December 13, 2022

Modeling NHL Goaltender Performance

DATA 603 L02 Group 1
Fall 2022

Shawn Dornan, Jordan Keelan, Abrie Le Roux, Kevin Treptau, Arthur Trim

Table of Contents

INTRODUCTION 2
VARIABLES AND DATASET 4
METHODOLOGY 7
RESULTS 9
MODEL DIAGNOSTICS 14
MODEL TRANSFORMATION AND RESULTS 21
PREDICTION 24
CONCLUSION 27
DISCUSSION 28
REFERENCES 29

INTRODUCTION

Ice hockey is the most popular sport in Canada, with over 500,000 registered players in 2022 (Statistica, 2022). In addition, the sport's popularity is growing in the United States (600,000 registered players) and has over 300,000 registered players in various other countries worldwide. Specifically, the North American professional ice hockey league, the National Hockey League (NHL) is the 6th most profitable sports league in the world with profits of $4.3 billion in 2019 (Randjelovic, 2020). Worldwide, this is only surpassed by the National Football League, Major League Baseball, National Basketball Association, Indian Premier League (Cricket) and English Premier League (soccer).

Amongst all positions in ice hockey, no position garners as much scrutiny as the goaltender. The goaltender is on the ice for the entire game and is the last line of defense between the opposing team's shooters and the goal. One of the statistics most often used to describe goaltender performance is save percentage (SV%), which is simply the fraction of shots on goal saved over total shots on goal, faced by that goaltender. This stat is typically reported on a game-by-game, season-by-season, and overall-career basis. The save percentage of a teams' goaltenders will in turn determine the number of goals a team surrenders, and over the course of a season, will play a critical factor in the success or failure of the team.

Additionally, drafting and scouting goaltenders has proven to be difficult for NHL teams (in other words, teams have trouble accurately evaluating young goaltender performance). An analysis of goaltender performance relative to their draft round determined that the "most optimal time" to pick a goaltender is after the first round (Gross, 2013). This is due the outsized value a first round draft pick has to team, the impact a generational talent can have on a team, and the subsequent competitive downside of failing to use that first round pick on a valuable asset. Due to the fact that goaltender development timelines and performance prediction is less consistent than skaters, historical evidence has shown that picking a goaltender in the first round is not an optimal use of a team's draft capital.

In recent years, more thought and effort has been placed into applying advanced statistics to hockey player evaluation. Acknowledging that in a hockey game, goals are not a common event. Initially statistical measures, such as CORSI and FENWICK, were generated by looking at the shot attempts and successful shots in a game to determine how control of the puck possession is balanced between the opposing teams (Woodley, 2014). These early measures have evolved in recent years into more advanced calculated measures, such as "expected goals", "score adjusted", "per 60 metrics", and "goals above replacement", among others (Goldman & O'Connor, 2022). The key fact of these advanced statistical measures is that they mostly do not apply to goaltenders. There is another key difference for a goaltender in hockey, compared to other sports that have critical single players, such as a quarterback (in football) or a pitcher (in baseball), who gets to control and dictate the play, goaltenders are recipients of the play, and as such there is a huge impact on how the team plays in front of them and statistical outcomes" (Woodley, 2014).

In fact, the statistical evaluation of goaltenders is still relatively straightforward, relying on simple measures such as save percentage and average goals against to construct more complex metrics such as "goals saved above expectation, GSAx" (Goldman & O'Connor, 2022). GSAx is said to be the natural progression from SV%, reflecting the quality of the shots the goaltenders face (each shot is assigned a numerical value based on the chance it has of going in, "quality"), and their performance therein, as opposed to the bulk volume of shots and saves (Zirhen, 2021). Ultimately, professionals and amateurs alike are building models using available public data to endeavour to unravel these issues, in an attempt for a more analytical approach to goaltender evaluation.

What we have set out to accomplish is to take an approach to goaltender evaluation and performance prediction using data specific to individual goaltenders in a specific season and excluding various team- and game-based factors that could not only vary from year to year, but team to team, and game to game for a specific goaltender. If we are successful in our model building, we could provide a confidence range to the save percentage (SavePct) of a goaltender based on season-specific performance (time-on-ice, shoutouts, etc.), career (years in minors, draft round, etc.), and personal attributes (weight, birth country, etc.). In this way, we can provide an additional goaltender analysis tool that could complement more advanced generated statistical attributes and ultimately blend significant player-specific metrics with team- and situation-specific game metrics in overall goaltender evaluation.

Initially, we wanted to establish that the save percentage of a goaltender had a relationship to team success in each season (team points percentage is the possible team points over total possible points characterized with 2 points for a win, 1 point for overtime/shootout loss and 0 points for a regulation time loss). If we can establish a positive relationship (with 95% confidence) between these two variables (namely, Team Points % and Goaltender SV%), we have a reason (a "why?") for our statistical modeling of goaltender SavePct.

We were able to model team points percentage vs. team save percentage, using data from the 1997/1998 season to the 2018/2019 season (NHL.com, n.d.). This will be the same dataset that we will use in our future goaltender statistical modeling. The scatterplot with the fitted model is shown in Figure 1.

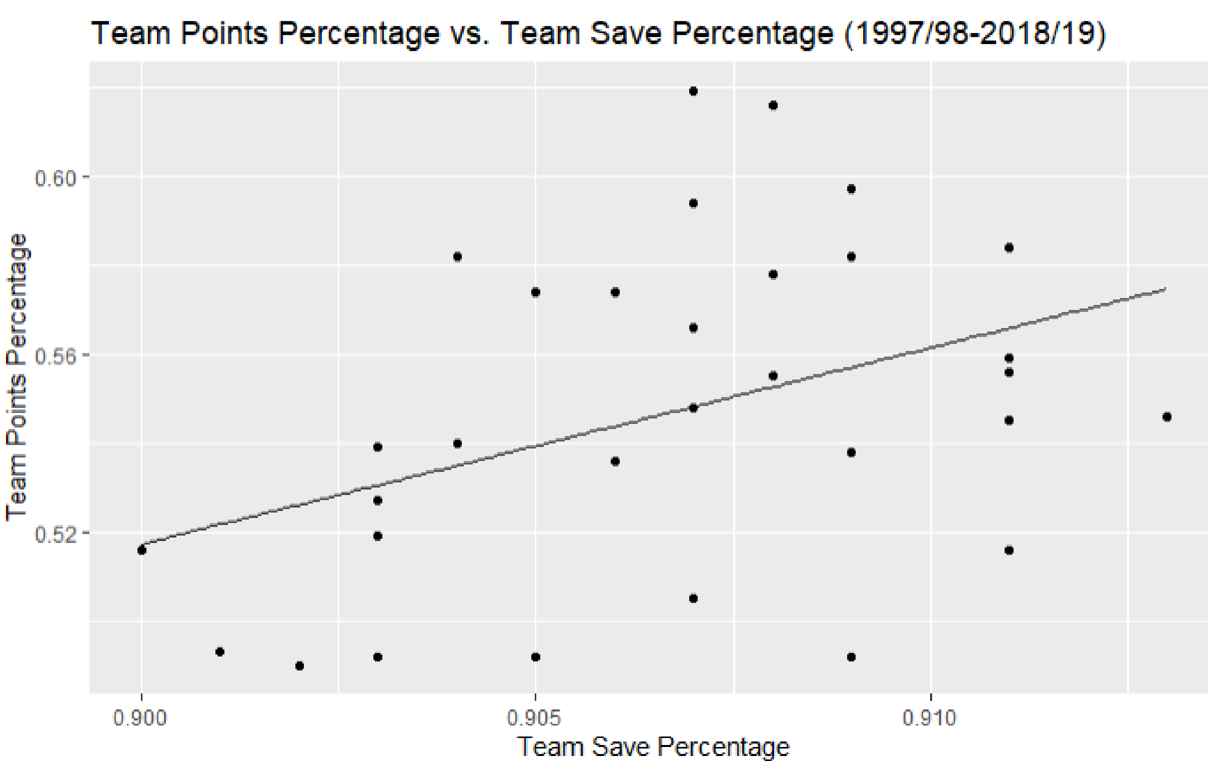


FIGURE 1: Scatterplot of Team Point % and Team Save % for the 1997/98 – 2018/19 seasons

We obtained a significant coefficient (P-value = 0.029) for the predictor variable "Team Save Percentage". The coefficient of the team SAVEPCT has a positive value of 4.407. In addition, the 95% confidence interval for the predictor coefficient ranges from 0.496 to 8.319. Therefore, we can say with a 95% confidence that there is a positive correlation between team points percentage and team SAVEPCT, with the formula is expressed below:

$$Poin\hat{t}PCT = -3.449 + 4.407SAVEPCT$$

As we can see that save percentage has a positive effect on a team's overall performance, we propose to determine if it is possible to predict a goaltender's save percentage based on various independent variables focusing on personal traits such as age, weight, height, country of birth, etc, excluding team performance variables.

