**Report on Analysis of Betting Strategy in Sports**

Logo

Description automatically generated

**Northeastern University, Boston**

**College of Professional Studies**

**Class:** ALY6050-21598

**Professor:** Richards He

**Student:** Durga Bhanu Nayak

**NUID:** 001532778

**Date of Submission:** 01/16/2022

# **PART 1**

Probability that Boston Red Sox **wins** a game in **Boston** = p = 0.59

Probability that Boston Red Sox **loses** a game in **Boston** = 1 – p = 0.41

Probability that Boston Red Sox **loses** a game in **New York** = q = 0.56

Probability that Boston Red Sox **wins** a game in **New York** = 1 – q = 0.44

1. **Probability that Red Sox team will win the best of three series where first match is played in New York**

Possible match scenarios for Red Sox:

Table

Description automatically generated

To win the best of three series **Red Sox** team needs to win at least **2** matches and can afford to lose at most **one** match.

Hence, the winning probability can be calculated as:

Table

Description automatically generated

Probability of winning first two matches consecutively + probability of winning first and third match and losing the second one + probability of losing first match and winning the rest of the two matches

= (0.44)\*(0.59) **+** (0.44)\*(0.41)\*(0.44) **+** (0.56)\*(0.59)\*(0.44)

= 0.484352

So, Boston Red Sox is having **48.44%** chances that it will win the series if the first match is played in New York.

1. **Probability Distribution for the calculation of expected net win and standard deviation**

Table

Description automatically generated

Here, X is the net amount calculated in of the series scenario played between Boston Red Sox and New York Yankees.

Using the above probability distribution table, we were able to calculate the expected net win and the standard deviation.

Text

Description automatically generated

E(X) = µ = -64.10

Std. Dev, σ = 1611.14

1. **Creating random variable for X**

Using MS Excel, we have created 10,000 random values for our X and named the values as Y. To create the random values, we have used RAND() function and chose the values among the X values.

Text, letter

Description automatically generated

From the generated Y values, we have tried to calculate the expected net win which is -59.44.

Here, we are getting the confidence interval of (-90.90, -27.98).

Our calculated E(X) = -64.10 lies between the limits of the confidence interval of 95%.

\*Note: The random values are subjected to change while working with Excel workbook.

1. **Checking for the data distribution using Chi-squared goodness of fit test**

Constructing a frequency distribution for Expected values (Theoretical values) and Observed values (from Random values Y) to compare the

Table, Excel

Description automatically generated

Here, we can observe,

ρ1 = Probability Distribution for Expected Values

ρ2 = Probability Distribution for Random Values

H0: ρ1 = ρ2

H1: ρ1 ≠ ρ2

Degree of Freedom = 3

Significant value, α = 0.05

Chi-Squared metric =

= 2.54

p-value = 0.46

As the p-value is greater than the significance value, α, we will not be rejecting our Null Hypothesis, H­0aswe are not having enough evidence against it. Hence, we can conclude that, the random values distribution, Y is very close to the expected values.

Also, we are 95% confident that our expected mean is present between the interval of (-90.90, -27.98).

1. **Betting Strategy**

From the above analysis, I can see that 95% chances are there that if I try to bet in this series, the odds will not be in my favor as my average winning value will lie between -90.9 and -27.98 which is a loss.

Hence, I will try to refrain from this betting strategy.

# **PART 2**

1. **Probability that Red Sox team will win the best of three series where first match is played in Boston**

Possible match scenarios for Red Sox:

Table

Description automatically generated

To win the best of three series **Red Sox** team needs to win at least **2** matches and can afford to lose at most **one** match.

Hence, the winning probability can be calculated as:

Table

Description automatically generated

probability of winning first two matches consecutively + probability of winning first and third match and losing the second one + probability of losing first match and winning the rest of the two matches

= (0.59)\*(0.44) **+** (0.59)\*(0.56)\*(0.59) **+** (0.41)\*(0.44)\*(0.59)

= 0.5610

So, Boston Red Sox is having **56.1%** chance that it will win the series if the first match is played in their home ground and so on.

1. **Probability Distribution for the calculation of expected net win and standard deviation**

Table

Description automatically generated

Here, X is the net amount calculated in of the series scenario played between Boston Red Sox and New York Yankees.

