

Spring 5-7-2020

## Get A Bang For Your Puck: Hockey, Statistics, and Betting

Kathleen Capozello

*Bowling Green State University, capozks@bgsu.edu*

Follow this and additional works at: <https://scholarworks.bgsu.edu/honorsprojects>



Part of the [Business Analytics Commons](#), [Business Intelligence Commons](#), and the [Sports Management Commons](#)

---

### Repository Citation

Capozello, Kathleen, "Get A Bang For Your Puck: Hockey, Statistics, and Betting" (2020). *Honors Projects*. 526.

<https://scholarworks.bgsu.edu/honorsprojects/526>

This work is brought to you for free and open access by the Honors College at ScholarWorks@BGSU. It has been accepted for inclusion in Honors Projects by an authorized administrator of ScholarWorks@BGSU.

**Get A Bang For Your Puck: Hockey, Statistics, and Betting**

**Kathleen Capozello**

**Honors Project**

**Submitted to the Honors College  
at Bowling Green State University in partial fulfillment of the  
requirements for graduation with  
University Honors May 16, 2020**

**Mike Good, Accounting and Information Systems, Advisor  
Kyle Moninger, Applied Statistics and Operations Research, Advisor**

**Table of Contents**

Background & Introduction	3
Research Questions	3
The Math Part 1:	
Gathering Data	6
Building the Equation	8
Explanation of the Equation	10
Accuracy of the Equation	14
The Math Part II:	
Sports Betting Revenues	16
NHL Revenues and Player Salaries	17
The Future of Sports Betting and the NHL	18
Conclusion	19
Annotated Bibliography	20

## **Background & Introduction**

Four years ago, as I finished up my senior year of high school and my Advanced Placement Statistics class, I was tasked with demonstrating some of the knowledge I had learned in the course, in any way I saw fit. I have been a hockey fan for the past six to eight years of my life and jumped at the chance to analyze how good my favorite players actually were. Using the averages of a certain metric, I judged whether National Hockey League defensemen were more effective in their 20's or 30's. Now, as I am finishing my degree in Business Administration, with a specialization in Business Analytics and Intelligence and have increased my knowledge in statistics and analytical methodologies, I thought that continuing this project in more depth would be the perfect use of my Honors Project. This time, I will be looking at more than just a few metrics and I will build a regression equation that will allow us to predict how many goals a skater will score in a future season, based on the statistics from the season prior.

Ahead I will explain the research questions, my methods of gathering data, the process of producing the equation, the amount of money in sports betting and the NHL to then explain the monetary value of the equation.

## **Research Questions**

**How can we use statistical software and sports analysis to predict the number of goals an NHL defenseman or forward will score in a future season?**

One of the goals for this project was to take hockey data and see what tools could be used to analyze it and bring forth value. Microsoft Excel and Minitab are two popular tools that were utilized in this project. Excel allows users to sort and organize data, make comparisons, create tables and charts, and prepare data before moving it into a more powerful tool to run more

complex analysis on it. Minitab allows users to make data transformations and look for patterns. The basis of this project is to create a regression equation that will help predict how many goals a forward or defenseman, also referred to as a “skater”, will score in a future NHL season based on a previous season’s information. The software sorts through the data, both categorical and continuous, and determines which statistics are the most effective predictors. From there, Minitab can be used to create a number of equations that will be narrowed down until the best one is found.

### **What applications does this type of information have? Who can use this information?**

This information is useful in a number of different settings. One application is for the National Hockey League, mostly team general managers, scouts, and coaches so they can understand player potential better and create the best lineups possible and make informed trades and signing decisions. This information is also useful to the players to some extent to see room for potential improvement, and to understand their relationship to the game and some of the factors that influence their gameplay. At the professional level, it is doubtful that any of the players have many misconceptions about hockey left, and as they have been playing hockey for most of their lives, they probably understand their gameplay pretty well. The equation might help them understand their teammates and competitors better and give an understanding of where they fit in on the team. Not everyone is a big goal scorer, there need to be players who are better at defense and better at setting up teammates than scoring themselves to have a well-rounded hockey team. Additionally, there have to be some players out there with a size factor who are good at checking and standing up for their team, though this type of player, once referred to as an

“Enforcer” is now required to have more skills than just hitting people, as the league has placed a bigger emphasis on player safety over the years.

Finally, the information uncovered in this project and the equation produced will be valuable to those involved in sports betting. This includes the gamblers themselves, and the people who run sites like FanDuel and DraftKings, as well as the casinos in Las Vegas, Nevada that offer sports betting. Knowing how many goals a player is projected to score in a season is useful when picking your fantasy team as a gambler, and is useful to the people who set the odds and the prices on the betting. Sports betting is a growing industry and equations like this one will become more popular as the trend grows (“A Look Inside the Numbers of Sports Betting in the U.S. and Overseas”).