VARIABLES AND DATASET

We collected our data in .csv format from the official website of the National Hockey League (NHL.com, n.d.). This website hosts open-source data from every season, dating back to the 1917-1918 seasons. This includes individual player statistics for both skaters and goaltenders, as well as team statistics.

As the data source contains both regular season and playoff statistics, the decision was made to analyze regular season data only. This is mainly because half of the teams in the league achieve playoffs each year, and the number of games played by any given team in the playoffs in unequal. One potential source of error in the data is inconsistent reporting, mainly due to technology changes and changes in the sport over the last 100 years. For example, video recording of games, paired with instant replay has allowed for statistics to be more accurately recorded in more recent decades. In addition, the NHL switched to an 82-game regular season in the 1995-1996 season. The number of games in a season is bound to affect player health and performance. On the contrary, sports science and player conditioning has improved vastly since the inception of the NHL, which will naturally result in more resilient, higher-performance athletes.

The dataset is generated by the league (NHL, n.d.) and is collected by team or league representatives for every game played under the purview of the league. As such, there are general standards that are league wide, but the actual recording of the statistical data in each game (shots, for example), are at the discretion of the individual collecting the data, which could impart some bias. The expectation is that over the course of a given season (82 games), the potential variability from rink-to-rink, or person-to-person will be averaged out. Personal player attribute data (height, weight, country, etc) are a single value for a single player throughout the length of their career (in other words, a player does not grow, or gain/lose weight once their biographical data is placed in the league record). In reality, players will change weights over the course of a career but given the fact that we are modeling professional athletes, we except that change to be minimal. After all, these players are paid large salaries to remain in shape and perform at the highest level.

The data used in the modeling was obtained from two different statistical reports provided from the NHL website, one report primarily focusing on the team statistics, including such information such as games played, recorded save percentage for a given season. The second report contained biographical data for each goaltender, including personal statistics like height, weight, country of birth, draft round and more. The websites built-in extraction application has a maximum export size into an excel sheet of 100 rows, we utilized UiPath web scraping for a more efficient method to export all datasets iteratively from 1995 to 2023 and automatically combine the exported data into a single raw datafile. These files consisted of 2396 rows of data for each statistical report, and would be cleaned and modified using R.

One of the main issues found when initially working with the unedited dataset was the misreporting of data within rows, which would impart Null and missing values into the data tables. It was found that the reason for most missing data was due to the undrafted players, containing empty data cells for the draft number variable. To handle this, rows which contained Null or missing values were removed entirely from the data frame. With a clean dataset, modification to the data frame continued in one of two ways, the first was to reformat variables into their proper data types. For example, the Time-On-Ice metric was recorded as a character with the format [minutes: seconds: milliseconds], therefore the character string was cut at the first identified colon, and casted as an integer resulting in a final transformed variable containing the value [minutes on ice]. The second modification made to the data frame was calculating and adding additional fields, primarily focusing on personal goaltender traits.

Each row of data was treated as a unique dataset, thereby if multiple seasons are played by one specific goaltender, each season became a unique value for the model. Static statistical variables were determined on a row-by-row basis, including their draft category – identifying what round of the NHL entry draft that they were selected (transformed into a categorical variable: early, middle, late), calculating the player's rookie year, and the number of years played in the minors (YIM: calculated by the difference of their first NHL season played and year in which they were drafted). Dynamic variables that were added to the data frame included their current age in a particular season, years since draft, years in the NHL, and their starting percentage (a ratio of games started divided by games played).

Additionally, a character variable identified as Era was input into the data frame and is used to identify different periods of time. The factor variable was generated to account for disruptions in the NHL season that resulted in less than 82 games played in that given season, and the introduction of new rules to the game that resulted in league-wide changes in game scoring dynamics. Specifically, the goaltender statistics relating to the NHL seasons 1994-1995 (48 games), 2004-2005 (0 games), 2012-2013 (48 games), and 2020-21 (56 games) were removed from the dataset to ensure that all goaltender season analysed contained the maximum possible games played as 82 (in otherwards, a "full" season). Subsequently, the Era factor variable was subdivided into continuous blocks of seasons to determine if there could be any significant difference to goaltenders during these times (Era = 1997/98-2003/04, 2005/06-2011/12, and 2013/14-2019/20). In this way, we have attempted to deal with the possible data bias imparted due to changes in the abnormal season length, game rules, player technology, player conditioning and the implementation of the salary cap.

Lastly, we wanted to establish a way to evaluate "statistically significant" goaltenders in a season, or what would be characterized as NHL-caliber goaltenders. This was an attempt to eliminate randomness in the data from sample goaltenders in each season who only played a few games, and whose performance might be more the affect of the situation in the given game, rather than the player themselves. We removed any goaltender who played less than 25 games in a season from the dataset. This is justified by the median [games_played] of the initial dataset being 25, the mean of [games_started] = 26.9, and the NHL rule that for a goaltender to be qualified of the Vezina trophy (awarded to the “to the goalkeeper adjudged to be the best at this position”) being set at 25 games (NHL.com, 2022). Subsequently, we chose 25 games as a reasonable cut-off for trimming our dataset to “NHL starting goaltenders”, and after removing the rows, the data frame was re-indexed. The effect of this 25-game cut-off is shown in Figure 2 with a scatter plot of Games_Played and SavePct before and after the cut-off was applied. One can observe the funnel-shaped scatter of data on the left plot starting at games_played < 20, and the relatively random pattern of the data distribution in the right scatter plot. Additionally, the removal data still resulted in a dataset size of 985 individual goaltender season entries to undertake our model building effort on.

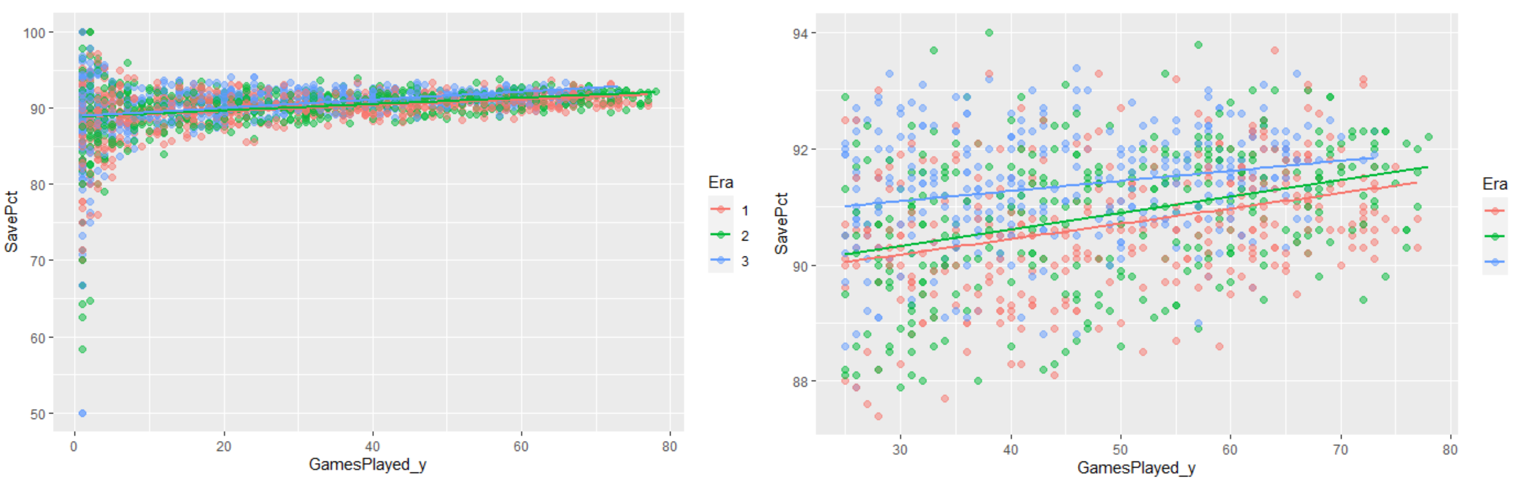


FIGURE 2: Plot of Games Played vs Save % before (left, n=1939) and after (right, n=985) 25 game cut-off

To effectively model and predict goaltender performance, we selected the following variables for investigation. Every row in the dataset is a given set of variables for a particular goaltender in a specific season. Thus, a goaltender who played 10 seasons in the NHL would be 10 separate row entries (given they have played > 25 games in each season), and each season can be treated independently, regardless of the team they played on, while other attributes remain constant (country of birth, for example).

