Graphical user interface, text

Description automatically generated with medium confidence

Logo

Description automatically generated

**Abhilash Dikshit**

ALY6050: Analysis of a Betting Strategy in Sports

Instructor: Soheil Parsa

Week 1: R Practice

Date: 2023/03/03

**Problem Statement:**

Suppose that Boston Red Sox and New York Yankees (two American League baseball teams) are scheduled to play a best of three series. The winner of the series will be the first team that wins two of the three games. The probability that the Red Sox win a game in their home stadium is 0.6 and the probability that Yankees win their home game is 0.57. Next, suppose that you place a bet on each game played where you win $500 if the Red Sox win and you lose $520 if the Red Sox lose the game.

Assume that the game results in sections 1-3 below are unrelated to one another.

**Part A:**

Set the odds of the Red Sox and Yankees winning at home as stated in the problem statement first.

Let's now determine the odds of each team winning the series.

A picture containing graphical user interface

Description automatically generated

1. Calculate the probability that the Red Sox will win the series.

Graphical user interface, text, application

Description automatically generated

1. 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.

Graphical user interface, text

Description automatically generated

Graphical user interface, text, application

Description automatically generated

1. Generate 10,000 random values of X.

Graphical user interface, text, application

Description automatically generated

1. 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.

Graphical user interface, text, application

Description automatically generated

1. Use your observations of parts (ii) and (iii) above to describe whether your betting strategy is favorable to you. Write a summary of your observations and analyses in the Word document.

The predicted net win indicates that the betting strategy is not in your favour because it is negative, according to the expected net win and the standard deviation calculated in part (ii). Your actual net win might, however, be positive based on the 95% confidence interval derived in part (iii), however this cannot be said with absolute certainty.

The distribution of Y does not closely approximate the distribution of X, according to the Chi-squared goodness of fit test carried out in part (iv), which raises the possibility that the simulation may not accurately reflect the real probabilities of the outcomes. In general, it's vital to exercise caution and be aware of the hazards while placing bets.

**PART B:**

1. Calculate the probability of the Red Sox winning the series.

Graphical user interface, text, application

Description automatically generated

1. Print the expected net win and the standard deviation.

Graphical user interface, text, application, chat or text message

Description automatically generated

1. Calculate the 95% confidence interval for the expected net win.

Graphical user interface, text, application, chat or text message

Description automatically generated

1. Create a frequency distribution for Y and perform the Chi-squared goodness of fit test.

Graphical user interface, text, application, email

Description automatically generated

1. Use your observations of parts (ii) and (iii) above to describe whether your betting  
   strategy is favorable to you. Write a summary of your observations and analyses in the Word document.

The predicted net win indicates that the betting strategy is not in your favour because it is negative, according to the expected net win and the standard deviation calculated in part (ii). Your actual net win might, however, be positive based on the 95% confidence interval derived in part (iii), however this cannot be said with absolute certainty.

The distribution of Y does not closely approximate the distribution of X, according to the Chi-squared goodness of fit test carried out in part (iv), which raises the possibility that the simulation may not accurately reflect the real probabilities of the outcomes. In general, it's vital to exercise caution and be aware of the hazards while placing bets.

PART C:

Repeat part 1 above but now assume that the series is a best-of-five series where the first team that wins three games wins the series with games alternating between Boston and New York, with the first game being played in Boston.

Create a matrix to store the probabilities of winning for each team in each game.

Text, letter

Description automatically generated

## **Aspects of assignment:**

Part 1:

With the first game taking place in Boston, the Boston Red Sox have a about 0.4284 chance of defeating the New York Yankees in a best-of-three series. Based on the results of each game, a betting strategy's projected net win would be roughly -$78.13, showing that the approach is unfavourable. With a standard deviation of $357.934, the net win distribution is biased to the left. According to the findings of our simulation, there is a 95% chance that the predicted net win will be between -$74.41 and -$60.35.

Part 2:

A best-of-three series between the Boston Red Sox and the New York Yankees, with the first game taking place in New York, has a probability of the Boston Red Sox winning of roughly 0.4240. A betting strategy based on the results of each game would forecast a net win of roughly -$82.55, which would make the strategy quite unfavourable. The standard deviation of the net win distribution is $357.537, which is biased to the left. According to the findings of our simulation, there is a 95% chance that the actual net win will be between -78.49 and -64.45 dollars.

Part 3:

The likelihood of the Boston Red Sox winning a best-of-five series against the New York Yankees, with the first game being held in Boston and games alternating between Boston and New York, is around 0.60. If a betting strategy is based on the outcomes of each game, the expected net win is approximately -$53.85, indicating that the strategy is unfavorable. The distribution of net wins is skewed to the left, with a standard deviation of $506.01. Our simulation results indicate that we can be 95% confident that the expected net win falls between -$43.93 and -$63.77.

