Relative Age Effects in the NHL

Fares Issa
Kaitlyn Richard
Brent Watling
Finnian Copp

Abstract:

The objective of this report is to prove the existence of the relative age effect in the NHL. Real world data was analyzed and the results were documented. These results served as the basis for a predictive model with simplified probabilities to ease the modelling of this system, proving the phenomenon in a simulation setting.

ACM Reference format:

2018. Relative Age Effects in the NHL. *RELATIVE AGE EFFECTS IN THE NHL* (April 2018), 7 pages.

1.0 Introduction

Studies in the past demonstrated relative age effects in different aspects of life, such as education, athleticism and mental health. [3]. These studies show those who are relatively younger within their age group are deemed less likely to succeed either due to age bias or actual disadvantage based on birth month [1]. This research implies that the birth quarter is directly correlated to the success of an individual; mainly in sports. While specific cause is unclear this phenomenon happens often in sports, especially hockey. A study was done on the National Hockey League (NHL) from 1980 to 2006 comparing those born in the first quarter to third and fourth quarter. The study found that those born in the first quarter are two times more likely to reach career milestones and those in the third and fourth quarter. Nearly 36% born in Jan to April while 14.5% from September to December [1]. Possible factors assuming everything is equal include that at a young age the youngest youth boy on a hockey team is smaller, weaker, less mature and less likely to be selected for elite teams [1]. In order to prove or disprove this research, data from the NHL was analyzed and studied and the results were discussed.

In the NHL, scouts and coaches tend to overlook players born in the later months of the year when it comes to drafting into the NHL. Many factors come into play with scouting decisions, some of these include the skill, height and strength of an individual. Two of these major factors are related to the biological development of the player and therefore, the younger players relative to the rest of those born in the same age group are at a clear disadvantage.

1.1 Relevance

The relative age effect is a phenomenon in hockey and other sports that while its effects are clear the cause is still variable. The purpose of this report is to prove the existence of bias in NHL based on age and birth quarter. Statistical analysis will be performed on the current data set of all players currently in the NHL. Once the correlation is proved on the real world data a predictive model is created to show this effect in a simulation setting. In Thinking Complexity Chapter 1.3 we follow the Predictive - explanatory model to assist with the assignment [2]. This model is used a simplified set of data and probabilities in order to predict the outcome of the effect of the month born on an athletes success in the NHL. Similarly, Chapter 4 focused on Scale-Free Networks which uses statistical analysis to show correlation in large data sets [2]. The aim is to take a large data set of all players and prove the correlation in month born and success in the NHL. This is similar to the Facebook data set correlating the degree of separation, however, in this case it's birth month to overall success in the NHL.

1.2 Problem Approach

The approach to this phenomenon was done in several steps. Firstly, the data set for all players currently in the NHL was parsed to a Jupyter Notebook. In order to prove this correlation, the data was analyzed and graphed to show generic trends between players born in different months. The data was an external .csv file that had all current NHL hockey players, with birthdate and country of origin. In order to make the data more succinct some assumptions were made. The year corresponding to the rate of birth year used was assumed to be the average age of the NHL players within data set. Thus, the average age of 28 was used, which corresponds to a birth year of 1990.

The first task was to separate the number of players by the month they were born in to see any trends within the NHL players. This is displayed below in Fig. 1.