The source CSV files had numerous other columns which we dropped from the model data frame, or we used to calculate new variables more suited to our analysis.

METHODOLOGY

DATASET VARIABLES

The following variables were **extracted** from the source data

1. **SavePct - Goaltender save percentage: The percentage of saved shots divided by total shots on goal faced by the goaltender. *(Dependent Variable)***
    - We multiplied SavePct by 100 to convert the source decimal value (0-1) into a percentage value that would have a value > 1
    - **Important Note:** The range of values for goaltender SavePct in our dataset all fall between 88-94%. So, despite the issue of performing linear regression on a proportion response variable, we do not expect our model to fit values that fall so far outside this range that they are impossible (ie. Fit > 1 OR Fit < 0)
2. HeightInches – Goaltender height in inches. *(Independent Variable)*
3. WeightPounds – Goaltender weight in pounds. *(Independent Variable)*
4. Country – Goaltender country of birth. *(Independent Categorical Variable)*
    - Represented in the data frame as a FACTOR with 5 levels
    - This was transformed into the following categories: “Canada”, “USA”, “Finland”, “Sweden” and “Other”
    - During our data wrangling, we identified 19 different countries in the source data, and created a set of factor variables based on the “top 4” countries in the data, and the grouped the remainder of the countries into “Other”
5. ShootsCatches_x – Which hand does the goaltender catch with, “L” or “R”? *(Independent Categorical Variable)*
    - The goaltender equipment consists of a blocker in one hand (holding the stick) and a glove in the other hand, this is the catching hand
6. PenaltyMinutes – How many penalty minutes did the goaltender accumulate that season? *(Independent Variable)*
7. TOI (time on ice) – Minutes played that season, floor rounding was applied to remove second and milliseconds, the minute was rounded to the nearest integer, equal to or below the actual value. *(Independent Variable)*
8. Shutouts_y – How many full-length games the goaltender played that season in which he did not allow any goals. *(Independent Variable)*

The following variables were **calculated** from existing variables

9. Age – Goaltender age in years. *(Independent Variable)*
    - Current season minus year of birth
10. DraftCategory – What round was the player drafted? Perceived “better” potential players are typically drafted earlier. *(Independent Categorical Variable)*
    - Round (1-2), Round (3-4), LaterRound (5-10), or Undrafted - ("EarlyRound", "MidRound", "LateRound", "Undrafted")
11. YSD (years since drafted) – How many years since this player was drafted? *(Independent Variable)*

    - Current season – Draft Year
12. YIM (years in minors) – How many years did this player spend in the minor leagues prior to playing their “first season” in the NHL? *(Independent Variable)*
    - First Season – Draft Year
13. Era – Which “Era” did this player play in? *(Independent Categorical Variable)*
    - 1996/97-2003/04, 2005/06-2011/12, 2013/14-2018/19 - ("1", "2” "3"; the breakdown of games per era is - Era 1: 338, Era 2: 325, Era 3: 322)
14. StartPct – What percentage of games did the player “start” the game as the goaltender? *(Independent Variable)*
    - If a goaltender performs poorly in a specific game, or is injured, they can be replaced by the back-up goaltender at the coach’s discretion

<u>SOURCE OF THE DATA</u>

- *Source of Data*:
    - Player and team statistical data is provided by the National Hockey League (NHL.com, n.d.)

- *Data Permission:*
    - NHL.com - You may access, use, and display the Services and print copies of the NHL Content only for non-commercial, informational, personal use, without modification or alteration in any way, and only so long as you comply with these Terms.

<u>APPROACH AND WORKFLOW</u>

The multiple linear regression model was fit using methods learned in DATA 603. Our initial, full model was composed of all 13 variables available. Iteratively, non-significant variables (α=0.05) with the highest p-value were removed. Until we were left with a first order model containing only significant predictors.
To improve this model, we started over with our full, 13 variables, model and applied interaction terms to all variables. Again, we iteratively removed main effects and interaction terms, starting with the largest p-values until settling on a valid model composed of significant (α=0.05) predictors and predictors included via hierarchy principal.

Once the final model was determined and all appropriate diagnostics/transformations were completed, the model was used to test predictability.

<u>DISTRIBUTION OF WORKLOAD</u>

Each team member contributed to all components of this analysis. Work was completed in person with all group members present.

RESULTS

FIRST-ORDER MODEL

The first order model that was generated from the individual t-tests was the following:

$$H_0 : \beta_i = 0$$
$$H_A : \beta_i \neq 0$$
$$i = predictor\ variables$$

$$H_0 : \beta_1 = \beta_2 = \beta_i = 0$$
$$H_a : \text{at least one } \beta_i \text{ is not zero}$$
$$i = x, y, z, \ldots [predictor\ variables]$$
$$\alpha = 0.05$$

Predictors were excluded from the model at a significance of α=0.05.

To test the impact of our key data wrangling effort of limiting the data to [games_played] ≥ 25, we ran a test of full model build using every variable on both datasets (n=1939 & n=985). The output statistics of the two models are below:

The full first-order model from full dataset:

- Adjusted R-squared value = 0.0805
- RMSE = 3.597%

The full first-order model from trimmed dataset yielded:

- Adjusted R-squared value = 0.3518
- RMSE = 0.8997%

By removing goaltenders with < 25 game played, we can see that the model has improved performance prior addressing non-significant variables in the dataset. After this we built the best first-order model based on all significant variables.

$$\widehat{SVPCT}_i = \begin{cases} 89.8055 + 0.2601 Shutouts & \text{if } i^{th} \text{ ERA is 1 (1996/97 - 2003/04)} \\ 90.0798 + 0.2601 Shutouts & \text{if } i^{th} \text{ ERA is 2 (2005/06 - 2011/12)} \\ 90.1156 + 0.2601 Shutouts & \text{if } i^{th} \text{ERA is 3 (2013/14 - 2018/19)} \end{cases}$$

The “best” first-order model yielded:

- Adjusted R-squared value = 0.3381
- RMSE = 0.9199%

INTERACTIVE MODEL

Next, we undertook an iterative approach to building a model with interactions. We decided to approach interaction next as: firstly, were not fully satisfied with a model that included only two variable predictors in our dataset that resulted from the first-order approach; and secondly, institutional knowledge about hockey would indicate many traits of a goaltender and how they influence performance over a season would interact, as shown in Figures 3 and 4.

