Proposal: An Analysis of Hockey's Most Controversial Statistic

Melanie Desroches
Department of Statistics
University of Connecticut

September 26, 2024

Introduction In hockey, a very common yet controversial statistic that is reported is the plus/minus statistic. Plus/minus is calculated by taking all the even strength goals that were scored by their team while they were on the ice and subtracting it by the total number of goals that were scored by the other team while they were on the ice. The idea behind the number is that it will show if a player has an overall positive or negative contribution when they are playing. However, the plus/minus has a lot of drawbacks. According to Brian MacDonald "The traditional plus-minus statistic in hockey is highly dependent on a player's teammates and opponents" Macdonald (2011). In ice hockey, players tend to play on "lines", meaning the same three forwards and the same two defencemen tend to play at the same time. Because of this, plus/minus is heavily influenced by who plays with who. Some argue that the plus/minus is more of a team statistic rather than an individual statistic. Many of the criticisms of this statistic comes from the fact that there are a lot of confounding variables that influence how the plus/minus is calculated. Alternatives to the plus/minus statistic have been developed, such as COSRI or Fenwick. Understanding the limitations and potential of plus/minus can help coaches, analysts, and fans better interpret player data. While advanced statistics are valuable, plus/minus is still popular, especially among casual fans. This project explores whether the plus/minus still holds relevance or whether it should be supplemented or replaced with other metrics.

Specific Aims The goal of this project is to evaluate the effectiveness of the plus/minus statistic. One of the topics of discussion is if the plus/minus statistic can be used to determine how good a player is both on offense and defense. I also want to see if the plus/minus has any validity being used as an individual statistic or as a team statistic. Based on my observations, I will compare the plus/minus statistic to other statistics that have a similar purpose such as CORSI or Fenwick. These questions are critical as plus/minus is used in scouting and analysis, but its reliability has been debated. By comparing it with modern metrics, I aim to shed light on its true value in player evaluation.

Data I am using data from the National Hockey League (NHL). The data is collected from the websites Natural Stat Trick and NHLEdge. I collected data from the 2021-2022, 2022-

2023, and 2023-2024 seasons. I am particularly interested in the advanced hockey statistics that are collected for each player in the NHL. These statistics include CORSI, Fenwick, high-danger chances, shooting percentages, and more. This data will also include more basic statistics such as points, goals, assists, plus/minus, time on ice, etc.

Research Design and Methods This study will evaluate the effectiveness of the plus/minus statistic in ice hockey to determine whether it can assess a player's offensive or defensive contributions and if it should be treated as an individual or team statistic. The analysis will also compare plus/minus with other advanced hockey metrics, such as Corsi and Fenwick, to assess its overall utility. The first step of the project is to collect all relevant data. In this case, I will be exporting the data from all players from the 2021- 2024 seasons into an Excel or csv file. Once the data has been collected, it needs to be cleaned. After the data has been prepared, I can start some initial analysis to see what variables are correlated and where issues of collinearity exist, since this will effect which models and variables I use. In order to determine if or how plus/minus can be used to assess offensive and defensive contribution, I will use a ridge regression. Since the plus/minus statistic tends to deal with collinearity, a ridge regression might be more appropriate for analysis. Similar work has been done in Macdonald (2012). The coefficients from the ridge regression model will reveal whether offensive or defensive metrics contribute more to explaining plus/minus. I plan on using a similar method for comparing the plus/minus with other advanced statistics, replacing the offesnive and defensive statistics with CORSI, Fenwick, and any other advanced statistics that are deemed relevant. When assessing if plus/minus is more influenced by individual performance or team performance, I am considering using a mixed-effects ridge regression model. By using ridge regression with mixed-effects, it will control for the potential influence of team quality on individual plus/minus values. This will help us determine how much of the variability in plus/minus is explained by the team versus individual players.

Discussion The hardest part of this project will be finding ways to get around collinearity. Many of these statistics are highly correlated. The main goal to get around this is to implement ridge regression models. According to Brian MacDonald in his ridge regression analysis of the plus/minus, "Ridge regression frequently reduces the error bounds in the estimates and improves the predictive performance of the model when collinearly exists in the data. Ridge regression introduces bias in the estimates, but the tradeoff is typically worthwhile." Macdonald (2012). Mixed-effects models, principal component analysis, or other models could be used if ridge regression proves to be ineffective. I am also assuming that the data that I am collecting from Natural Stat Trick is reliable and accurate. While Natural Stat Trick provides advanced metrics, the quality and completeness of the data may vary. I will have to be extra dilligent in the data cleaning process to ensure that variables of interest are available and consistent across the dataset.

References

Macdonald, B. (2011, January). A regression-based adjusted plus-minus statistic for nhl players. *Journal of Quantitative Analysis in Sports* 7(3).

Macdonald, B. (2012, January). Adjusted plus-minus for nhl players using ridge regression with goals, shots, fenwick, and corsi. *Journal of Quantitative Analysis in Sports 8*(3).