### **What value does this information have?**

This project is not only about building an equation to see how many goals a player will score in a future season based on a past season, but also to determine what value this information will have. As was previously mentioned, this equation has value to a few different groups for different reasons, but most of them stem back to money. For general managers and coaches, this will help them decide what players are worth, and help them avoid hitting the salary cap sooner rather than later. Among players it will help them get a sense of their worth and how they might be able to improve, as well as help them gauge other players’ potentials on the ice and help them know who to look out for. In the sports betting arena, this equation will help betters make educated picks and help the industry better understand the factors in play. Understanding the value of the equation will be easier with more knowledge of how much money flows through sports betting and how much revenue the NHL and the teams make in a year.

## **The Math: Part I**

### **Gathering Data**

The primary goal was to acquire a season's worth of data that included all of the potential factors that might make accurate predictors for goals scored in a future season. Some statistics up for consideration included player age, position, and the plus/minus statistic. In order to check the equation's accuracy, there needed to be another complete data set following the one chosen for building the equation. With that in mind, the equation was built off of data from the 2017-2018 season so that the accuracy could be tested with the data from the completed 2018-2019 season. Testing against the 2019-2020 season would not have been feasible as the season would not have wrapped up in time to allow for full and accurate results. It also would have been made impossible by the COVID-19 crisis shutting down sports for the foreseeable future.

Following the selection of a particular season, individual statistics had to be chosen to test their importance in influencing goals scored. Most of the common game play statistics were included in the data gathering. These included total points scored, which is a combination of goals and assists throughout the season, plus goals and assists made on the power play and while short-handed. Some defensive statistics such as hits and blocked shots were also included.. Additionally, positions were listed in the data. Position was a categorical variable marked as either forward or defenseman despite the fact that forwards are distinguished as left wing, right wing, and center based on where they start at puck drop. While there are a lot of human factors in hockey that cannot be statistically measured yet, a player's leadership status as either a Captain or Alternate Captain can be recorded and was listed as a categorical variable of either

“Yes” or “No” in regards to whether or not the player held a leadership position. Player age was also included in the dataset as arguments can be made for both sides of whether or not younger players or more experienced veteran players are more skilled. The last statistic included was country of origin. Though it is called the National Hockey League and is played in the United States and Canada, there are players from all over the world playing. The countries with the most players include the United States and Canada, as well as Russia, Finland, Sweden, and the Czech Republic. There are also players from other European countries such as France, Germany, and Slovakia, as well as one player from Australia. Overall there were 17 countries represented in the 2017-2018 season (“2017-2018 NHL Scoring Leaders”).

After gathering the data, it had to be cleaned. This process of cleaning refers to removing incomplete records, removing outliers, and making it easier to process and analyze. In this case, there were no incomplete records, but there were a lot of records that would skew the results and those were removed to make the data easier to use. First, the data was narrowed down to the players who played in at least 21 of the 85 games. Then, a new category was created for country possibilities that narrowed the number down from 17 countries to 7 countries and an “Other” category. This was done by determining how many players were from each country and if there were fewer than 10 players from a country, those players were put into the “Other” Category. The countries that remained are the United States, Canada, Russia, Finland, Sweden, Czech Republic, and Switzerland. This was done to test if the smaller countries made a difference in the equation or not. The blocked shots statistic was removed because it did not appear to be a significant variable after the initial tests in Minitab. Additionally, the Team statistic was not used because players can be traded.



## **Building the Equation**

A regression equation allows users to pick a factor that they want to predict and then input variables they believe to have significant impact on the prediction and then the software will weigh them and determine what is a good predictor and what is not. The software then assigns the factors a coefficient that represents the influence it has in the equation based on the values provided in the data. What type of regression the user wants will also help determine the steps involved. In this case, a simple linear regression was used, which means if it were graphed, it would look like a straight line. Minitab's regression capabilities allow for a couple methods of equation building but all of them provide you with the information of how important different variables are, how much of the data is accounted for in the equation, and measures of accuracy ("Overview for Simple Regression").

The data was entered into Minitab and the testing of equations began. Usually, when categorical variables are utilized, they are coded into a binary system of 1 and 0 to make it easier for the program to analyze them. Minitab is capable of coding categorical variables either in a binary state, wherein the user chooses in the coding menu which of the two should be 1 and which should be 0. It also allows for a range of coding from -1 to 0 to 1. Coding helps give the categorical variables a numerical explanation or weight when building the equation and thereby aids in describing how important the variable is to the regression model by assigning a coefficient or not including it in the equation. In this case, the -1, 0, 1 coding was used, as the variables regarding the country of origin have more than two possible entries.