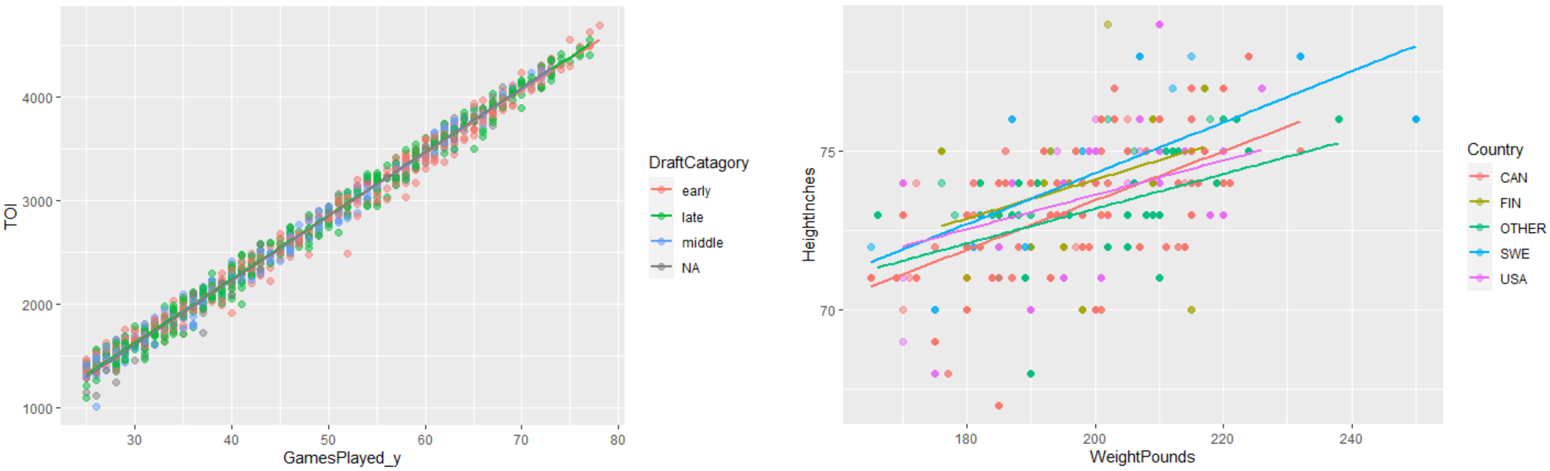


FIGURE 3: Plot of Games Played vs TOI (left) and Weight vs Height (right) to illustrate the different impacts of specific independent variables on the dependent variable, for different values of a different independent variable.

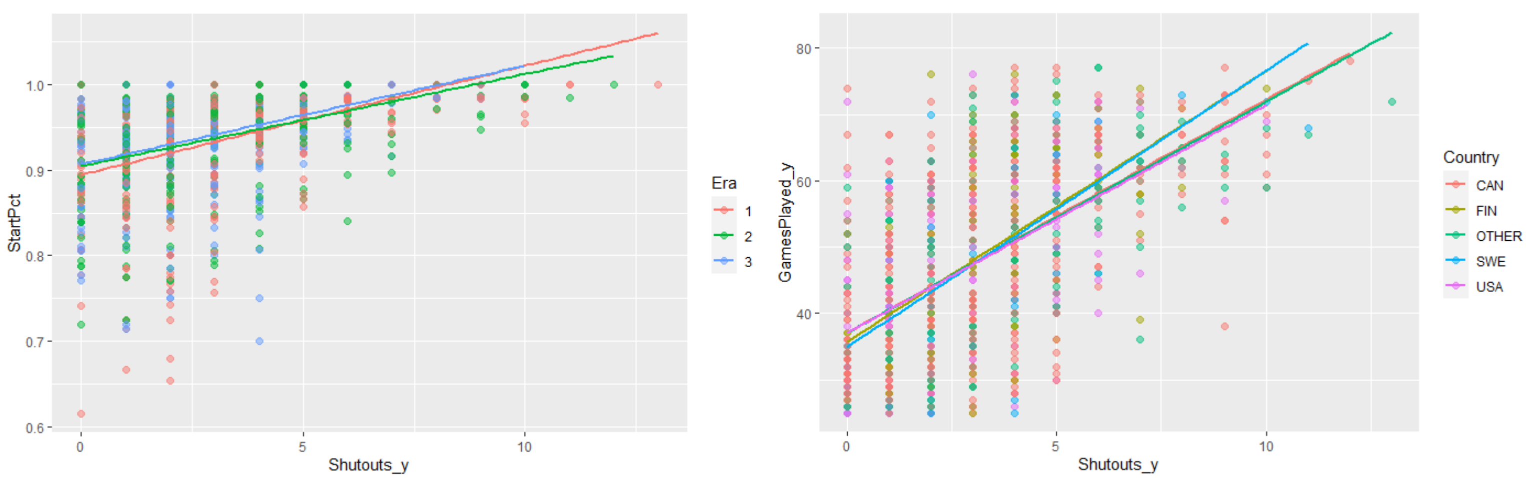


FIGURE 4: Plot of Shutouts vs StartPct (left) and Shoutouts vs GamesPlayed (right) to illustrate the interactions.

The iterative approach to finding the best interactive model yielded the following model (Figure 5):

```
Call:
lm(formula = SavePct ~ Shutouts_y + factor(Era) + factor(Country) +
    factor(DraftCatagory) + YIM + WeightPounds + TOI + factor(DraftCatagory):factor(Era) +
    factor(DraftCatagory):YIM + factor(Country):factor(Era) +
    factor(Country):WeightPounds + TOI:Shutouts_y, data = goalies_new)

Residuals:
     Min       1Q   Median       3Q      Max
-2.83410 -0.60151 -0.00723  0.59493  2.71598

Coefficients:
                                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                              8.716e+01  6.398e-01 136.232  < 2e-16 ***
Shutouts_y                               5.291e-01  5.374e-02   9.845  < 2e-16 ***
factor(Era)2                             2.712e-01  1.226e-01   2.211 0.027287 *
factor(Era)3                             7.877e-01  1.361e-01   5.788 1.01e-08 ***
factor(Country)FIN                       4.063e+00  1.717e+00   2.366 0.018223 *
factor(Country)OTHER                     1.685e+00  1.157e+00   1.456 0.145703
factor(Country)SWE                       2.337e+00  1.407e+00   1.660 0.097213 .
factor(Country)USA                       2.389e+00  1.128e+00   2.117 0.034572 *
factor(DraftCatagory)late               -6.176e-02  1.705e-01  -0.362 0.717281
factor(DraftCatagory)middle              1.411e-02  1.708e-01   0.083 0.934185
YIM                                     -3.729e-02  3.687e-02  -1.011 0.312131
WeightPounds                             1.016e-02  3.190e-03   3.186 0.001498 **
TOI                                      2.632e-04  6.156e-05   4.274 2.14e-05 ***
factor(Era)2:factor(DraftCatagory)late  -5.213e-01  1.778e-01  -2.931 0.003472 **
factor(Era)3:factor(DraftCatagory)late  -3.355e-01  1.888e-01  -1.776 0.076027 .
factor(Era)2:factor(DraftCatagory)middle -6.965e-01  2.192e-01  -3.177 0.001544 **
factor(Era)3:factor(DraftCatagory)middle -1.850e-01  1.972e-01  -0.938 0.348498
factor(DraftCatagory)late:YIM            9.417e-02  4.508e-02   2.089 0.037003 *
factor(DraftCatagory)middle:YIM          5.679e-02  5.365e-02   1.059 0.290135
factor(Era)2:factor(Country)FIN         -6.969e-01  3.695e-01  -1.886 0.059673 .
factor(Era)3:factor(Country)FIN         -8.405e-01  3.772e-01  -2.228 0.026133 *
factor(Era)2:factor(Country)OTHER        2.167e-01  2.082e-01   1.041 0.298255
factor(Era)3:factor(Country)OTHER        5.509e-02  2.188e-01   0.252 0.801281
factor(Era)2:factor(Country)SWE          8.317e-01  3.925e-01   2.119 0.034373 *
factor(Era)3:factor(Country)SWE          1.081e+00  4.117e-01   2.626 0.008809 **
factor(Era)2:factor(Country)USA          8.200e-01  2.142e-01   3.827 0.000139 ***
factor(Era)3:factor(Country)USA          4.737e-01  2.166e-01   2.187 0.028985 *
factor(Country)FIN:WeightPounds         -1.604e-02  8.600e-03  -1.866 0.062451 .
factor(Country)OTHER:WeightPounds       -8.429e-03  5.775e-03  -1.460 0.144760
factor(Country)SWE:WeightPounds         -1.550e-02  7.440e-03  -2.083 0.037597 *
factor(Country)USA:WeightPounds         -1.330e-02  5.773e-03  -2.303 0.021514 *
Shutouts_y:TOI                          -8.828e-05  1.601e-05  -5.516 4.64e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8736 on 830 degrees of freedom
  (123 observations deleted due to missingness)
Multiple R-squared:  0.4108,    Adjusted R-squared:  0.3888
F-statistic: 18.67 on 31 and 830 DF,  p-value: < 2.2e-16
```

FIGURE 5: Model output and coefficients for best interactive model

In this final interactive model, all variables were significant, or included in the model as they were part of a significant interaction term. The model yielded an Adjusted R-squared value of 0.3888 and a RMSE of 0.8736%. This model was used to check for higher order terms and to run full diagnostics assumptions checks.

