**ALY6050 80478 Intro to Enterprise Analytics SEC 09 Spring 2023 CPS**

**Module 1 Assignment — Analysis of a Betting Strategy in Sports REPORT**

Logo

Description automatically generated

**NORTHEASTERN UNIVERSITY**

**College of Professional Studies, Boston, MA, 02215.**

**Submitted by**

Zihan Ma

ma.zihan1@northeastern.edu

**Instructor**

 Prof. Richard He

**Date**

05/30/2023

**Analysis of a Betting Strategy in Sports**

**Assignment Summary:**

The assignment involves analyzing a betting strategy in a best-of-three or more series between the Boston Celtics and the Miami Heat in the NBA. It includes four parts that need to calculate probabilities and construct probability distributions.

**Problem 1:**

Calculate the probability that the Boston Celtics will win the best-of-three series against the Miami Heat.

1. Calculate the probability that the Boston Celtics will win the series  
     
   There are three situations for Celtics to win:  
   Winning the first two games (WW)

Winning the first and third games (WLW)

Losing the first game and winning the next two (LWW)  
So, P(Celtics Win Series) = p(C) \* p(C) + p(C) \* p(H) \* p(C) + p(H) \* p(C) \* p(C)  


1. Construct a probability distribution for your net win (X) in the series. Calculate your expected net win (E(X)) and the standard deviation of X.  
     
   All possible results are:  
   Celtics win the series   
   (net win: 2 \* bet\_celtics\_win)

Celtics lose the series   
(net win: 2 \* bet\_celtics\_lose)

Series goes to the third game and Celtics win  
 (net win: bet\_celtics\_win + bet\_celtics\_lose)

Series goes to the third game and Celtics lose

(net win: bet\_celtics\_lose + bet\_celtics\_win)

Celtics win the first two games and the series ends   
(net win: bet\_celtics\_win + bet\_celtics\_win)

Celtics lose the first two games and the series ends   
(net win: bet\_celtics\_lose + bet\_celtics\_lose)  
  
E(X) = (2 \* bet\_celtics\_win \* P1) + (2 \* bet\_celtics\_lose \* P2) + (bet\_celtics\_win + bet\_celtics\_lose) \* P3 + (bet\_celtics\_lose + bet\_celtics\_win) \* P4 + (bet\_celtics\_win + bet\_celtics\_win) \* P5 + (bet\_celtics\_lose + bet\_celtics\_lose) \* P6  
  
SD(X) = sqrt( ( (X1 - E(X))^2 \* P1 )+... + ( (Xn - E(X))^2 \* Pn ) )  


1. Use Excel or R to create 5,000 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 the E(X) in (ii)?  
     
     
   The result is false, which indicates that the betting outcomes may deviate from what was initially anticipated. The observed average net win from the random values differs significantly from the expected net win.
2. Construct a frequency distribution for Y. Next, use the Chi-square goodness of fit test to verify how closely the distribution of Y has estimated the distribution of X.  
     
   A black text on a white background

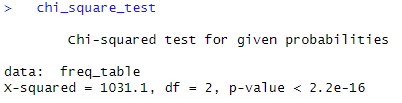
   Description automatically generated with low confidence  
   The result of the Chi-square goodness-of-fit test indicates a significant difference between the observed frequencies in the frequency distribution of Y and the expected frequencies based on the distribution of X. The p-value, which is less than 0.05, suggests strong evidence against the similarity of the two distributions.
3. Use your observations in 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.  
     
   Based on the observations and analyses in parts (ii) and (iii), the assessment of the betting strategy is as follows:

The expected net win (E(X)) provides an estimate of the average outcome, while the standard deviation indicates the level of risk involved. The estimated confidence interval is used to assess the accuracy of our estimation.

If the estimated confidence interval contains the E(X) calculated in part (ii), it suggests that the betting strategy is performing as expected. However, the interval we find does not contain the E(X), which means this strategy has very high risk.

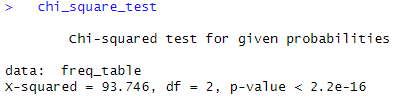
**Problem 2:**

Repeat the analysis from Problem 1, but with a different game order where the first game is played in Boston, the second game is played in Miami, and the third game (if necessary) is played in Boston.  
The formulas and calculations used in Problem 2, 3, 4 are similar to those in Problem 1, with the main difference being the probabilities and game order considered. The formulas and calculations serve the same purpose of determining the probability of winning, expected net win, and standard deviation of net win.