Another option Minitab provides for creating regression equations is the Stepwise function. Stepwise functions go through each variable and use an entrance criteria to determine whether or not a variable should be included in the equation. Two methods of Stepwise are very

common: starting with nothing and adding variables based on acceptance criteria in the forward selection method, or starting with all the variables and removing some of them based on the acceptance criteria, resulting in a backwards elimination regression (“Overview for Simple Regression”). Both methods were utilized in the construction of this equation with varying entry criteria, also referred to as  $\alpha$  (alpha). To enter, the p-value of the variable has to be less than or equal to alpha to be included in the equation. If the p-value is greater than alpha then the variable is not a significant predictor and therefore is not included in the equation (“Overview for Simple Regression”).

Once the data is entered and the program is run with the specifications of coding, Stepwise, and entrance criteria, an equation is produced along with statistics to explain the factors and the equation as a whole. Two equations were produced, based on which country classification was included in the data, among other factors. If “Country of Origin”, the variable that had all the countries listed in it, was used, then it was deemed statistically significant and included in the equation. If “Country”, the variable that had the seven most represented countries and the ‘Other’ category, was used, the variable did not meet the entry criteria and was not included in the equation. The readouts from Minitab of the entrance criteria are pictured below. Again, to be deemed a significant predictor, the P-Value had to be less than the alpha level, which was 0.15. The constant does not follow this as it is not a predictor, but calculated by Minitab specifically to fit into the equation.

Equation A:

Equation B:

## Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.139	0.370	0.38	0.707	
Games	-0.02797	0.00809	-3.46	0.001	2.42
Pts	0.3031	0.0148	20.47	0.000	11.53
SOG	0.02098	0.00337	6.23	0.000	5.70
GWG	0.6376	0.0733	8.70	0.000	2.33
PP Goal	0.7838	0.0476	16.46	0.000	2.82
PP Assist	-0.4737	0.0310	-15.29	0.000	4.10
SH Goal	0.626	0.149	4.21	0.000	1.20
SH Assist	-0.437	0.161	-2.71	0.007	1.11
Position					
Defenseman	-1.222	0.114	-10.71	0.000	1.36

## Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	1.729	0.750	2.31	0.021	
Player Age	-0.0335	0.0222	-1.51	0.132	1.08
Games	-0.02702	0.00810	-3.33	0.001	2.51
Pts	0.2976	0.0147	20.18	0.000	11.82
SOG	0.02153	0.00335	6.44	0.000	5.81
GWG	0.6104	0.0732	8.33	0.000	2.40
PP Goal	0.8023	0.0473	16.96	0.000	2.88
PP Assist	-0.4736	0.0309	-15.31	0.000	4.23
SH Goal	0.644	0.149	4.34	0.000	1.24
SH Assist	-0.398	0.162	-2.46	0.014	1.15
Country of Origin					
Austria	5.97	1.34	4.45	0.000	42.03
Canada	-0.777	0.363	-2.14	0.033	10.58
Czech Republic	-1.282	0.550	-2.33	0.020	8.92
Denmark	-2.170	0.978	-2.22	0.027	23.15
Finland	-0.740	0.554	-1.34	0.182	9.05
France	-1.02	1.34	-0.76	0.447	41.99
Germany	-1.55	1.17	-1.32	0.188	32.65
Latvia	-1.45	2.28	-0.63	0.526	119.34
Norway	1.55	2.28	0.68	0.496	118.77
Russia	0.013	0.563	0.02	0.981	9.28
Slovakia	-0.898	0.919	-0.98	0.329	20.65
Slovenia	4.57	2.30	1.99	0.047	121.32
Sweden	-0.752	0.427	-1.76	0.079	7.53
Switzerland	-0.933	0.793	-1.18	0.240	15.86
Position					
Defenseman	-1.225	0.113	-10.82	0.000	1.39

### **Explanation of the Equation**

By this point, there were two equations. The goal was to find which of the equations, which did not have all the same factors in them, or even the same number of factors in them, was the better one. Some may believe that the one with more factors is inherently better because more factors means more things you are using to predict your outcome with. This is not necessarily the case as it poses a risk for over-explaining the data and including variables of little influence to the overall equation (“Overview for Simple Regression”). Most commonly the R-squared, R-squared adjusted, and R-squared predictive statistics are examined to aid in choosing a representative and effective equation. All three statistics are presented as percents and express how much of the data is represented or explained by the equation. The R-squared statistic is the raw percentage. The R-squared Adjusted explains how much of the data is represented by the equation like R-Squared does, but it takes into account the number of variables as well, which is helpful when equations like the ones in question do not have the same number of variables in them. The R-squared Predictive statistic shows what percent of the data can be predicted using the equation, which can be a measure of accuracy (“Overview for Simple Regression”).

