distribution for net win (X)

Suppose A= wins of Boston, B= wins of Yankees
X= net win

X	P(X)	X*P(X)	$(X-\mu)^2$	$(X-\mu)^2*P(X)$
-1560	0.0912	-142.272	2677654.6	244202.097
-1060	0.1952	-206.878	1291300.5	252020.54
-560	0.1527	-85.5331	404946.47	61850.5526
460	0.2291	105.3889	147184.22	33720.846
980	0.1770	173.4482	816576.01	144524.154
1500	0.1548	232.2	2026767.8	313743.655
SUM	1	Mean 76.35		VAR 1050061.8
				SD 1024.73

Interpretation

1. Converge the number of X to 6 in all cases which means lose 1560, lose 1060, lose 560, earn 460, earn 980, and earn 1500
2. Each probability is the probability of adding the number of each case in 20 possible cases which contain Boston win and Yankees win scenario
3. The mean was then obtained as the sum of the values of X*P(X), Mean is 76.35 which means if I bet on Boston in every game, the expected earning is 76.35
4. The variance is the sum of (X-μ)2*P(X). Root-squared VAR is SD which is 1024.73
5. These are the expected values; the population mean and the variance of population

PART 3-3. 10,000 random values for X & Confidence interval

10 Head of 10,000 random values		Table for VLOOKUP	
RAND()	X	CUMULATIVE	X
0.381938	-560	0	-1560
0.749361	980	0.0912	-1060
0.21185	-1060	0.286368	-560
0.1308	-1060	0.439106	460
0.512752	460	0.668212	980
0.104981	-1060	0.8452	1500
0.203772	-1060	1	
0.659979	460		
0.11599	-1060		
0.58224	460		

10,000 Random Values

1. (Left table, r) Using RAND(), Create 10,000 random values
2. (Left table, x) Using VLOOKUP(target, cumulative table) assign specific X
3. (Right table) Table for VLOOKUP which contains the X and cumulative values
4. Since I conduct simulation by creating random values between 0 and 1, I perform VLOOKUP with cumulative values

OBSERVED		CONFIDENCE	95%
average	83.946	20.19	
variance	1061096.3	CI	
SD	1030.0953	63.76	104.14

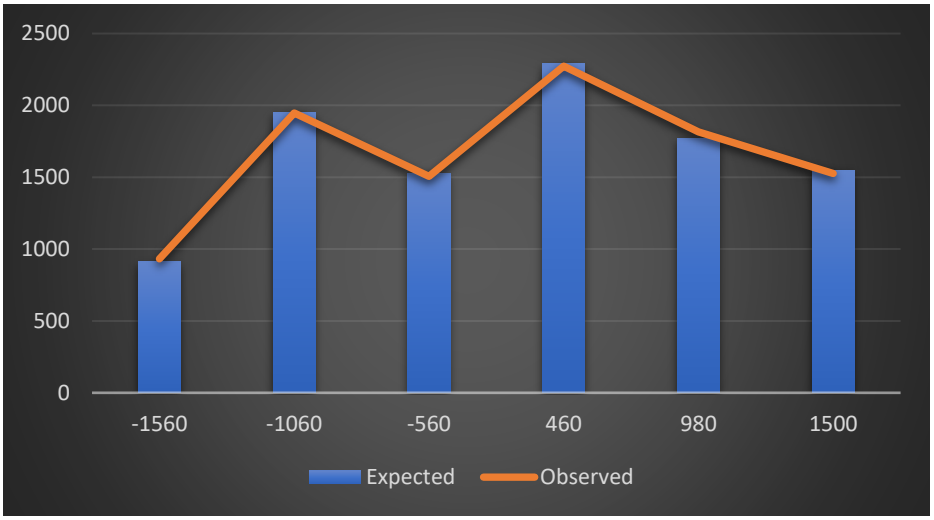
EXPECTED		Df = M - m - 1	
average	76.35	M	6
variance	1050061.84	m	0
SD	1024.73	Df	5

Interpretation

1. Mean is 76.35 and SD is 1024.73 which is expected value calculated through probability
2. This means that the average expected amount is 76.35 when betting on Boston's victory in all games in the above 5 series and home & away condition
3. As a result of repeating this 10,000 times, the average was 83.95 and the SD was 1030.10
4. The Confidence interval is from 63.76 to 104.14, so the expected value of 76.35 is within the CI range