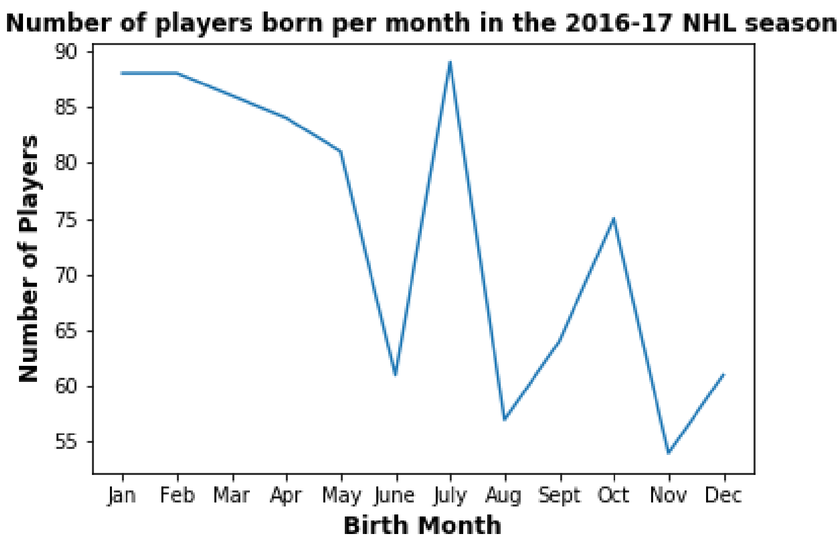

Fig. 1. NHL Players per Birth Month vs Total Players in the NHL

Studies suggest that there are a large number of players who are born in the first quarter of the year; January to April, this was evident in the data within this experiment. When compared to the fourth quarter; August to December, the totals are significantly lower [1]. For simplicity, this is further analyzed in the pie chart in Fig. 2 to more easily depict the divide. Fig. 2 shows that the highest percentage of athletes is 29.5% in the first quarter and 21.4% in the fourth quarter, this is approximately an 8% difference.

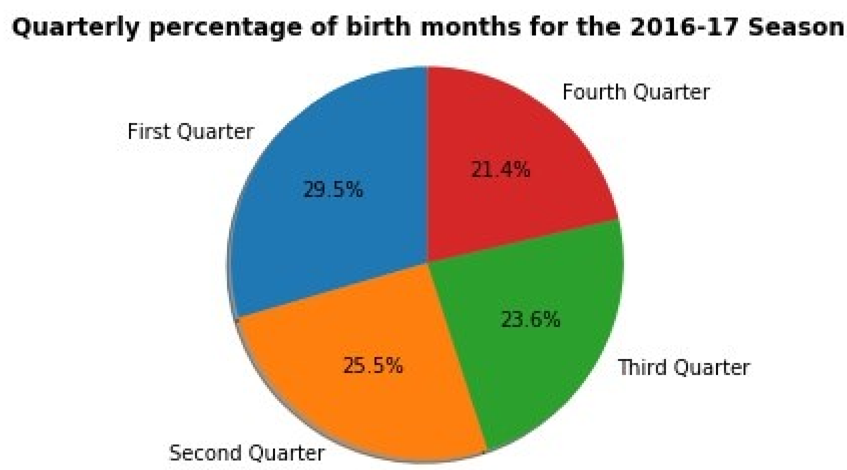

Fig. 2. Visual Representation of players born in each quarterly year

This number does not directly correlate with the nearly 36% born between January and April, while 14.5% from September to December from the article: *Born at the Wrong Time*, however, their data was done over a much larger scope between 1980 and 2006. The data obtained for this experiment simply looks at a single season [1]. Additionally, the relative age effect was revisited in a study focusing on players in the NHL between 2008 and 2009. The resulting analysis of the report had a slightly different composition for the first and fourth quarters, however, the overall data continued to show the existence of the Relative Age Effect in the NHL [6]. Similarly, while Fig. 2 is not a direct relationship to research studies, the overall age effect still holds true.

When proving a correlation, one must consider other external factors. In this case it cannot be assumed that there are the same number of births across each month. Therefore, using the average year of 1990 the average birth rate per month was calculated using data from United Nations and shown below in Fig. 3 [4]. The graph shows that February, November and December have relatively low birth rates when compared to the rest of the year.