The two equations contain nine variables in common which are: Games (Played), Points, Shots on Goal, Game Winning Goals, Power Play Goals, power play Assists, Shorthanded Goals, Shorthanded Assists, and Position. Equation A contains only these variables, and Equation B also includes Country of Origin and Player Age. Both equations are pictured below with their model summaries.

Equation A:

### Regression Equation

$$\begin{aligned} \text{Goals} = & 0.139 - 0.02797 \text{ Games} + 0.3031 \text{ Pts} + 0.02098 \text{ SOG} + 0.6376 \text{ GWG} + 0.7838 \text{ PP Goal} \\ & - 0.4737 \text{ PP Assist} + 0.626 \text{ SH Goal} - 0.437 \text{ SH Assist} - 1.222 \text{ Position\_Defenseman} \\ & + 1.222 \text{ Position\_Forward} \end{aligned}$$

### Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
2.44027	93.24%	93.15%	92.94%

Equation B:

### Regression Equation

$$\begin{aligned} \text{Goals} = & 1.729 - 0.0335 \text{ Player Age} - 0.02702 \text{ Games} + 0.2976 \text{ Pts} + 0.02153 \text{ SOG} + 0.6104 \text{ GWG} \\ & + 0.8023 \text{ PP Goal} - 0.4736 \text{ PP Assist} + 0.644 \text{ SH Goal} - 0.398 \text{ SH Assist} \\ & + 5.97 \text{ Country of Origin\_Austria} - 0.777 \text{ Country of Origin\_Canada} \\ & - 1.282 \text{ Country of Origin\_Czech Republic} - 2.170 \text{ Country of Origin\_Denmark} \\ & - 0.740 \text{ Country of Origin\_Finland} - 1.02 \text{ Country of Origin\_France} \\ & - 1.55 \text{ Country of Origin\_Germany} - 1.45 \text{ Country of Origin\_Latvia} \\ & + 1.55 \text{ Country of Origin\_Norway} + 0.013 \text{ Country of Origin\_Russia} \\ & - 0.898 \text{ Country of Origin\_Slovakia} + 4.57 \text{ Country of Origin\_Slovenia} \\ & - 0.752 \text{ Country of Origin\_Sweden} - 0.933 \text{ Country of Origin\_Switzerland} \\ & - 0.535 \text{ Country of Origin\_USA} - 1.225 \text{ Position\_Defenseman} + 1.225 \text{ Position\_Forward} \end{aligned}$$

### Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
2.40009	93.60%	93.37%	*

Initially, both of these factors appeared to be very significant as the ages of players range from 18-45 and the players come from more than a dozen countries around the world. However, upon inspecting both models' R-squared, R-squared Adjusted, and R-Squared Predictive statistics, there was not a significant difference between them, despite one equation having more

variables. In Equation A, the R-squared Adjusted was 93.15% and the R-squared Adjusted for Equation B was 93.37%.

Reexamining the data provided an explanation for how close the R-Squared Adjusted scores were, despite the addition of two variables to Equation B. 318 of the 686 players who were included in the research were Canadian, which is almost 50% of the data. This would mean that while all 318 Canadian players are professional hockey players and they are better than the majority of people at hockey, all 318 of them cannot possibly be the best of the best and therefore will have a range of other data that would make being from Canada an insignificant predictor of how many goals a player could score. There were six other countries with more than ten players, including the United States, which only had 175 of the 686. On top of this, while it seems that age would matter because it could mean differing skill levels, this is not necessarily the case. It turns out that the players in the dataset range from 18-45, or really 18-38, with only three of the 686 being 40 years of age or older. As was previously mentioned, these are the best hockey players in the world, or else they would have found another profession, however, their age does not necessarily mean that they are a better goal scorer, though it may mean they are a better defenseman, or they may be a leader on the ice. Player ages also depend on the teams. Some teams have enough depth in their lineups that they can keep younger players at their associated minor league teams for longer until they need them. Other teams are not that fortunate and need to play young players immediately, whether or not they are a top goal scorer when they get drafted. The needs of the team, along with the innate and learned skills of a player make more of a difference than age, and therefore age does not prove to be an explanatory variable for goal scoring.

Based on the above explanations of the ages and nationalities of the players, along with the lack of a significant difference, which in this case would have been a difference of 5% or more, Equation A was selected as the better equation. It is pictured again below in a readout from Minitab.

