Northeastern University
ALY-6050

ALY-6050   Module One Project

Project: Analysis of a Betting Strategy in Sports

The project consists of three parts. The submission of this project will consist of two attachments:

1. A Word document that is prepared according to the APA standards of formatting – this refers to the title page, page numbers, spacing, and citations. In the Word document, explain the experiments and their respective conclusions, and additional information as indicated in each problem. Your report should be submitted as a pdf titled as follows: ALY6050_MOD1Project_ LastName.pdf
2. You should complete this project as an R script (.R) or as an Excel workbook (.xls) and submit your R script named as follows: ALY6050_MOD1Project_ Lastname.R (or .xls)

Problem:

Suppose that Boston Red Sox and New York Yankees (two Major 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 at their home stadium is 0.6 and the probability that the Yankees win their home game is 0.57. Suppose that you bet on the outcome of each game. If the Red Sox win a game, you win $500. If the Red Sox lose a game, you lose $520 In parts 1-3 below, assume that the outcomes of the games are independent of each other.

Part 1: (60 % of the grade)

Consider the series where the games are played in Boston, New York, and then Boston (if necessary). Complete in R or Excel and report on each step below.

(i)   Calculate the probability that the Red Sox will win the series.

(ii)   Construct the theoretical probability distribution for your _net winnings in dollars_ (X) for the series. From this theoretical calculation you should also compute and record the expected value of the net winnings (the mean of X) and the theoretical standard deviation of X.

(iii)   Create a simulation of 10,000 different 3 game series by using R to create 10,000 random values for X. Let these random values be denoted Y. Each Y value denotes an outcome of the series as defined by the probability distribution. These 10,000 outcomes represent a statistical sample of possible outcomes. Use this sample of outcomes to estimate the expected net win by using a 95% confidence interval. Does this confidence interval contain the theoretical E(X)?

(iv)   Construct a frequency distribution for Y. Use the Chi-squared goodness of fit test to verify how closely the distribution of Y has estimated the distribution of X.

(v) 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.

Part 2: (30% of the grade)

Repeat part 1 above but assume that the first game is played in New York, the second game is played in Boston, and the third game (if it becomes necessary) is in New York.

Part 3: (10% of the grade)

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, and with the first game played in Boston.

Reference:

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

Northeastern University
ALY-6050

Rubric ALY 6050 Project 1

Category Score	Characteristics
A Range *Excellent* **90–100 points**	• **R/Excel** Accurate completion of 90%–100% of all R/Excel requirements. Code and/or excel is well formatted and easily readable. • **Report (Content)** Complete presentation and analysis of key results. Contains all required tables, and visualizations. Provides a precise description of the analytical concepts and theories used in the analysis. • **Report (Style and Submission)** Title page, introduction and conclusion/recommendation included; accurate APA citations; minor grammar or spelling errors; page numbers.
B Range *Good* **80–90 points**	• **R/Excel.** Accurate completion two parts of project. Code and/or Excel is poorly formatted or difficult to read. • **Report (Content)** At most one major required component missing. Report shows gaps in reasoning or conclusions not supported by the data. • **Report (Style and Submission)** Missing one of the required elements (introduction, conclusion, etc.); incomplete or incorrect citations; occasional grammar or spelling errors. Imprecise.
C Range *Satisfactory* **70–80 points**	• **R/Excel.** Completion of only a single part of the project. Major deficiencies in readability. • **Report (Content)** Report missing major required elements; evidence for recommendations is unclear or inaccurate; lack of organization. • **Report (Style and Submission)** Missing more than one of the required elements; few or no citations; frequent grammatical and spelling mistakes.
F Range *Unsatisfactory* **0–70 points**	• **R/Excel.** Accurate completion of fewer than 70% of all requirements. Disorganized and incomplete code. • **Report (Content)** Mostly missing. • **Report (Style and Submission)** Missing most required elements; major formatting or grammatical errors.

Table 1 Project 1 Rubric