Our final fitted model consists of 4 independent quantitative variables, 3 factor variables, and 5 interaction terms is reported in Figure 6:

- Shutouts
- YIM
- Weight
- TOI
- Factor (Era) – 3 levels
- Factor (Country) – 5 levels
- Factor (DraftCategory) – 3 levels
- Shutouts * TOI (interaction)
- Era * DraftCategory (interaction)
- DraftCategory * YIM (interaction)
- Era * County (interaction)
- Country * Weight (interaction)

```
                          (Intercept)                            Shutouts_y                        factor(Era)2                          factor(Era)3
                         8.715889e+01                          5.290569e-01                        2.711822e-01                          7.876982e-01
                  factor(Country)FIN                  factor(Country)OTHER                  factor(Country)SWE                    factor(Country)USA
                         4.062684e+00                          1.684564e+00                        2.336889e+00                          2.388523e+00
            factor(DraftCatagory)late           factor(DraftCatagory)middle                                 YIM                          WeightPounds
                        -6.175927e-02                          1.410992e-02                       -3.728579e-02                          1.016369e-02
                                  TOI  factor(Era)2:factor(DraftCatagory)late factor(Era)3:factor(DraftCatagory)late factor(Era)2:factor(DraftCatagory)middle
                         2.631494e-04                         -5.212621e-01                       -3.354643e-01                         -6.964826e-01
factor(Era)3:factor(DraftCatagory)middle     factor(DraftCatagory)late:YIM     factor(DraftCatagory)middle:YIM        factor(Era)2:factor(Country)FIN
                        -1.849962e-01                          9.417021e-02                        5.678907e-02                         -6.968860e-01
      factor(Era)3:factor(Country)FIN     factor(Era)2:factor(Country)OTHER     factor(Era)3:factor(Country)OTHER        factor(Era)2:factor(Country)SWE
                        -8.404557e-01                          2.167438e-01                        5.509127e-02                          8.317028e-01
      factor(Era)3:factor(Country)SWE       factor(Era)2:factor(Country)USA       factor(Era)3:factor(Country)USA      factor(Country)FIN:WeightPounds
                         1.080940e+00                          8.199799e-01                        4.737210e-01                         -1.604456e-02
 factor(Country)OTHER:WeightPounds       factor(Country)SWE:WeightPounds      factor(Country)USA:WeightPounds                       Shutouts_y:TOI
                        -8.429158e-03                         -1.549516e-02                       -1.329702e-02                         -8.827919e-05
```

FIGURE 6: Model coefficients for best interactive model

The “best” interactive model yielded:

- Adjusted R-squared value = 0.3888
- RMSE = 0.8736%

We can see that by undertaking an iterative approach to building an interactive model, we improved the adjusted R^2 (0.3381 to 0.3888) and lowered the RMSE (0. 9199% to 0.8736%) of the model output, resulting in a more accurate predictive model. In addition, the interactive model incorporates more variables from the dataset, based on the hierarchical principle of including non-significant terms that have significant interactions. The possibility of higher order terms in our model was then assessed with a pairs plot (Figure 7).

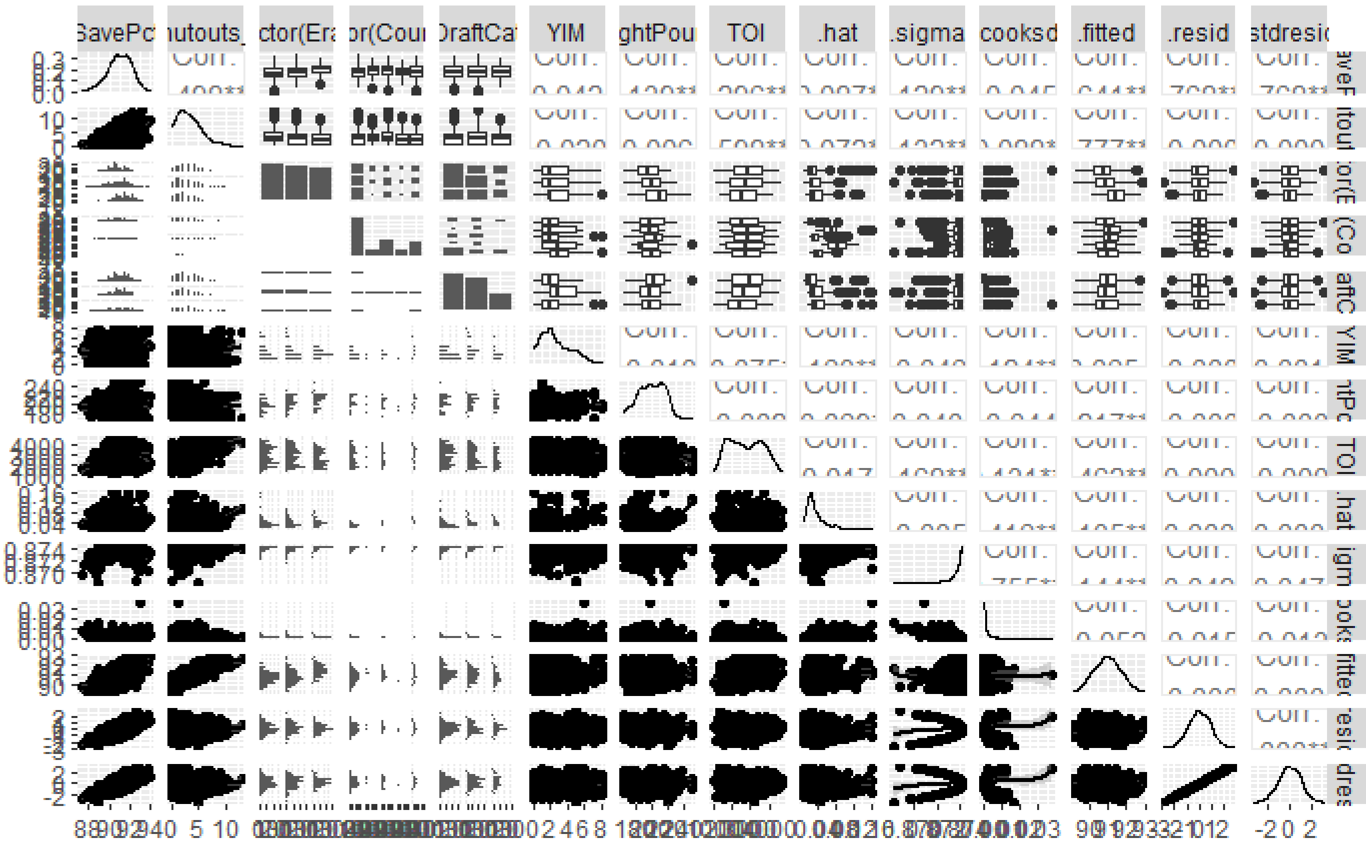


FIGURE 7: Plot of predictor and variables

Examination of the scatterplot matrix of predictor and variables, allow for visualizing the correlation between variables (Figure 7) and does not show any patterns that would reason to explore the possibility of the model requiring polynomial terms.

In addition to no evidence in the scatter plots suggesting a higher order relationship between SavePct and any of the predictors; we know of no contextual reason that a higher order model should be investigated. This exhausts the resources available to the scope of this report for fitting our linear regression model. Next, we will perform regression diagnostics to ensure our model satisfies all necessary assumptions for its results to be trustworthy.

MODEL DIAGNOSTICS

SUMMARY OF MODEL ASSUMPTIONS

To assess the validity and trustworthiness of the model, it is necessary to validate six assumptions that the model must satisfy. These assumptions and determining tests are:

1. **Linearity:** Residuals vs. Fitted Plot
2. **Independence:** Not time series data, N/A
3. **Equal Variance Assumption:** Residuals vs. Fitted Plot and Breusch-Pagan test
4. **Normality of Error Terms:** Q-Q plot, histogram of residuals and Shapiro-Wilk normality test
5. **Multicollinearity:** Variance inflation factors (VIF). This was performed on the main effects from the final interactive model.
6. **Influential Points and Outliers:** Cook's distance and leverage distances

Note: all statistical hypothesis' testing was based off α = 0.05.