### Regression Equation

$$\begin{aligned} \text{Goals} = & 0.139 - 0.02797 \text{ Games} + 0.3031 \text{ Pts} + 0.02098 \text{ SOG} + 0.6376 \text{ GWG} + 0.7838 \text{ PP Goal} \\ & - 0.4737 \text{ PP Assist} + 0.626 \text{ SH Goal} - 0.437 \text{ SH Assist} - 1.222 \text{ Position\_Defenseman} \\ & + 1.222 \text{ Position\_Forward} \end{aligned}$$

### Accuracy of the Equation

With the equation created, accuracy becomes the next concern. As the Minitab output described, Equation A had an R-Squared Predictive of 92.94%. That is very high and means that 92.94% of the data in the set used to build the equation can be predicted by it. However, the data used to create the equation only accounts for one season of professional hockey. By utilizing the coefficients and the data from the season that followed, the accuracy of the equation can be put to the test. In this case, fifteen randomly selected players were chosen and their data from the 2018-2019 season was entered in the equation and the output was compared to the actual number of goals they scored during the season.

Player Name	Equation Output	Goals Scored
Michael Amadio	4.90409 ± 0.586	6
Tom Wilson	18.42433 ± 0.586	22
Christian Fischer	7.31289 ± 0.586	11

Filip Forsberg	$27.19 \pm 0.586$	28
Brad Richardson	$12.863392 \pm 0.586$	19
Mathew Barzal	$18.773124 \pm 0.586$	18
Pavel Zacha	$13.586862 \pm 0.586$	13
Matt Cullen	$8.915022 \pm 0.586$	7
Ryan McDonagh	$11.156344 \pm 0.586$	9
Danny O'Regan	$1.374982 \pm 0.586$	0
Jacob Trouba	$10.812064 \pm 0.586$	8
Robert Hagg	$5.126684 \pm 0.586$	5
Tim Schaller	$4.000154 \pm 0.586$	3
Anders Lee	$26.988624 \pm 0.586$	28
Nicklas Backstrom	$25.24578 \pm 0.586$	22

Based on the above table, the equation was fairly accurate. A 90% confidence interval was calculated based on the mean number of goals in the 2017-2018 season in Excel and that is where the margin of error of 0.586 comes in. This means that there was a 90% confidence level that the number of goals scored was in the range calculated by the equation (“Confidence Intervals”). While it was not 100% accurate for all 15 players, it was quite close on some and this is only the first set of data that the equation has been tested against. With more data and more



trials, the equation should become more accurate. Overall, this equation is accurate when transferring from one season to another. As hockey is not a controlled experiment and it is played by humans, there will always be factors that cannot be controlled, such as team chemistry, trades, injuries, and emotional and mental conditions, so there will always be a margin of error and a chance that the equation will not be 100% accurate. Eventually, the goal is to continue feeding information into Minitab using averages and updated rosters, to create a better understanding of what factors are the most important and build a better equation.

## **The Math Part II:**

### **Sports Betting Revenues**

One of the segments that this equation would be most valuable to is the sports betting arena. The most popular sports in the United States for betting are football, basketball, and baseball, in that order. Hockey falls somewhere in the “Other” category, but this category still accounted for \$428 million in revenue in 2017 (“A Look Inside the Numbers of Sports Betting in the U.S. and Overseas”). The total sports revenue from June of 2018 until March 22, 2020 was \$1,390,427,457, but there was about 20 times that amount placed in wagers over that time (“US Sports Betting Revenue And Handle”)

All of this goes to show that sports betting is on the rise and has a lot of money in it in the future. As the legalization across the states spreads, and as sports come back after the COVID-19 crisis, we will see an increase in the revenue coming from this, as well as an increase in wagers and payouts as those three things correspond with each other.

## **NHL Revenues and Player Salaries**

The NHL brought in \$5.09 billion during the 2018-2019 season (Gough, “NHL - Total League Revenue”). This comes from air time, tickets, merchandise, and more. The team with the largest revenue is the New York Rangers with \$240 million in 2019. Even the team with the smallest revenue, the Arizona Coyotes, made \$102 million in 2019 (“The Business of Hockey.”). This means that there is a lot of money flowing through the NHL. One might think that with an increased revenue, teams can spend as much as they can afford to build the best team and buy the best players. This scenario would lead to a spending war between teams as players are offered increasingly higher salaries and teams with lower revenue are left with players who are not nearly as good. Thankfully, this is not the system, or the sport would become boring as the richest teams would consistently stay good and there would be no surprises. Due to the lottery draft, trade deadlines, and salary caps, the teams can all compete for top talent. The salary cap is what the governing body of the league sets as a ceiling for how much a team can spend on player salaries in a given year. The cap for the 2018-2019 season was \$79.5 million and the cap for the 2019-2020 season was \$81.5 million (Rosen). This is the same across the whole league, which limits unfair competition for players between organizations. If the teams can use equations like mine to predict a player’s production in future seasons, among other things, they can make more informed signings and better manage their money so they do not encounter the salary cap as quickly as they might without the information. Additionally, this will allow them to create the perfect mixture of players with salaries that accurately reflect the players’ values to the organization. This is extremely helpful because the average salaries across the NHL by team range from \$1.54 million to \$3.54 during the 2018-2019 season (Gough, “Annual Average Player Salary”).