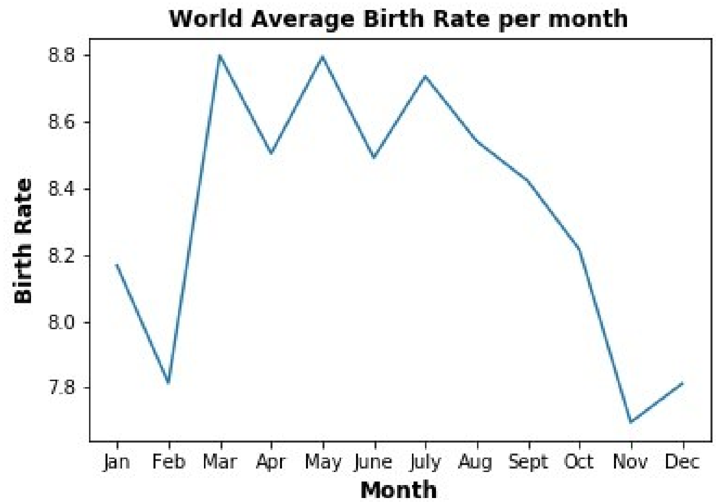

Fig. 3. World Average Birth Rate in 1990

Using this World Birth Rate Fig. 1 was adjusted to accommodate this factor and the adjusted graph is shown in Fig. 4 below. While there are some minor differences between the two graphs the overall trend remains. Those in the first quarter still outnumber players born in the fourth quarter.

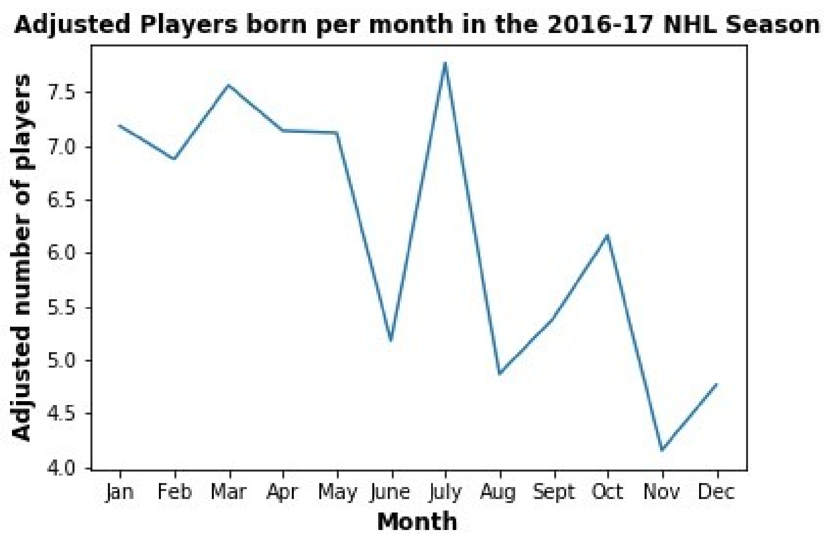

Fig. 4. NHL players per Birth Month with adjusted World Birth Rate

After the birth month divide was determined, the next step was to see this effect on a players' chances for success in their career, specifically the number of games played. Figure 5 depicts the total games played by the athletes, separated by their birth month. For the purposes of this report the primary focus is players born in the first quarter and fourth quarter of the calendar year. It is evidently shown below that there is a significant difference in the number of games played by those born in the first quarter compared to those born in the fourth quarter. This would be partially due to the number of players in each respective quarter. Additionally, studies suggest that there is a cutoff age within sports that exists over a calendar year. At a young age those born earlier in the year have a large overall advantage to those born later in the same year [5]. There is a large skew in youth and national league hockey players to the amount of players born in earlier years to those born in later years [5]. This cutoff is the primary cause to this relative age effect in sports, specifically in hockey [5].