PART 3-4. A Frequency Distribution for Y

Y	Expected. Freq.	Observed. Freq.	Chi-Squared	
-1560	912	936	0.631578947	Chi-squared Metric
-1060	1951.7	1888	2.077770126	7.26
-560	1527.4	1559	0.654768293	p-value
460	2291.1	2252	0.666064368	0.20
980	1769.9	1749	0.246329921	
1500	1548	1616	2.987080103	



Chi-squared goodness of Fit Test

Step 0 Assumption

Distribution about categorical variable which are -1560, -1060, -560, 460, 980, and 1500. Sample was randomly selected, and over 900 observations for each group

Step 1 Hypothesis.

Null: The population follows the specified distribution which is that -1560, -1060, -560, 460, 980 and 1500 is distributed as 9.12%, 19.52%, 15.27%, 22.91%, 17.7%, and 15.48%
Alternative: The population does not follow the specified distribution

Step 2 Find the critical value.

- The p-value is $\alpha = 0.05$

Step 3 Compute the test value.

- The p-value with 1-CHISQ.DIST() in excel is 0.20,
- Because CHISQ.DIST() calculate the area on the left. Chi-squared test is always right area test, so I have to minus chi-squared from the one

Step 4 Make the decision.

- There is not enough evidence to reject null hypothesis
  since 0.20 (p-value) > 0.05

Step 5 Summarize the results.

- There is not enough evidence to reject the claim that the population follows the specified distribution, which is that -1560, -1060, -560, 460, 980 and 1500 is distributed as 9.12%, 19.52%, 15.27%, 22.91%, 17.7%, and 15.48%

Interpretation

1. Theoretically, applying the probabilities, when X= -1560, -1060, -560, 460, 980, and 1500, the numbers are 912, 1952, 1527, 2291, 1770, and 1548 out of 10,000
2. According to our randomly generated simulations, the Observed Frequencies are 936, 1888, 1559, 2252, 1749, and 1616
3. The metric calculated by chi-squared is 7.26 and the p-value is 0.20
4. When Df is 5 in 0.05 alpha, the critical value of chi-squared table is 11.070. In this case the chi-squared metric is 7.26, I cannot reject the null hypothesis and also p-value is greater than 0.05

PART 3-5. My betting Strategy

Strategy Analysis

1. In common sense, if I had to choose whether or not to bet, I would of course choose to bet only on home matches and not bet on away matches
2. This means betting on the first, third and last match, not betting on the second and forth matches
3. If I could bet on just one team in the series not in each game, I would have to bet on Boston and the resulting expected value would be 76.35. In sampling, when simulating 10,000 times, the mean is 83.95, which justifies my strategy because the expected value in the between confidence interval [63.76, 104.14].

PART 4. Answers for Questions

PART 1

1. Calculate the probability that the Red Sox will win the series
   The probability that the Red Sox win the series, in the three games which are held Boston, Yankees, and Boston, is 0.5664.

2. 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.

Probability distribution for your net win (X)				
X	-1040	-540	480	1000
P(X)	0.228	0.2056	0.3084	0.258
Mean of X	57.888		SD	795.15

3. Estimate your expected net win by using a 95% confidence interval. Does this confidence interval contain E(X)?
   The Confidence interval is 39.34 and 70.57, so it contains E(X) which is 57.89.

4. 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.
   The p-value of Chi-squared goodness of Fit Test is 0.87, so there is not enough evidence to reject the claim that the population follows the specified distribution, which is that -1040, -540, 480 and 1000 is distributed as 22.8%, 20.6%, 30.8%, and 25.8%. The the distribution of Y is similar to the estimated distribution of X.

Frequency distribution for Y				
Y	-1040	-540	480	1000
Observed	2293	2071	3045	2591
Expected	2280	2056	3084	2580

5. Describe whether your betting strategy is favorable to you
   Boston's betting strategy is favorable. Because by making that choice, I have an expected value of 57.89. Also, according to what was confirmed by simulation, the expected value exists within the confidence interval [39.34, 70.57]. This supports my strategy.

PART 2

1. Calculate the probability that the Red Sox will win the series
   The probability that the Red Sox win the series, in the three games which are held Yankees, Boston, and Yankees, is 0.4790. This is 0.0874 lower than when playing

2 games in Boston home. This also brings the overall expected value down to less than 50%.