### **The Future of Sports Betting and the NHL**

Sports betting just became legal in all 50 states in 2018. This means it is a growing market and is steadily becoming more popular outside of Las Vegas, Nevada and other pockets of the country where it was already legal and popular (“US Sports Betting Revenue And Handle”). According to the American Gaming Association in 2018, the National Hockey League stood to gain more than \$200 million following the legalization of sports betting due to an increase in publicity and fans across the nation (“NHL Could See \$216 Million in Revenue Annually Due to Legalized Sports Betting”). While hockey is not the biggest sport for betting right now, with both the addition of an NHL team in Las Vegas, the biggest gambling center in the country, and the legalization of sports betting all across the country in 2018, we will see a rise in gambling centered around hockey.

Currently, when it comes to hockey, most bets are based on which team is going to win each game and by how much, but with the rise of sites such as FanDuel and DraftKings, there will be more places to create hockey fantasy teams and bet on those (“NHL Betting”). The fantasy team builders will be the ones who appreciate this equation the most, because fantasy teams rely on gamblers drafting a team of players from across the league and facing off with peers to see whose picks have the better statistics. Being able to predict beforehand how many goals a player might score in a season will help people build better fantasy teams. This will entice people to bet higher and more frequently and bring more revenue to the companies that manage the gambling. Additionally, if those companies can get a better understanding of players’

worths, they can more accurately manage the prices of bids and bring in more revenue for themselves.

## **Conclusion**

Hockey is a very human sport, and there is a lot that cannot be predicted at this time. However, using the statistics from past seasons, we can begin to build and improve equations that will aid in the prediction of outcomes for seasons in the future. These equations will help general managers, coaches, and scouts assemble better teams as well as help fans build fantasy teams and help grow the culture around NHL betting. While some variables may not be as influential to the game play outcomes as we might expect, by continuously analyzing the information collected over time, we can learn more about the statistics and the game and help others use data to make informed decisions.

### Annotated Bibliography

“A Look Inside the Numbers of Sports Betting in the U.S. and Overseas.” *Sports Business*

*Journal*, 16 April 2018, <https://www.sportsbusinessdaily.com/Journal/Issues>

[/2018/04/16/World-Congress-of-Sports/Research.aspx](https://www.sportsbusinessdaily.com/Journal/Issues/2018/04/16/World-Congress-of-Sports/Research.aspx). Accessed 21 April 2020.

This article helped me understand how popular sports betting is here in the United States and elsewhere in the world. It also helped me understand how popular hockey betting was in relation to the betting on other sports as well as the opportunities for growth and the growth trends that it has seen so far. This relates to the other sources on sports betting and helps me to synthesize them and gain a better understanding about sports betting overall and how my equation might impact it.

Alamar, Benjamin and Mehrotra, Vijay. “Beyond Moneyball: The future of sports analytics.”

*Analytics Magazine*, Aug. 2018, <http://analytics-magazine.org/beyond-moneyball-the>

[-future-of-sports-analytics/](http://analytics-magazine.org/beyond-moneyball-the-future-of-sports-analytics/). Accessed 20 April 2019.

This article was one of the first I found and helped point and direct my research and project ideas. It detailed some of the history of sports analytics and the growing importance of it to the sports industry. It also talked about how the profession will be growing in the future which is insightful to me as it is something I have considered doing for a living. The thought of the sports analytics industry growing is exciting as this project could help me break into the industry.

“Confidence Intervals.” *Math is Fun*, [https://www.mathsisfun.com/data/confidence-](https://www.mathsisfun.com/data/confidence-interval.html)

[interval.html](https://www.mathsisfun.com/data/confidence-interval.html).

Accessed 29 April 2020.

This source gave me a refresher on confidence intervals, their uses, and how to calculate one. Confidence intervals allow one to judge accuracy to a certain degree of one’s choosing by seeing how close things are to the expected goal. This source relates to the Minitab source as both helped me with building my equation and testing its accuracy. I had learned about confidence intervals before, but I needed to be refreshed on the formula for calculations.

Gough, Christina. “Average annual player salary in the National Hockey League in 2018/19, by

team(in million U.S. dollars).” *Statista*, 9 Aug. 2019, [https://www.statista.com/](https://www.statista.com/statistics/675382/average-nhl-salary-by-team/)

[statistics/675382/average-nhl-salary-by-team/](https://www.statista.com/statistics/675382/average-nhl-salary-by-team/) . Accessed 21 April 2020.

This source is from *Statista* which is a trustworthy place to find statistics and graphs. This source helped me to understand player salaries and the impacts they can have on salary caps.