LINEARITY ASSUMPTION

Our model assumes that there is a straight-line relationship between the predictors and the response. Looking at the residuals vs. fitted plot (Figure 8), there is no discernable pattern, confirming the linearity assumption.

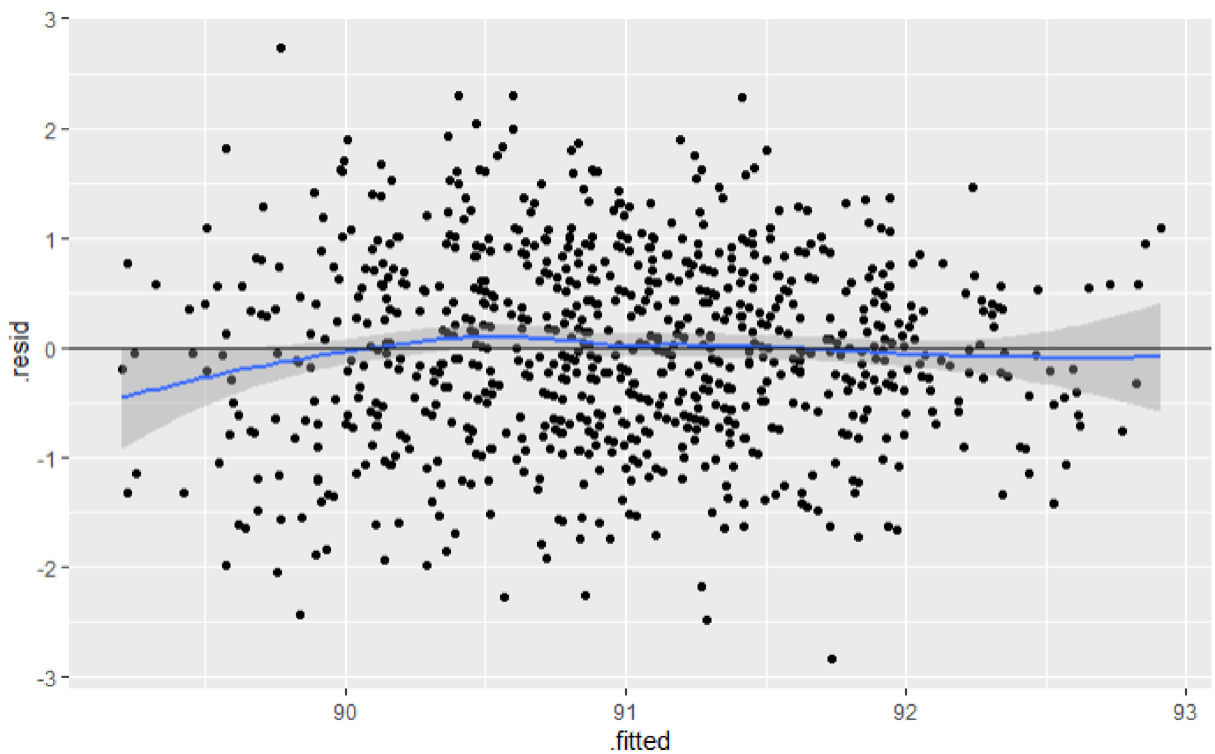


FIGURE 8: Residuals vs Fitted plot

INDEPENDENCE
As the data is not related to time (time has been modified into a categorical variable, categorizing the dataset into three different Eras).

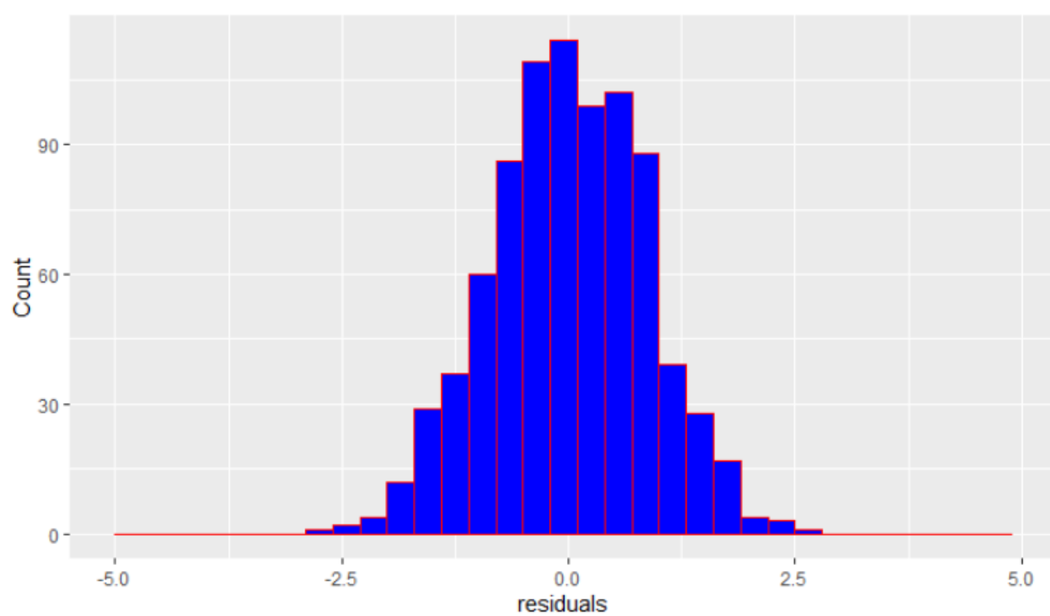
EQUAL VARIANCE ASSUMPTION
The residual vs. fitted plot (Figure 8, above) suggests the dataset is randomly spread, however this assumption can be quantitively tested with the Breusch-Pagan test to provide a definitive answer for the equality of variance spread. This is a measurement of how similar variance is across all values of the independent variables.

$$H_0 = heteroscedasticity\ is\ not\ present\ (homoscedasticity)$$
$$H_A = heteroscedasticity\ is\ present$$

```
        studentized Breusch-Pagan test

data:  best.model
BP = 61.722, df = 31, p-value = 0.000842
```

The Breusch-Pagan test yielded a P-value of 0.000842, indicating that we should reject the null hypothesis stated above. **This indicates that the dataset does show heteroscedasticity.** Later in our methodology, we will be exploring a removing any outliers and a Box-Cox transformation to improve the homoscedasticity.

NORMALITY ASSUMPTION
Both the histogram of residuals and the normal probability plot (Figure 9) indicate normality of error terms.



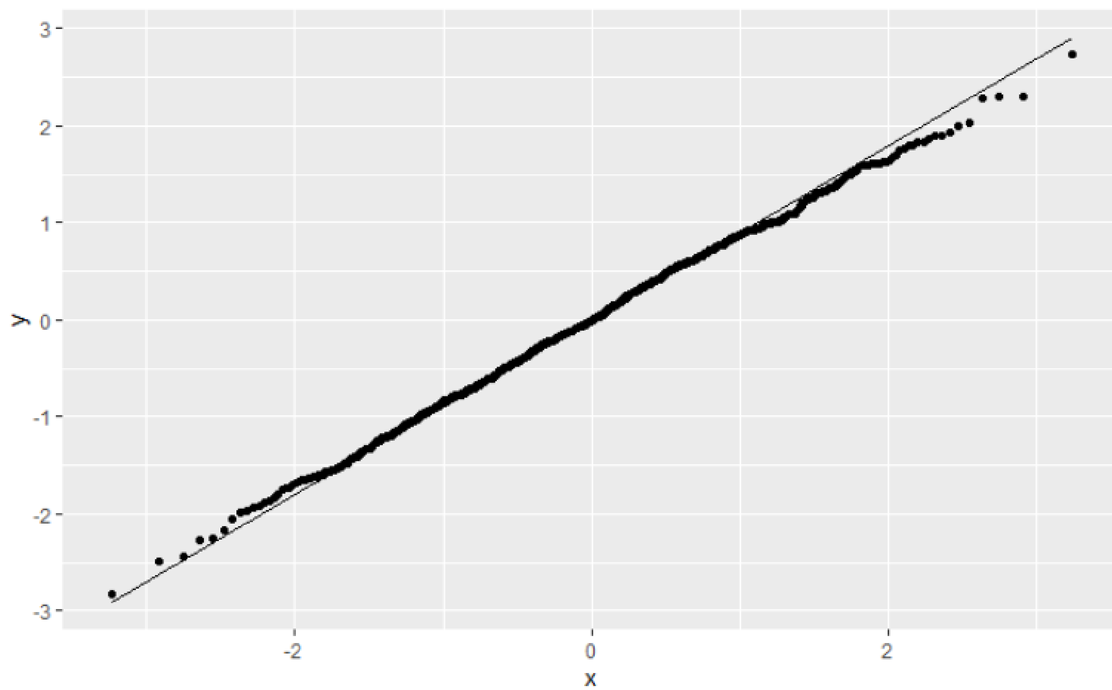


FIGURE 9: Histogram of residuals (above) and Q-Q normal probability plot (below)

