

**STAT 1341**  
**Final Project**  
**Due: December 17, 2021**

**PROJECT OVERVIEW**

Your final project is an individual project that involves implementing the Elo rating systems in conjunction with a Monte Carlo simulation. For this project, you will do the following:

- Choose a sport and season of interest to you out of the following options: MLB from 1969-2019, NFL from 1970-2019, NHL from 1967-2019, or NBA from 1971-2019.
- Use the datasets posted on Canvas to obtain the necessary data for the season you have chosen.
- Obtain each team's preseason rating using Elo. Use the optimal values from Lecture 18 unless you want to try a different scaling function. In that case, justify your choice and specify the new scaling and weighting parameters.
- Use these ratings to simulate the regular season and playoffs 10,000 times, similar to the second problem from Homework 10. Be sure to check the playoff structure for the season and sport you chose.
- Along the way, save the number of wins/points each team accumulated in each of the 10,000 simulated seasons as well as the playoff results and division winners.

Once the simulation is complete, you will use those results to do a post-hoc analysis on the season. You will compare the actual regular season and playoff results against the results of the Monte Carlo simulation. You will also identify three teams to do a closer analysis on: one that overachieved, one that underachieved, and one that performed close to expected. The final product will be a PowerPoint presentation complete with your analysis, tables, and graphs. Your presentation must contain the following sections:

- **Season Overview:** Describe any major events that occurred during the season and/or describe why you thought the season might be interesting to simulate
- **Season Format:** Describe the regular season format and playoff format
- **Preseason Elo Ratings:** Insert tables/graphics of the preseason Elo ratings for all teams in the league
- **Simulated Regular Season Results:** Insert tables/graphics of simulated and actual win totals along with win residuals
- **Comparison of Simulated and Actual Regular Season:** Compare the simulated results with the actual final regular season standings graphically and numerically
- **Discussion of Regular Season Results:** Discuss the simulated results in relation to the actual final regular season standings
- **Simulated Playoff Results:** Insert tables/graphics of playoff appearances, division titles, championship game appearances, and championships
- **Actual Playoff Bracket:** Insert a graphic of the actual playoff bracket from the season
- **Comparison of Simulated and Actual Playoff Results:** Compare the simulated playoff results with how the playoffs actually played out
- **Overview of Three Selected Teams:** Choose three teams of interest to analyze closer - one that performed close to expectations in the simulation, one that overachieved, and one that underachieved
- **Relevant Statistics:** Choose relevant statistics that can be used to further analyze why these three teams performed as they did. (e.g. OPS and FIP in baseball, QBR in football, the four factors in basketball, Fenwick and PDO in hockey). Summarize these statistics across the league using graphical and numeric summaries.
- **Analysis of Average Team:** Use the relevant statistics from above to describe or hypothesize why the first team performed similar to expectations

- **Analysis of Overachieving Team:** Use the relevant statistics from above to describe or hypothesize why the second team overachieved
- **Analysis of Underachieving Team:** Use the relevant statistics from above to describe or hypothesize why the third team underachieved
- **Conclusions/Takeaways:** Summarize the major talking points or results from the simulation

I encourage you to make use of the code from Lectures 18 and 21-23 to run these simulations. You will probably need to adjust the playoff structure in the code from Lecture 22. The final product will consist of a PowerPoint presentation, the R file containing the code, and any data used to analyze the question. These materials must be submitted no later than December 17, 2021 on Canvas.

Your initial Elo rating CSV file must:

- Be updated to reflect the correct conferences and divisions for each team in the league for the season you are simulating
- Include defunct/folded teams with 'NA' as their conference and division

Your R code must:

- Run with no errors (Warnings are fine)
- Include clear comments describing what each line/chunk does. (You can leave my comments in, but be sure to include new comments detailing any changes that you make, especially to the playoff structure.)

#### GRADING

Your final project will be graded according to the following criteria:

Component	Points
<b>Initial Elo Ratings CSV</b>	5
<b>R Code</b>	10
Runs	5
Well-Documented	5
<b>Presentation</b>	75
Season Overview	5
Season Format	5
Preseason Elo Ratings	5
Simulated Regular Season Results	5
Comparison of Simulated and Actual Regular Season	5
Discussion of Regular Season Results	5
Simulated Playoff Results	5
Actual Playoff Bracket	5
Comparison of Simulated and Actual Playoff Results	5
Overview of Three Selected Teams	5
Relevant Statistics	5
Analysis of Average Team	5
Analysis of Overachieving Team	5
Analysis of Underachieving Team	5
Conclusions/Takeaways	5
<b>Spelling/Grammar</b>	5
<b>Overall Impression and Quality</b>	5
<b>Total</b>	100

When you are ready to submit, upload (1) the final presentation, (2) the R code, and (3) the initial Elo rating CSV to Canvas.