This relates to the article on salary caps from the NHL as some of the player averages are higher than the increase in the salary cap. The equation I created will be helpful to the people who make decisions about player salaries as they will be able to determine how much the players are worth based on goal production.

Gough, Christina. "National Hockey League - total league revenue from 2005/06 to 2018/19 (in billion U.S. dollars)\*." *Statista*, 14 Jan. 2020, <https://www.statista.com/statistics/193468/total-league-revenue-of-the-nhl-since-2006/>. Accessed 21 April 2020.

This chart helped showcase the revenue that came into the National Hockey League over the past several years. It helped me to understand how valuable this information could be to the league as they make decisions about players and advertising and how they will interact with the gaming institutions. *Statista* is a trusted source for graphs and statistics. This source ties in with the other one from *Statista* as I could compare player salaries and league revenues.

Gouker, Dustin. "How Much Money Did DraftKings, FanDuel And The DFS Industry Make In The Past Year? Now We Know Almost Exactly." *Legal Sports Report*, 20 Oct. 2017, <https://www.legalsportsreport.com/16152/draftkings-fanduel-dfs-revenue/>. Accessed 21 April 2020.

This article showcased the value that can be found in online sports betting, which is where a lot of fantasy teams take place. Fantasy teams are the biggest gambling population that would be impacted by the equation created in this project. This source defined very clearly how much money gamblers have spent over time to participate and it shows a continuous growth. Now, while that is being hindered by the lack of sports due to COVID-19, we would hope that it will continue to grow again once sports resume.

King, Bill. "Hockey Antes Up The NHL is taking advantage of new gambling opportunities, and its odds of winning keep getting better.." *Sports Business Journal*, 29 Oct. 2018, <https://www.sportsbusinessdaily.com/Journal/Issues/2018/10/29/Leagues-and-Governing-Bodies/NHL-gambling.aspx>. Accessed 21 April 2020.

It explained the increase in hockey betting that will come following the legalization of sports betting. It relates to the other articles that also deal with this topic from the Legal Sports Report. Additionally, it went into detail about how more money can be made in the future. It discussed the decisions made by different sporting leagues regarding official league data and how that will impact the betting. It gave a background on how sports betting works from the side of the gambling institutions and the league organizations.

“NHL Betting.” *The Sports Geek.*, <https://www.thesportsgeek.com/sports-betting/nhl/>. Accessed 16 April 2020.

This source discussed sports betting in terms of the National Hockey League and the sites that are most common among those who bet on the NHL proceedings. It also explained the most common types of bet placed on NHL games such as over under bets. This gave a quick crash course on how to bet on the NHL and is valuable for understanding how it works and how the equation might be beneficial.

“NHL Could See \$216 Million in Revenue Annually Due to Legalized Sports Betting.”

*American Gaming Association*, 2 Oct. 2018, <https://www.americangaming.org/new/nhl-could-see-216-million-in-revenue-annually-due-to-legalized-sports-betting-4/>. Accessed 21 April 2020.

This source from the American Gaming Association is trustworthy as it is a press release from the organization. It provided information regarding the profits the NHL could expect to see after the legalization of sports betting in the United States. It also explained where these revenues could come from. The source provided insights into how valuable this equation could be in the future as sports betting grows. It relates to the other articles about the rise of sports betting from the *Legal Sports Report*.

“NHL Stats.” *Rotowire*, <https://www.rotowire.com/hockey/stats.php>. Accessed 2 March 2020.

*Rotowire* is a trusted source for fantasy sports players and that is why I chose to gather the majority of my data here. It provided detailed statistics in all of the areas that I wanted to examine aside from age and country of origin. This source relates to the *Quant Hockey*, NHL Teams, and NHL Captains sites and they all provide a lot of similar information about the same topics and when combined can create powerful tools for analytics. This source was also the easiest to obtain data from as it allows you to export it directly to an Excel document. This source is where I obtained the data for both the 2017-2018 and 2018-2019 seasons for the equation’s preparation and testing.

Nover, Scott. “The rise of daily fantasy and sports betting has created an economy of its own.”

*Vox*, 29 Jan. 2020, <https://www.vox.com/2020/1/29/21112491/daily-fantasy-sports-betting-dfs-merch-analysis-weatherman>. Accessed 21 April 2020.

*Vox* is a trusted source for reporting in all categories. This article dealt with the growth in the sports betting arena and how much money is flowing in and out of it. It helped me gain a better understanding about how popular sports betting is and potentially could be. It also helped

me to understand how valuable this equation could be in the future as sports betting continues to grow in popularity after the COVID-19 crisis.

“Official League Data.” *Legal Sports Report*, 14 Feb. 2020, <https://www.legalsportsreport.com/official-league-data/>. Accessed 21 April 2020.