Additionally, the Shapiro-Wilks test to test for normality using the following statistical hypothesis.

$$H_0 = sample\ data\ are\ significantly\ normally\ distributed$$
$$H_A = sample\ data\ are\ not\ significantly\ normally\ distributed$$

```
        Shapiro-Wilk normality test

data:  residuals(best.model)
W = 0.99871, p-value = 0.8012
```

The Shapiro-Wilks test yielded a P-value of 0.8012, indicating that we fail to reject the null hypothesis stated above. Thus, confirming that the residuals are normally distributed.

MULTICOLLINEARITY ASSUMPTION

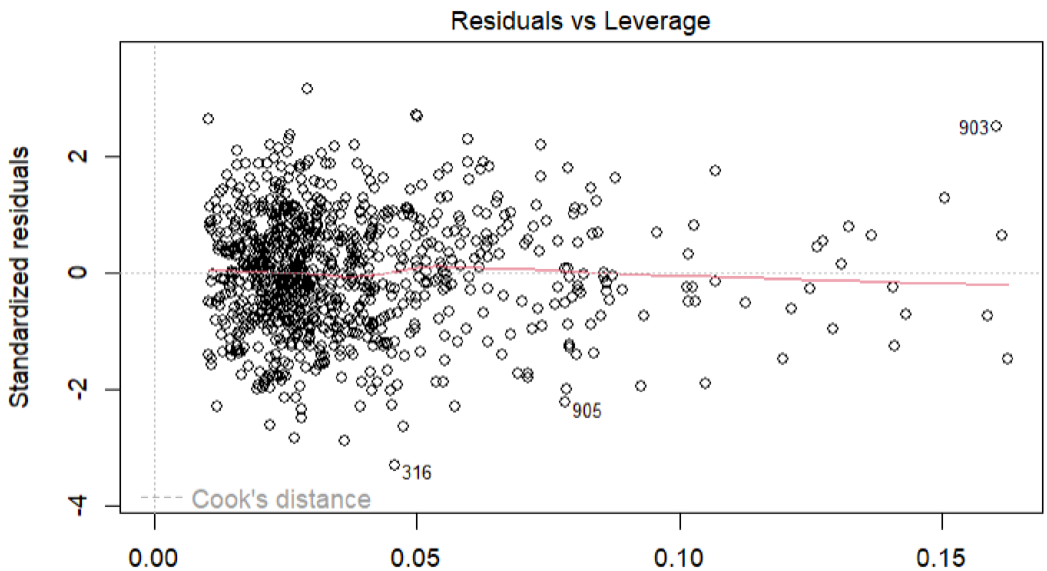
We undertook the VIF test on the main independent variables that were present in our final interactive model (Figure 10) to test for multicollinearity. The result is that the main effects do not yield any variance inflation factors to suggest multicollinearity.

```
Call:
imcdiag(mod = goalies.maineffects, method = "VIF")


 VIF Multicollinearity Diagnostics

                                 VIF detection
Shutouts_y                    1.5915         0
factor(Era)2                  1.3790         0
factor(Era)3                  1.5360         0
factor(Country)FIN            1.1358         0
factor(Country)OTHER          1.2773         0
factor(Country)SWE            1.1420         0
factor(Country)USA            1.2051         0
factor(DraftCatagory)late     1.6725         0
factor(DraftCatagory)middle   1.2355         0
YIM                           1.3239         0
WeightPounds                  1.1577         0
TOI                           1.6117         0

NOTE:  VIF Method Failed to detect multicollinearity


0 --> COLLINEARITY is not detected by the test
```

FIGURE 10: VIF Diagnostics of final interactive model

OUTLIERS AND INFLUENTIAL POINTS

The residuals vs. leverage plot shows no residuals approaching the Cook's Distance lines of 0.5 or 1.0. The highest Cook's distance in our dataset approaches 0.04 (Figure 11). This outcome was not surprising, as one would presume that that the athletes playing in the NHL are the elite-level of all that participate in the sport professionally and would be reasonably expected to result in less significant outliner. In addition, the decision we made during wrangling to only include the goaltenders with ≥ 25 games played in a season predictably resulted in a dataset that had removed one of the primary sources of outliers in the dataset.



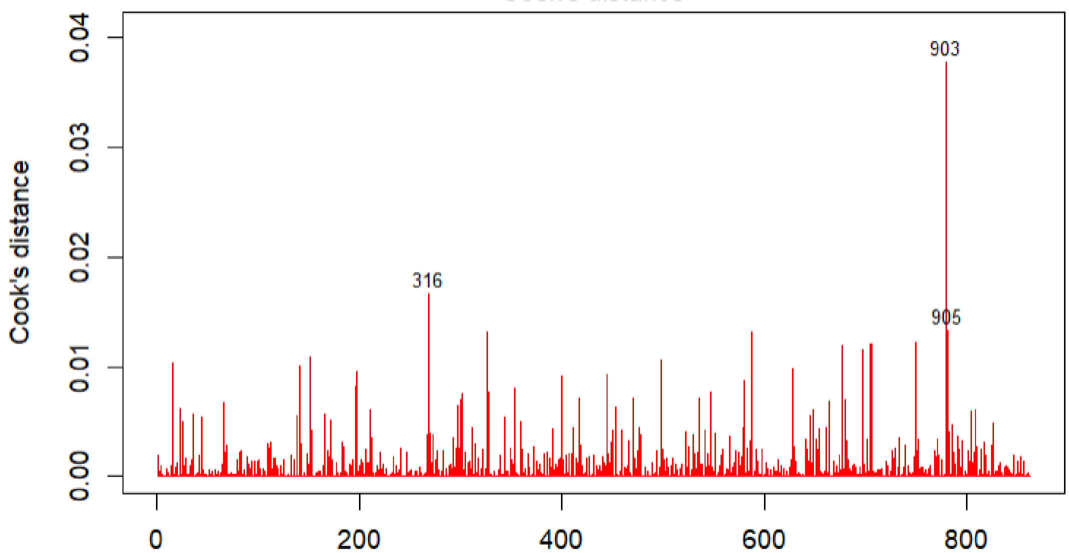
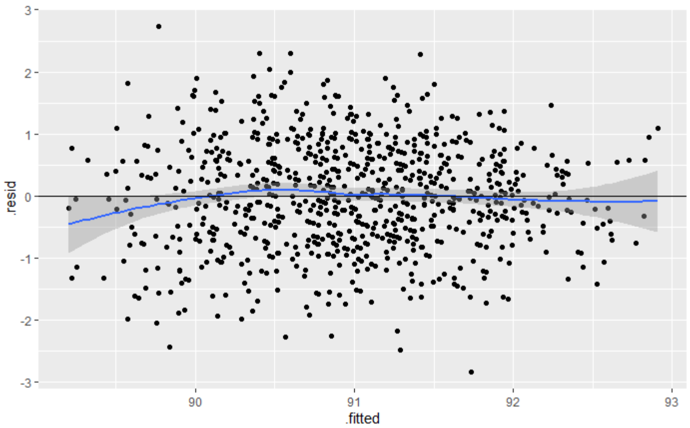
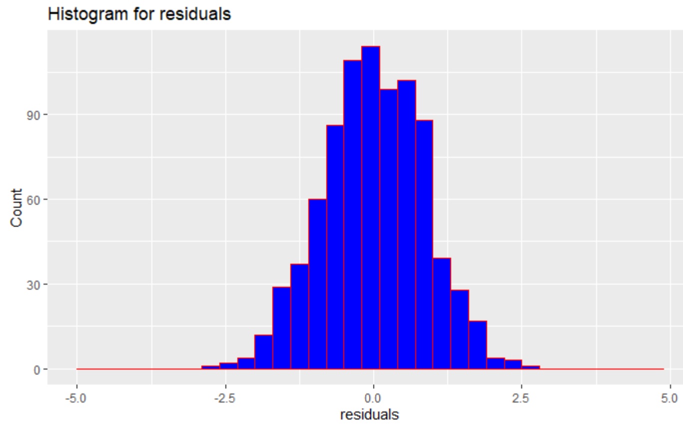


FIGURE 11: Standardized residual vs leverage plot and Cook's Distance number.