2. 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.

Probability distribution for your net win (X)				
X	-1040	-540	480	1000
P(X)	0.228	0.2930	0.2210	0.258
Mean of X	-31.24		SD	799.99

3. Estimate your expected net win by using a 95% confidence interval. Does this confidence interval contain E(X)?
   The Confidence interval is from -55.81 to -24.47, so it contains E(X) which is -31.24.

4. 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.
   The p-value of Chi-squared goodness of Fit Test is 0.48, so there is not enough evidence to reject the claim that the population follows the specified distribution, which is that -1040, -540, 480 and 1000 is distributed as 22.8%, 29.3%, 22.1%, and 25.8%. The the distribution of Y is similar to the estimated distribution of X.

Frequency distribution for Y				
Y	-1040	-540	480	1000
Observed	2330	2915	2229	2526
Expected	2280	2929.8	2210.2	2580

5. Describe whether your betting strategy is favorable to you
   Boston's betting strategy is not favorable. Because by making that choice, I can lose 31.24 which is expected value. Also, according to what was confirmed by simulation, the expected value exists within the confidence interval [-55.81, -24.47]. This means that the Boston betting strategy is not favorable.

PART 3

1. Calculate the probability that the Red Sox will win the series
   The probability that the Red Sox win the series, in the five games which are held Boston, Yankees, Boston, Yankees, and Boston is 0.5609.

2. 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.

Probability distribution for your net win (X)

X	-1560	-1060	-560	460	980	1500
P(X)	0.0912	0.1952	0.1527	0.2291	0.1770	0.1548
Mean of X	76.35				SD	1024.73

3. Estimate your expected net win by using a 95% confidence interval. Does this confidence interval contain E(X)?
   The Confidence interval is from 63.76 to 104.14, so it contains E(X) which is 76.35.
4. 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.
   The p-value of Chi-squared goodness of Fit Test is 0.20, so there is not enough evidence to reject the claim that the population follows the specified distribution, which is that -1560, -1060, -560, 460, 980 and 1500 is distributed as 9.12%, 19.52%, 15.27%, 22.91%, 17.7%, and 15.48%. The the distribution of Y is similar to the estimated distribution of X.

Frequency distribution for Y

Y	-1560	-1060	-560	460	980	1500
Observed	936	1888	1559	2252	1749	1616
Expected	912	1951.7	1527.4	2291.1	1769.9	1548

5. Describe whether your betting strategy is favorable to you
   Boston's betting strategy is favorable. Because by making that choice, I have an expected value of 76.35. Also, according to what was confirmed by simulation, the expected value exists within the confidence interval [63.76, 104.14]. This supports my strategy.

PART 5. CONCLUSION

In this project, I learned how to make decisions using Excel. The most important thing to solve the problem is to find out what's exact problem. Based on this, I gather the necessary information and do analysis. Finally, make decisions based on that.

In this betting example, our task is determining whether the betting strategy is favorable or not. I have a strategy that bet on Boston to win. The information which is needed is Boston's odds of winning at each stadium. Based on this, I create a Probability distribution for my win (X). Then, calculate the mean and standard deviation and after simulation with 10,000 samples, they were compared with the expected values. Based on the confidence interval and frequency distribution, I finally draw a conclusion about my strategy.

REFERENCE

ExtendOffice. (2018). How to generate random value based on assigned probability in Excel. Retrieved from https://www.extendoffice.com/documents/excel/4486-excel-random-with-probability.html

LibreTexts. (September 11, 2021). 5.3: Expectation, variance and standard deviation. Retrieved from https://stats.libretexts.org/Courses/City_University_of_New_York/Introductory_Statistics_with_Probability_(CUNY)/05%3A_Discrete_Random_Variables/5.03%3A_Expectation_Variance_and_Standard_Deviation

CFI Team. (November 29, 2022). Confidence function. Retrieved from https://corporatefinanceinstitute.com/resources/excel/confidence-interval-excel-function/

Diana, M. & Phoebe, B. (n.d.). The chi square test. Retrieved from https://www.westga.edu/academics/research/vrc/assets/docs/ChiSquareTest_LectureNotes.pdf

Shaun Turney. (May 24, 2022). Chi-square goodness of fit test | formula, guide & examples. Scribbr. Retrieved from https://www.scribbr.com/statistics/chi-square-goodness-of-fit/