Taken together, these findings imply that the Boston Red Sox are more likely to win the series in each situation, but a betting strategy based on the outcomes of each game is unlikely to be profitable over the long run.

**References:**

* Bevans ([2022, November 11](http://127.0.0.1:37113/rmd_output/2/#ref-R-Career));Datanovia ([2019, December 26](http://127.0.0.1:37113/rmd_output/2/#ref-R-Action));Linear Regression Example in r Using Lm() Function ([n.d.](http://127.0.0.1:37113/rmd_output/2/#ref-R-Cran));Zach ([2021, September 29](http://127.0.0.1:37113/rmd_output/2/#ref-R-Material1))
* Bevans, R. 2022, November 11. Hypothesis Testing | a Step-by-Step Guide with Easy Examples. <https://mathstat.slu.edu/~speegle/_book/RData.html>.
* Datanovia. 2019, December 26. How to Do a t-Test in r: Calculation and Reporting. <https://www.datanovia.com/en/lessons/how-to-do-a-t-test-in-r-calculation-and-reporting/>.
* Linear Regression Example in r Using Lm() Function. n.d. <https://www.learnbymarketing.com/tutorials/linear-regression-in-r/>.
* Zach, Z. 2021, September 29. How to Perform Logistic Regression in r (Step-by-Step). <https://www.statology.org/logistic-regression-in-r/>.

## **Appendix:**

prob\_rs\_win\_home = 0.6

prob\_nyy\_win\_home = 0.57

prob\_rs\_win\_series = prob\_rs\_win\_home^2 \* (1 - prob\_nyy\_win\_home) + prob\_rs\_win\_home \* (1 - prob\_rs\_win\_home) \* prob\_nyy\_win\_home \* 2

cat("Probability of Red Sox winning the series:", prob\_rs\_win\_series, "\n")

outcomes =c(-520, 0, 500)

probabilities =c((1 - prob\_rs\_win\_series)^2, 2 \* prob\_rs\_win\_series \* (1 - prob\_rs\_win\_series), prob\_rs\_win\_series^2)

cat("Probabilities:", probabilities)

mean\_X =sum(outcomes \* probabilities)

variance\_X =sum((outcomes - mean\_X)^2 \* probabilities)

sd\_X =sqrt(variance\_X)

cat("Expected net win:", mean\_X, "\n")

cat("Standard deviation of net win:", sd\_X, "\n")

set.seed(1)

Y =sample(outcomes, 10000, replace = TRUE, prob = probabilities)

*# Calculate the 95% confidence interval for the expected net win*

lower\_ci =mean(Y) - qt(0.975, df = length(Y) - 1) \* sd(Y) / sqrt(length(Y))

upper\_ci =mean(Y) + qt(0.975, df = length(Y) - 1) \* sd(Y) / sqrt(length(Y))

cat("95% Confidence interval for expected net win:", lower\_ci, "-", upper\_ci, "\n")

**if** (lower\_ci <= mean\_X & mean\_X <= upper\_ci) {

cat("The confidence interval contains E(X).\n")

} **else** {

cat("The confidence interval does not contain E(X).\n")

}

freq\_table =table(Y)

expected\_counts =length(Y) \* probabilities

chisq\_test =chisq.test(freq\_table, p = expected\_counts, rescale.p = TRUE)

*# Print the Chi-squared test results*

cat("Chi-squared test results:\n")

print(chisq\_test)

cat("The predicted net win indicates that the betting strategy is not in your favour because it is negative, according to the expected net win and the standard deviation calculated in part (ii). Your actual net win might, however, be positive based on the 95% confidence interval derived in part (iii), however this cannot be said with absolute certainty.

The distribution of Y does not closely approximate the distribution of X, according to the Chi-squared goodness of fit test carried out in part (iv), which raises the possibility that the simulation may not accurately reflect the real probabilities of the outcomes. In general, it's vital to exercise caution and be aware of the hazards while placing bets.")

prob\_rs\_win\_series =prob\_nyy\_win\_home^2 \* (1 - prob\_rs\_win\_home) + prob\_nyy\_win\_home \* (1 - prob\_nyy\_win\_home) \* prob\_rs\_win\_home \* 2

cat("Probability of Red Sox winning series:", prob\_rs\_win\_series, "\n")

*# Calculate the possible outcomes and their probabilities*

outcomes =c(-520, 0, 500)

probabilities =c((1 - prob\_rs\_win\_series)^2, 2 \* prob\_rs\_win\_series \* (1 - prob\_rs\_win\_series), prob\_rs\_win\_series^2)