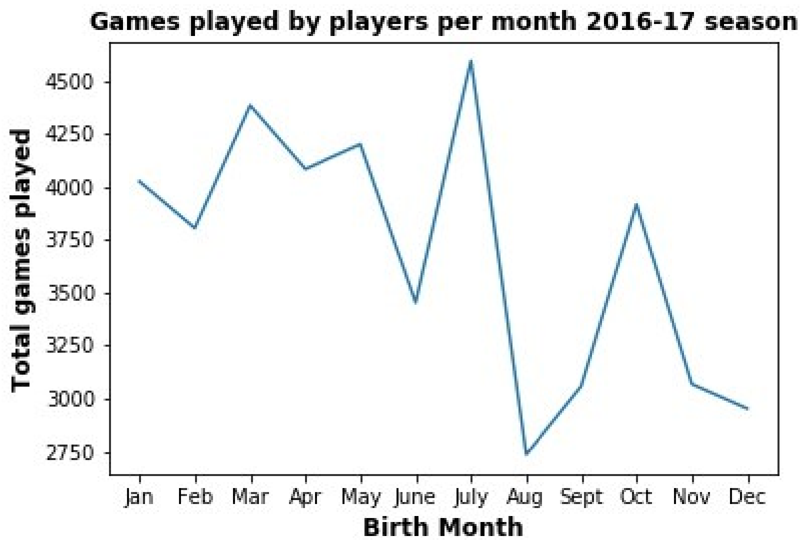

Fig. 5. Number of NHL games played vs Birth Month

2.0 Simulation Methodology

2.1 Statistical Models

In order to generate the statistical models above, we first needed to find data on the birth months of players currently in the NHL. We were able to find data on 888 NHL players from the 2016-17 season. From this data, we created a function to read the data from the .CSV file, and organize it into the number of players born in each month. From this we were able to create a plot that represents the number of players born per month, as well as a pie chart of the percentage of players born in each quarter.

We also found data on the birth rates of various countries for each month. Using another data reading function we created for this data, we found an approximate birth rate for each month for the world. This allowed us to create the plot in Fig. 4.

2.2 Predictive Models

In the predictive model, it was decided that the best way to determine whether a player will be drafted was to use an integer value to represent the players rating starting at a value of 10. After each time step, the player either gains or loses a rating point based on the chance of injury and birth month. Each time step represents one year, and once a player reaches 18, their rating will be compared to the 99.9th percentile. If their rating is above this value, they will be drafted, otherwise, they won't.

In order to create a hockey player, we created a class such that each instance of the class would store the attributes of each of the hockey players created in our simulation. Based on our class, each player has an injury rate, birth month and based on these two variables, their rating is determined. After each time step, the attributes that are defined and the players creation will be used to update the rating of the player.

The objective of this model was to make it as accurate as possible, so the decision to add two different factors that would affect the players' rating after each time step was met. The first factor is birth month and it will have the biggest factor of the two. The second factor will be a random injury chance.

The birth month of a player will be randomly assigned based on the percentage of current NHL players born in each month. The percentages will be taken from the statistical analysis we completed in the section above. In this simulation, a player will receive a rating of 1 for every month that is simulated. This method will give the players born in the first quarter a slight advantage over those in the last quarter.

For the random chance of injury, we found data on the number of injury's per season and used this to create an injury percentage per person. Using this, we can randomly determine if a player is injured after every time step. If a player is injured, they will lose 1 point from their rating, putting them at a slight disadvantage, but not disqualifying them from the race to be drafted.

3.0 Simulation Results

The results shown in Fig. 6 below shows a simulation of 1,200,000 players and the simulation was run for 18 time steps, where each time step corresponds to a year until the player is old enough to be drafted into the NHL at the age of 18. After running this simulation where all players start at a rating of 10, over 18 years, only 1657 have a likelihood of entering the NHL. This percentage was derived given an injury rate of 30.4% being implemented on the players within this model, where a player loses a rating point if they are injured at each time step. The injury rate is based on data collected from an NHL injury database [7]. As for the rating of the player, the percentage of NHL players born on each month is calculated, then a random number is generated, if this random number happens to be below the percentage of players in the NHL for that respective month, their rating increases by one.

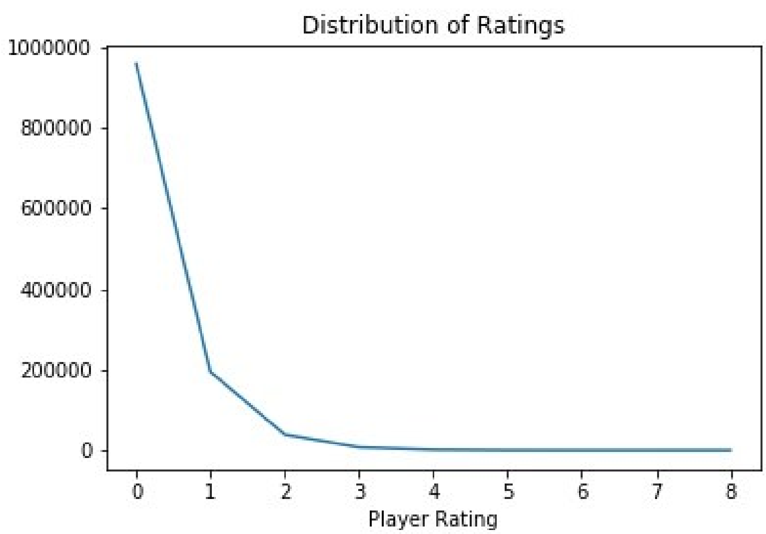

Fig. 6. Distribution of ratings for 12,000 players.

