Aidan Gross, Owen Gallagher & Ross Brancati
Stat 625 Final Project Summary
December 8, 2022

Optimizing Player Performance - UMass Men's Ice Hockey Data Exploration

Introduction:

Ice hockey is a fast, dynamic game played on an ice surface where two teams are competing in games to score goals against each other. At the highest levels, players will often practice for multiple days leading up to a game. Proper training levels, for example putting the correct amount of load/exertion on players, during practices is critical for optimizing in game performance. Too much player load may be detrimental for performance, but too little load may not adequately prepare the athletes for in game scenarios. Luckily, we can quantify metrics related to the overall player load using wearable technology such as inertial measurement units. Catapult Sports developed a product to record movement parameters of athletes during practices and games to capture data related to player load. The UMass men's ice hockey team recently bought into this system and uses metrics derived from catapult sensors to record movement parameters during practices in games. Using a dataset from practice sessions from January - October of 2022, our group sought to perform an exploratory analysis to relate some of the sensor derived metrics to factors controllable by coaching staff. The high level goal is to relay this information in a very simplified manner to the coaching staff in hopes of properly planning practices to optimize in-game performance.

Project Benefits:

This project benefited us in multiple ways, but two stand out. Firstly it allowed us to use real world data and the challenges with getting real world data ready to model. Secondly, the project offered our group the opportunity to try the methods we learned in the course, but in a setting where there is no correct answer unlike the homeworks.

One of the obstacles we faced in this project was dealing with real world data. To begin with, we had to read the ice hockey teams practice journal and enter the number of zones that a drill used manually since that was not provided in the data set. Furthermore, we had to understand what many of the variables actually represented and how we could use them. Once that was completed, we then had to normalize all our variables to drill duration since not all drills were conducted for the same amount of time. Finally, we had to remove some drills called "Rehab Skates" since they were not actual drills, but instead recovery skates for the injured players. All of these challenges allowed us to experience what working with real data is like and how different the process of building a model is compared to when we are given a clean dataset in the homeworks.

Using real data also created benefits in the model building experience. In the homeworks we learn how and when to use different methods and model building techniques. Nonetheless the homework questions lack the open endedness of a real world problem that we experienced in this project. For instance, in the homeworks we are often told which variables to include, and given hints on how to proceed. In the project, and real world, there are no guidelines or hints, but instead a blank slate. Because of this we had to figure out our own variable selection, transformations, and even determine if the regression assumptions were met. In this project we also had to consider more than just building the best model, but a model that could be simple for a hockey coach, with no statistical knowledge, to interpret so

they could build better hockey practices. In the end the real world data allowed us to experience what modeling in the real world is truly like.

Overall the main benefit of this project was it allowed us to get an idea of what being a statistical consultant is like, from cleaning and understanding the data to the final product of building a model. In conclusion, the project allowed us to see that real world statistical modeling is a combination of an art and a science.

Team Member Contributions:

Owen worked significantly in the data exploration phase and helped to make box plots and normalize the data. In the model building phase he brought unique ideas including building a multinomial model to see if we could use the Catapult data to predict the "zone_count" of certain drills. After implementation and analyzing the results he decided that the model was not useful. Owen next made the call to instead build a model with Player Load as the dependent variable and use Zone Count along with a few other predictors as the independent variable which significantly helped us achieve our goals.

Ross compiled the initial dataset from Openfield (Catapult's cloud based data storage system), which involved creating custom data fields to meet our needs. He helped with exploring the data by creating scatterplot matrices and excluding outliers. Ross worked significantly to help his teammates understand the root questions we were trying to get at. As such he generated models using variables that are more "controllable" by the coaches, such as "zone_count", total number of strides, explosive efforts, and force per stride. Through this process, we plan to inform the coaching staff of the variables that are important for explaining high intensity work duration, a variable that is crucial for optimizing player performance in games.

Aidan helped to annotate our initial dataset such that we could analyze our synthetically engineered "zone_count" variable. He contributed to data exploration specifically in the areas of initial data visualization, outlier exclusion, variable selection and some initial stepwise models. He worked closely with his teammates to brainstorm and research the best paths forward to deal with the highly collinear dataset we were working with. From this Aidan took on the challenge of implementing ridge regression to explore and synthesize a model which we will be used to translate some of our findings into understandable "controllable" conclusions for the UMass coaching staff. Lastly he worked closely with teammates to construct and polish our work into a final presentation.

Future Direction:

In the future, the sports science team should look into how some of these metrics relate to in game performance. While providing information about practices is beneficial to the coaching staff, it would be better to relate some of these findings to in game performance. Ultimately, practices don't really mean much for the team's overall record or ranking, so some sort of game data analysis would be critical. One metric that could help quantify this is the acute to chronic work ratio, which is a metric that has been extensively researched and has been shown to effectively relate practice workload to in-game performance. Investigating other metrics and how they relate to game performance would expand the benefits of this project for the UMass hockey team.

References

1). Douglas, A., Johnston, K., Baker, J., Rotondi, M. A., Jamnik, V. K., & Macpherson, A. K. (2019). On-Ice Measures of External Load in Relation to Match Outcome in Elite Female Ice Hockey. Sports (Basel, Switzerland), 7(7), 173. https://doi.org/10.3390/sports7070173

2). Douglas, A. (2021). Catapult Ice Hockey Webinar Series: Changing the Game with Wearable Technology - Part 2. Catapult Webinar Series. Catapult Sports. Melbourne, Australia.

3). Taboga, Marco (2021). "Ridge regression", Lectures on probability theory and mathematical statistics. Kindle Direct Publishing. Online appendix.
https://www.statlect.com/fundamentals-of-statistics/ridge-regression.

4). Douglas, A. (2022). Bridging the Gap: Being a 'Pracademic' in the World of Elite Sport. Conference Proceedings. North American Congress on Biomechanics. August 20-25, 2022.

5). Weisberg, S. (2014). Applied Linear Regression. New York. Wiley.

6). Catapult Sports Website. Accessed December 5, 2022. catapultsports.com.