This source pertained to the data used in sports gambling and helped me understand the different bets that can be conducted. It also helped me determine the relationship between analytics and betting and gave me a better understanding of how my skill sets as an analyst are useful to gamblers and the people running the gaming institutions and websites. There are a few other pages from this site used in this project and I believe it to be a knowledgeable and trustworthy source on the subject of sports betting.

“Overview for Simple Regression.” *Minitab*,

<https://support.minitab.com/en-us/minitab-express/1/help-and-how-to/modeling-statistics/regression/how-to/simple-regression/before-you-start/overview/>. Accessed 21 April 2020.

This source gave me a refresher on my statistical knowledge and helped me navigate the Minitab software as I was using a newer version than I had in the past. Additionally, it helped me understand the meaning of different outputs and be able to explain them in less statistical terms so that any reader would be able to understand the process and the outputs. This is a trusted source as it is the support site for the software.

Rosen, Dan. “NHL announces salary cap for next season.” *NHL*, 22 June 2019,

<https://www.nhl.com/news/nhl-announces-salary-cap-for-2019-20-season/c-308008530>. Accessed 21 April 2020.

This source provided information about the National Hockey League salary cap, which is the ceiling on how much teams may spend in a year on player salaries. This information helped me to understand the money aspects of professional hockey and aided in putting a monetary value to the equation in regards to how it would be useful to team general managers, coaches, and scouts, and anyone else involved in deciding which players get signed where. This is a reliable source as it comes directly from a senior writer at the NHL.

Steinberg, Leigh. “CHANGING THE GAME: The Rise of Sports Analytics.” *Forbes*, 18 Aug.

2015, <https://www.forbes.com/sites/leighsteinberg/2015/08/18/changing-the-game-the-rise-of-sports-analytics/#a014aa44c1fd>. Accessed 20 April 2019.



Forbes is a trusted source with a good reputation. This article relates to the “Beyond Moneyball” article as they both deal with the increasing importance of analytics in sports. Both articles served to pique my interest on the topic that I would be researching and helped inspire the project, combining my love of both hockey and analytics. This article detailed a bit of the history in sports analytics and describes some of the newer trends and provides examples from multiple major sporting leagues.

“Teams.” *NHL*, <https://www.nhl.com/info/teams>. Accessed 15 April 2020.

This site provided detailed information on rosters for each team, including information about which players served as captains and alternate captains during the seasons. This source is accurate as it is coming directly from the National Hockey League. It ties in with the Hockey Reference source as both provided information on the leadership of the teams during the seasons in question. Though this site provides statistics of the players, it was not the site that was used to procure my data.

“The Business of Hockey.” *Forbes*, <https://www.forbes.com/nhl-valuations/list/>. Accessed 21 April 2020.

This source provided insight into how valuable different teams are across the National Hockey League. There are 31 teams listed here and they are ordered from greatest to least amount of revenue. This helped me understand how much revenue goes to each of the teams and how much money comes in and out of the sport of hockey. It also helped show how helpful the equation could be to the teams and their front office staff who decide which players get signed to the teams and how much they get paid. *Forbes* is a trusted business source.

“US Sports Betting Revenue And Handle.” *Legal Sports Report*, 23 April 2020,

<https://www.legalsportsreport.com/sports-betting/revenue/>. Accessed 23 April 2020.

This source discussed revenue from sports betting across the United States and discussed potential opportunities following the legalization of sports betting across the country. Prior to 2018, sports betting was not legal in all 50 states. It helped me to gain a better understanding of the money that goes into sports betting and the value that the equation might have in a monetary form. There were two other pages from the *Legal Sports Report* that were used in this project as well. This particular article also defined some terms that helped me have a better understanding of the way money flows through gaming organizations and helped me have a better understanding of those organizations and how they worked so that I would know when reading other articles about gambling and sports.

“2017-18 NHL Captains.” *Hockey-Reference*, [https://www.hockey-reference.com/leagues/NHL\\_2018\\_captains.html](https://www.hockey-reference.com/leagues/NHL_2018_captains.html). Accessed 15 April 2020.

This was a data source for discovering who the captains on each team were during the 2017-2018 NHL season. It ties in with the NHL source that also provided this information. This source did not match the NHL's data exactly and as the NHL source is the official website of the league, it would be the more accurate of the two. This source did not provide the names of the alternate captains on the teams, which meant it was not as useful as the NHL source.

"2017-2018 NHL Scoring Leaders." *Quant Hockey*, <https://www.quanthockey.com/nhl/seasons/2017-18-nhl-players-stats.html>. Accessed 2 March 2020.

This source has similar information to the Rotowire source in that it has many statistics from the 2017-2018 season that were used in the creation of the equation. This source was sought out for the two categories Rotowire lacked which were player age and country of origin. While these factors were not significant in the equation, they were still important to examine.