This simulation was further run to show of those who were drafted into the NHL; those scoring a high player rating within the 99.9th percentile, there are more players born in the earlier months of the year. This is seen in Fig. 7 below.

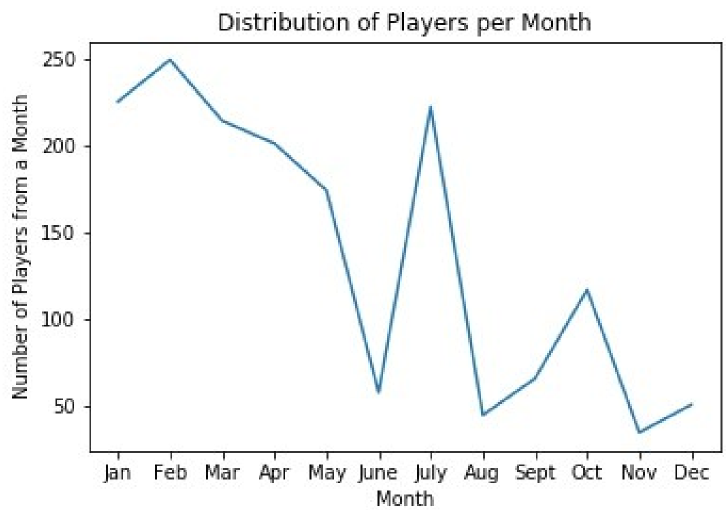

Fig. 7. Player rating based on birth month

4.0 Conclusion

The purpose of this project was to prove the Relative Age Effect in the NHL. Through statistical analysis of real world data, a simplified simulation was created. Multiple plots of the amount of players and the amount of games played showed that players born in the first quarter year had a much higher success rate than those born in the fourth quarter of a year. The simplified simulation proved that regardless of what month a player was born in the likelihood of that player making it to the NHL is very slim at only 0.138%. From the players who were selected to be drafted into the NHL, it is evident that there are more players born in the earlier months of the year compared to the later months. This trend proves that in a simulation setting that the Relative Age Effect holds.

5.0 References

[1] R. O. Deaner, A. Lowen, and S. Cobley, "Born at the Wrong Time: Selection Bias in the NHL Draft," Feb. 2013. [Online]. Available: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0057753 [Accessed: April 13, 2018].

[2] A. B. Downey, *Think Complexity 2e*. Sebastopol: OReilly, 2018. [Online].

[3] T. E. Elder, "The importance of relative standards in ADHD diagnoses: Evidence based on exact birth dates," *Journal of Health Economics*, vol. 29, no. 5, pp. 641–656, 2010. [Online]. Available: https://doi.org/10.1016/j.jhealeco.2010.06.003 [Accessed: April 13, 2018]

[4] "Live Births by Month of Birth," *United Nations*, 2018. [Online]. Available: http://data.un.org/Data.aspx?d=POP&f=tableCode:55. [Accessed: 13-Apr-2018].

[5] J. Musch and S. Grondin, "Developmental Review," *Unequal Competition as an Impediment to Personal Development: A Review of the Relative Age Effect in Sport*, vol. 21, no. 2, pp. 147–167, Jun. 2001. [Online serial]. Available: https://doi.org/10.1006/drev.2000.0516 [Accessed: April 13, 2018].

[6] J. E. Nolan and G. Howell, "Hockey success and birth date: The relative age effect revisited," *International Review for the Sociology of Sport*, vol. 45, no. 4, pp. 507–512, Jul. 2010. [Online]. Available: https://doi.org/10.1177/1012690210371560 [Accessed: April 13, 2018].

[7] LW3H, "NHL injury database (2009/10 to 2017/18)," *NHL Injury Viz*, 06-Nov-2015. [Online]. [Available]: http://nhlinjuryviz.blogspot.ca/2015/11/nhl-injury-database.html. [Accessed: 13-Apr-2018].