*# Calculate the expected net win and the standard deviation*

mean\_X =sum(outcomes \* probabilities)

variance\_X =sum((outcomes - mean\_X)^2 \* probabilities)

sd\_X =sqrt(variance\_X)

cat("Expected net win:", mean\_X, "\n")

cat("Standard deviation of net win:", sd\_X, "\n")

*# Generate 10,000 random values of X*

set.seed(1)

Y =sample(outcomes, 10000, replace = TRUE, prob = probabilities)

lower\_ci =mean(Y) - qt(0.975, df = length(Y) - 1) \* sd(Y) / sqrt(length(Y))

upper\_ci =mean(Y) + qt(0.975, df = length(Y) - 1) \* sd(Y) / sqrt(length(Y))

cat("95% Confidence interval for expected net win:", lower\_ci, "-", upper\_ci, "\n")

check\_confidence\_interval <- **function**(lower\_ci, upper\_ci, mean\_X) {

**if** (lower\_ci <= mean\_X & mean\_X <= upper\_ci) {

**return**("The confidence interval contains E(X).")

} **else** {

**return**("The confidence interval does not contain E(X).")

}

}

freq\_table =table(Y)

expected\_counts =length(Y) \* probabilities

chisq\_test =chisq.test(freq\_table, p = expected\_counts, rescale.p = TRUE)

*# Print the Chi-squared test results*

cat("Chi-squared test results:\n")

print(chisq\_test)

cat("The predicted net win indicates that the betting strategy is not in your favour because it is negative, according to the expected net win and the standard deviation calculated in part (ii). Your actual net win might, however, be positive based on the 95% confidence interval derived in part (iii), however this cannot be said with absolute certainty.

The distribution of Y does not closely approximate the distribution of X, according to the Chi-squared goodness of fit test carried out in part (iv), which raises the possibility that the simulation may not accurately reflect the real probabilities of the outcomes. In general, it's vital to exercise caution and be aware of the hazards while placing bets.")

*# rows represent Red Sox and columns represent Yankees*

prob\_matrix =matrix(nrow = 5, ncol = 2)

prob\_matrix[1, ] =c(prob\_rs\_win\_home, 1 - prob\_rs\_win\_home)

prob\_matrix[2, ] =c(1 - prob\_nyy\_win\_home, prob\_nyy\_win\_home)

prob\_matrix[3, ] =c(prob\_rs\_win\_home, 1 - prob\_rs\_win\_home)

prob\_matrix[4, ] =c(1 - prob\_nyy\_win\_home, prob\_nyy\_win\_home)

prob\_matrix[5, ] =c(prob\_rs\_win\_home, 1 - prob\_rs\_win\_home)

simulate\_series <- **function**() {

red\_sox\_wins <- 0

yankees\_wins <- 0

net\_win <- 0

**for** (i **in** 1:5) {

*# Sample the winner of the game based on the probabilities*

winner <- sample(c("Red Sox", "Yankees"), size = 1, prob = prob\_matrix[i, ])

*# Add a win to the winner's count*

**if** (winner == "Red Sox") {

red\_sox\_wins <- red\_sox\_wins + 1

} **else** {

yankees\_wins <- yankees\_wins + 1

}

*# If one team has won 3 games, end the series and calculate the net win*

**if** (red\_sox\_wins == 3) {

net\_win <- 500

**break**

} **else** **if** (yankees\_wins == 3) {

net\_win <- -520

**break**

}

}

**return**(net\_win)

}

set.seed(1)

net\_wins =replicate(10000, simulate\_series())

mean\_net\_win =mean(net\_wins)

sd\_net\_win =sd(net\_wins)

cat("Expected net win:", mean\_net\_win, "\n")

cat("Standard deviation of net win:", sd\_net\_win, "\n")

conf\_interval = t.test(net\_wins)$conf.int

cat("95% confidence interval for expected net win:", conf\_interval, "\n")

check\_confidence\_interval <- **function**(mean\_net\_win, conf\_interval) {

**if** (mean\_net\_win >= conf\_interval[1] && mean\_net\_win <= conf\_interval[2]) {

**return**("The confidence interval contains the expected net win.")

} **else** {

**return**("The confidence interval does not contain the expected net win.")

}

}

freq\_table =table(net\_wins)

freq\_dist =data.frame(net\_win = as.numeric(names(freq\_table)), count = as.numeric(freq\_table))

options(scipen = 100)

chisq.test(freq\_dist$count, p = dnorm(freq\_dist$net\_win, mean = mean\_net\_win, sd = mean\_net\_win), rescale.p = TRUE)

*## NA*