Using the above probability distribution table, we were able to calculate the expected net win and the standard deviation.

Text

Description automatically generated

E(X) = µ = 92.97

Std. Dev, σ = 1602.4

1. **Creating random variable for X**

Using MS Excel, we have created 10,000 random values for our X and named the values as Y. To create the random values, we have used RAND() function and chose the values among the X values.

Text

Description automatically generated

From the generated Y values, we have tried to calculate the expected net win which is 95.87.

Here, we are getting the confidence interval of (64.41, 127.33).

Our calculated E(X) = 92.97 **lies between** the limits of the confidence interval of 95%.

\*Note: The random values are subjected to change while working with Excel workbook.

1. **Checking for the data distribution using Chi-squared goodness of fit test**

Constructing a frequency distribution for Expected values (Theoretical values) and Observed values (from Random values Y) to compare the

Table

Description automatically generated

Here, we can observe,

ρ1 = Probability Distribution for Expected Values

ρ2 = Probability Distribution for Random Values

H0: ρ1 = ρ2

H1: ρ1 ≠ ρ2

Degree of Freedom = 3

Significant value, α = 0.05

Chi-Squared metric =

= 1.25

p-value = 0.7391

As the p-value is greater than the significance value, α, we will not be rejecting our Null Hypothesis, H­0aswe are not having enough evidence against it. Hence, we can conclude that, the random values distribution, Y is very close to the expected values.

Also, we are 95% confident that our expected mean is present between the interval of (64.41, 127.33).

1. **Betting Strategy**

From the above analysis, as I can see my odds of winning are high as I am 95% confident that my average winning will lie in the interval of 64.41 and 127.33.

Hence, I will try to take my chances in this series and would bet on Boston Red Sox.

# **PART 3**

1. **Probability that Red Sox team will win the best of five series where first match is played in New York**

Possible match scenarios for Red Sox:

Table

Description automatically generated

To win the best of five series **Red Sox** team needs to win at least **3** matches and can afford to lose at most **two** matches.

Probability grid for each win and lose scenario based on the matches happened:

Table

Description automatically generated

Hence, the winning probability of the series can be calculated as the winning scenarios highlighted in the above picture.

= 0.4999

So, Boston Red Sox is having **49.99%** chances that it will win the best of five series if the first match is played in New York and so on.

1. **Probability Distribution for the calculation of expected net win and standard deviation**

A screenshot of a computer

Description automatically generated with medium confidence

Here, X is the net amount calculated in of the series scenario played between Boston Red Sox and Ney York Yankees.

Using the above probability distribution table, we were able to calculate the expected net win and the standard deviation.

Text

Description automatically generated

E(X) = µ = -71.98

Std. Dev., σ = 2068.72

1. **Creating random variable for X**

Using MS Excel, we have created 10,000 random values for our X and named the values as Y. To create the random values, we have used RAND() function and chose the values among the X values.

Text, letter

Description automatically generated

From the generated Y values, we have tried to calculate the expected net win which is -69.75.

Here, we are getting the confidence interval of (-110.11, -29.38).

Our calculated E(X) = -71.98 lies between the limits of the confidence interval of 95%.

\*Note: The random values are subjected to change while working with Excel workbook.

1. **Checking for the data distribution using Chi-squared goodness of fit test**

Constructing a frequency distribution for Expected values (Theoretical values) and Observed values (from Random values Y) to compare the

Table, Excel

Description automatically generated

Here, we can observe,

ρ1 = Probability Distribution for Expected Values

ρ2 = Probability Distribution for Random Values

H0: ρ1 = ρ2

H1: ρ1 ≠ ρ2

Degree of Freedom = 5

Significant value, α = 0.05

Chi-Squared metric =

= 7.956

p-value = 0.1587

As the p-value is greater than the significance value, α, we will not be rejecting our Null Hypothesis, H­0aswe are not having enough evidence against it. Hence, we can conclude that, the random values distribution, Y is very close to the expected values.

Also, we are 95% confident that our expected mean is present between the interval of (-110.11, -29.38).

