Module 1 Assignment | Project: Analysis of a Betting Strategy in Sports

Trang Tran

CPS, Northeastern University

ALY6050 | Introduction to Enterprise Analytics

Professor Andrew Kinley

May 30, 2023

Analysis and Summary

Part 1 # Series: Boston - NY – Boston

(i) The probability that the Bruins will win the series is 0.550144.

1. We assign the winning probabilities for the Bruins and Rangers at their homes at 0.56 and 0.52 respectively.
2. The Bruins can win the series in the following outcomes:

   ## O1: W-W-

   ## O2: W-L-W

   ## O3: L-W-W

   Note that W(Bruins) = L(Rangers) and vice versa

   The probability that the Bruins will win the series is the probability of O1 or O2 or O3. So we compute the sum of all three events => *0.550144*

(ii) Theoretical Probability Distribution:

Description: df [4 × 2]

Net_Winnings <dbl>	Probability <dbl>
1040	0.268800
480	0.281344
-600	0.221056
-1120	0.228800

E[X] = 25.70752

Theoretical standard deviation = 849.4008

By listing all 06 total outcomes of the series (on the Bruins side), we easily collapse down into 04 net winnings outcomes and their corresponding probabilities as follows.

```
> net_winnings
[1]  1040   480  -600 -1120
```

```
> p_outcomes
[1] 0.268800 0.281344 0.221056 0.228800
```

After that, we can compute the expected value and standard deviation in the theoretical probability distribution.

(iii) Simulation and Estimation of Expected Net Win:

A simulation of 10,000 different 3-game series was performed, generating random values (Y) according to the probability distribution. The expected net win of the simulated outcomes is (-130.132). A 95% confidence interval was computed to provide a range within which the true population mean is likely to fall. This interval is from 13.55261 to 46.94339. And the 95% confidence interval contains the theoretically expected value (25.70752) that determines the closeness of the simulation estimate to the theoretical value.

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

To assess how closely the distribution of the simulated outcomes (Y) matches the distribution of the theoretical net winnings (X), a chi-squared goodness-of-fit test was performed. The observed frequencies of the simulated outcomes and the expected frequencies based on the theoretical probabilities were compared. The chi-squared test yielded a p-value that is much smaller than the significant level of 0.05 so we have enough evidence to reject the null hypothesis. This concludes a statistically significant difference between the distributions. This suggests that the

distribution of the simulated outcomes may not closely estimate the distribution of the theoretical net winnings.

```
chi squared goodness of fit p-value: 4.23544431262762e-99
```

(v) Overall, based on the observations of parts (ii) and (iii) above, the betting strategy appears to be reasonably favorable, with the simulation results aligning closely with the theoretical expectations, and the theoretically expected value is positive. However, the statistical significance in the chi-squared test suggests some deviation between the observed and expected distributions.

Part 2 # Series: NY - Boston - NY

We did the whole process the same as in part 1 and have the summary below:

(i) The probability that the Bruins will win the series is 0.509952 (slightly smaller than the result in Part 1)

(ii) Theoretical Probability Distribution:

Description: df [4 × 2]

Net_Winnings <dbl>	Probability <dbl>
1040	0.268800
480	0.241152
-600	0.261248
-1120	0.228800

E[X] = -17.69984

Theoretical standard deviation = 852.6654

(iii) Simulation and Estimation of Expected Net Win:

The expected net win of the simulated outcomes is -84.7

95% Confidence Interval: -32.24504 to 1.085036

The 95% confidence interval contains the theoretically expected value (-17.69984) that determines the closeness of the simulation estimate to the theoretical value.

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

P-value << 0.05 (alpha) so we have enough evidence to reject the null hypothesis and conclude that there is a statistically significant difference between the distributions. This suggests that the distribution of the simulated outcomes may not closely estimate the distribution of the theoretical net winnings.

```
chi squared goodness of fit p-value: 3.551714424806 31e-30
```

(v) Overall, based on the observations of parts (ii) and (iii) above, the betting strategy seems to be not favorable, with the simulation results aligning closely with the theoretical expectations, and both the theoretically expected value and simulation expected value of net winnings are negative.

References

Albright, S. (2016) Business Analytics. Sixth Edition. Cengage Learning. Boston, MA.