1.   
   The probability of the Boston Celtics winning the series with the modified game order (Boston, Miami, Boston) is calculated as 0.146072. This probability differs from the previous calculation because the game order affects the likelihood of winning.
2.   
     
   With the updated probabilities and game order, the expected net win is now estimated as 35.35832, representing the average net win from the betting strategy. The standard deviation is calculated as 274.2661, reflecting the variability or risk associated with the net win.
3.   
   By generating 5,000 random values for X (denoted as Y) based on the modified probabilities and game order, we estimate the mean of Y (mean\_Y) and calculate the standard error (se). The resulting 95% confidence interval is computed, and in this case, it contains the expected net win (E\_X\_q2ii), indicated by the value TRUE. This suggests that the estimated average net win falls within the confidence interval.
4.   
   Constructing a frequency distribution for Y and performing the Chi-square goodness-of-fit test shows that the observed frequencies differ significantly from the expected frequencies based on the distribution of X. The X-squared value is 1031.1, indicating a substantial deviation from the expected distribution. The extremely small p-value (< 2.2e-16) further supports the evidence against the similarity of the observed and expected frequencies.
5. With the revised game order, the expected net win and confidence interval have changed. The betting strategy appears to have a positive expected net win, but the substantial standard deviation and the Chi-square goodness-of-fit test results suggest a notable risk and a deviation from the expected distribution of net wins.

**Problem 3:**

Repeat the analysis from Problem 1, but with a best-of-five series where the first team to win three games wins the series, with games alternating between Boston and Miami, and the first game being played in Miami.

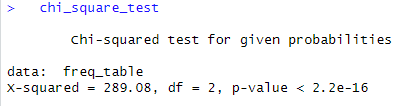
1. 
2.   
   
3. 
4. 
5. The deviation of the distribution of net wins from the expected distribution in Part 3 is primarily due to the change in the game order and series length. In Part 3, we have a best-of-five series where the first team to win three games wins the series, with alternating games between Boston and Miami and the first game played in Boston.

The in-game change order and series length introduce different probabilities for each outcome, resulting in a different distribution of net wins. The specific probabilities for each outcome, such as Celtics winning the series, Celtics losing the series, or the series going to the fifth game, are calculated based on the updated game order and series length. These probabilities influence the frequency of different net win values, and when compared to the expected distribution, they may deviate significantly.

In contrast, Part 2 had a different game order but was still a best-of-three series. The probabilities for each outcome were adjusted accordingly, but the series length remained the same. This similarity in series length between the expected distribution and the generated distribution of net wins in Part 2 could contribute to a better result where the observed distribution approximates the expected distribution more closely.

**Problem 4:**

Repeat the analysis from Problem 1, but with a best-of-seven series where the first team to win four games wins the series, with the team with home-court advantage hosting games 1, 3, 4, and 7, and the opponent hosting games 2, 5, and 6. Assume the Boston Celtics have the home-court advantage against the Miami Heat in the NBA Finals.

1. 
2.   
   
3. 
4. 
5. In Part 4, the probabilities, net win values, and calculations were adjusted for a best-of-seven series in the NBA Finals, where the Boston Celtics have a home-court advantage against the Miami Heat. However, the results indicate that the estimated expected net win and the observed distribution of net wins deviate significantly from the expected values based on the given probabilities. This suggests that the betting strategy in Part 4 may not be favorable, as the actual outcomes differ from the expected outcomes.

**Final Conclusions:**

The analysis of the betting strategy in different scenarios reveals the following conclusions:

Problem 1: The initial betting strategy in a best-of-three series showed a significant deviation between the expected net win and the observed outcomes, indicating a high level of risk and inconsistency.

Problem 2: Modifying the game order in a best-of-three series resulted in a closer approximation between the observed and expected net win distributions, but some deviation still existed.

Problem 3: Transitioning to a best-of-five series introduced further adjustments and increased risk, with the significant deviation between observed and expected net win distributions.

Problem 4: Expanding to a best-of-seven series with home-court advantage showed substantial deviation between observed and expected net win distributions, indicating an ineffective strategy.

Overall, the analyzed betting strategy is not consistently effective. We have to improve the betting strategy requires refining probability estimates, incorporating comprehensive data and factors, and adjusting bets based on series progress and changing probabilities.