1. **Betting Strategy**

From the above analysis, I can see that 95% chances are there that if I try to bet in this series, the odds will not be in my favor as my average winning value will lie between -110.11 and -29.38 which is a loss.

Hence, I will try to refrain from this betting strategy.

# **PART 4**

1. **Probability that Red Sox team will win the best of seven series where first match is played in Boston**

Possible match scenarios for Red Sox:

Table

Description automatically generated

To win the best of seven series **Red Sox** team needs to win at least **4** matches and can afford to lose at most **three** matches.

Probability grid for each win and lose scenario based on the matches happened:

Table

Description automatically generated

Hence, the winning probability of the series can be calculated as the winning scenarios highlighted in the above picture.

= 0.5568

So, Boston Red Sox is having **55.68%** chances that it will win the best of seven series if the first match is played in Boston and so on.

1. **Probability Distribution for the calculation of expected net win and standard deviation**

Table

Description automatically generated

Here, X is the net amount calculated in of the series scenario played between Boston Red Sox and New York Yankees.

Using the above probability distribution table, we were able to calculate the expected net win and the standard deviation.

Text

Description automatically generated with low confidence

E(X) = µ = 120.90

Std. Dev, σ = 2446.98

1. **Creating random variable for X**

Using MS Excel, we have created 10,000 random values for our X and named the values as Y. To create the random values, we have used RAND() function and chose the values among the X values.

Text, letter

Description automatically generated

From the generated Y values, we have tried to calculate the expected net win which is 107.385.

Here, we are getting the confidence interval of (59.54, 155.22).

Our calculated E(X) = 120.90 lies between the limits of the confidence interval of 95%.

\*Note: The random values are subjected to change while working with Excel workbook.

1. **Checking for the data distribution using Chi-squared goodness of fit test**

Constructing a frequency distribution for Expected values (Theoretical values) and Observed values (from Random values Y) to compare the

Table

Description automatically generated

Here, we can observe,

ρ1 = Probability Distribution for Expected Values

ρ2 = Probability Distribution for Random Values

H0: ρ1 = ρ2

H1: ρ1 ≠ ρ2

Degree of Freedom = 7

Significant value, α = 0.05

Chi-Squared metric =

= 2.25

p-value = 0.9444

As the p-value is greater than the significance value, α, we will not be rejecting our Null Hypothesis, H­0aswe are not having enough evidence against it. Hence, we can conclude that, the random values distribution, Y is very close to the expected values.

Also, we are 95% confident that our expected mean is present between the interval of (59.54, 155.22).

1. **Betting Strategy**

From the above analysis, as I can see my odds of winning are high as I am 95% confident that my average winning will lie in the interval of 59.54 and 155.22

Hence, I will try to take my chances in this series and would bet on Boston Red Sox.

# **SUMMARY**

* From the above game scenario, we have observed, for the team Boston Red Sox, if matches are first played in New York, the series winning probabilities are lower than 50%. And, if the series starts with the first match in Boston, then Red Sox are having more than 50% chances that they will win the series.
* For best of three series, possible net amount values are four i.e. (2000, 950, -1100, -2100).
* For best of five series, possible net amount values are six i.e. (3000, 1950, 900, -1150, -2150, -3150).
* For best of seven series, possible net amount values are eight i.e. (4000, 2950, 1900, 850, -4200, -3200, -2200, -1200).
* Using Chi-squared test we validated that our data distribution of random values are similar to the data distribution of our expected values calculated using respected probabilities.
* For series where matches are starting from New York, we are 95% confident that our net win amount will lie in negative values which is not favorable to us. For such series, I will not choose to place my bets on Boston Red Sox team.
* For series where matches are starting from Boston, we are 95% confident that our net win amount will lie among positive values. And even though the variation in the data is somewhat high, I would like to take the risk and place my bets on Boston Red Sox for these series.

# **REFERENCES**

* Guttman, I., Jayalath, K. P., & Gupta, B. C. (2020). Statistics and Probability with Applications for Engineers and Scientists Using MINITAB, R and JMP (2nd ed.). Wiley.
* Evans, J. (2012). Statistics, Data Analysis, and Decision Modeling (5th ed.). Pearson.
* Bluman, A. G. (2018). Elementary statistics: A step by step approach. New York, NY: McGraw-Hill Education.