Understanding this, in the effort to continue to improve our models and satisfy all the assumptions to the best of our efforts, we did want to test if we could improve our dataset further, limiting any potential influence of any outliers in the dataset, even after trimming. We ran a test of outliers that had a Leverage Distance of 3p/n and 2p/n to understand what outliers may still exists. We tested the removal of both the 3p/n and 2p/n groups of outliers and found that when we removed the 3p/n outliers' group (total n = 27 influential observations) from the dataset, we improved the model performance. Conversely, the removal of the "2p/n group" reduced the model performance. **Therefore, the dataset was modified to remove these influential points defined by the cut-off of 3p/n.** The previous-discussed diagnostic tests were again performed on the final model from this modified dataset, with the following results displayed in Figure 12.






```
        studentized Breusch-Pagan test

data:  best.model
BP = 59.81, df = 29, p-value = 0.0006522


        Shapiro-Wilk normality test

data:  residuals(best.model)
W = 0.99868, p-value = 0.8059
```

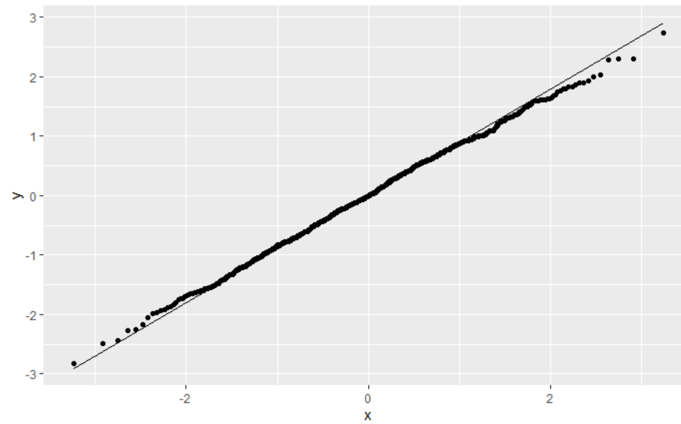


FIGURE 12: Model Assumption plots on interactive model after removal of 3p/n outliers

When running the final interactive model on the revised dataset with the outlier removed, all variables from before were still significant, or included in the model as they were part of a significant interaction term, so we did not need to adjust the model (the only term that was of note was [DraftCategory * YIM], with a P-value that changed from 0.0370 to 0.05268; we decided the interaction still was significant enough to retain, and it also allowed for model consistency). The model output is shown in Figure 13.

The model yielded an Adjusted R-squared value of 0.3918 and a RMSE of 0.874%. The dataset used in this interactive model was then used for the Box-Cox transformation.

<u>Summary of Model Assumption Tests</u>

1. **Linearity:** Yes
2. **Independence:** Not time series data, N/A
3. **Equal Variance Assumption:** Heteroscedasticity evident. Breusch-Pagan test, P-value = 0.00065 (unchanged, requires transformation)
4. **Normality of Error Terms:** Yes, Shapiro-Wilk test, P-value = 0.8059 (previously was 0.8012)
5. **Multicollinearity:** None

```
Call:
lm(formula = SavePct ~ Shutouts_y + factor(Era) + factor(Country) + 
    factor(DraftCatagory) + YIM + WeightPounds + TOI + factor(DraftCatagory):factor(Era) + 
    factor(DraftCatagory):YIM + factor(Country):factor(Era) + 
    factor(Country):WeightPounds + TOI:Shutouts_y, data = goalies_no_outliers)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.83297 -0.60600 -0.00879  0.60377  2.73215 

Coefficients: (2 not defined because of singularities)
                                          Estimate Std. Error t value Pr(>|t|)    
(Intercept)                              8.715e+01  6.419e-01 135.781  < 2e-16 ***
Shutouts_y                               5.305e-01  5.460e-02   9.716  < 2e-16 ***
factor(Era)2                             2.683e-01  1.229e-01   2.182  0.02937 *  
factor(Era)3                             7.859e-01  1.368e-01   5.746 1.30e-08 ***
factor(Country)FIN                       2.374e+00  1.779e+00   1.334  0.18247    
factor(Country)OTHER                     1.863e+00  1.163e+00   1.602  0.10957    
factor(Country)SWE                       5.323e+00  2.296e+00   2.318  0.02070 *  
factor(Country)USA                       2.417e+00  1.130e+00   2.140  0.03269 *  
factor(DraftCatagory)late               -7.091e-02  1.752e-01  -0.405  0.68581    
factor(DraftCatagory)middle              2.847e-02  1.726e-01   0.165  0.86900    
YIM                                     -3.334e-02  3.702e-02  -0.901  0.36806    
WeightPounds                             1.022e-02  3.197e-03   3.198  0.00144 ** 
TOI                                      2.583e-04  6.290e-05   4.106 4.43e-05 ***
factor(Era)2:factor(DraftCatagory)late  -4.968e-01  1.794e-01  -2.769  0.00576 ** 
factor(Era)3:factor(DraftCatagory)late  -3.440e-01  1.922e-01  -1.790  0.07379 .  
factor(Era)2:factor(DraftCatagory)middle -7.101e-01  2.226e-01  -3.189  0.00148 ** 
factor(Era)3:factor(DraftCatagory)middle -1.812e-01  2.036e-01  -0.890  0.37392    
factor(DraftCatagory)late:YIM            8.883e-02  4.578e-02   1.940  0.05268 .  
factor(DraftCatagory)middle:YIM          5.538e-02  5.693e-02   0.973  0.33089    
factor(Era)2:factor(Country)FIN          1.437e-01  2.627e-01   0.547  0.58441    
factor(Era)3:factor(Country)FIN                 NA         NA      NA       NA    
factor(Era)2:factor(Country)OTHER        1.735e-01  2.106e-01   0.824  0.41039    
factor(Era)3:factor(Country)OTHER        2.016e-02  2.207e-01   0.091  0.92724    
factor(Era)2:factor(Country)SWE         -4.062e-01  3.732e-01  -1.088  0.27679    
factor(Era)3:factor(Country)SWE                 NA         NA      NA       NA    
factor(Era)2:factor(Country)USA          8.175e-01  2.146e-01   3.810  0.00015 ***
factor(Era)3:factor(Country)USA          4.708e-01  2.170e-01   2.170  0.03033 *  
factor(Country)FIN:WeightPounds         -1.173e-02  8.909e-03  -1.317  0.18818    
factor(Country)OTHER:WeightPounds       -9.104e-03  5.797e-03  -1.571  0.11669    
factor(Country)SWE:WeightPounds         -2.486e-02  1.121e-02  -2.217  0.02693 *  
factor(Country)USA:WeightPounds         -1.343e-02  5.781e-03  -2.323  0.02044 *  
Shutouts_y:TOI                          -8.854e-05  1.628e-05  -5.438 7.15e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.874 on 805 degrees of freedom
  (123 observations deleted due to missingness)
Multiple R-squared:  0.4129,	Adjusted R-squared:  0.3918 
F-statistic: 19.53 on 29 and 805 DF,  p-value: < 2.2e-16
```

FIGURE 13: Model output and coefficients for best interactive model with outliers removed

MODEL TRANSFORMATION AND RESULTS

BOX-COX TRANSFORMATION

To attempt to remove the heteroscedasticity from our final interactive model, a Box-Cox transformation was evaluated. A “best lambda” value of 10.70707 (Figure 14) was found and used to transform the response variable.

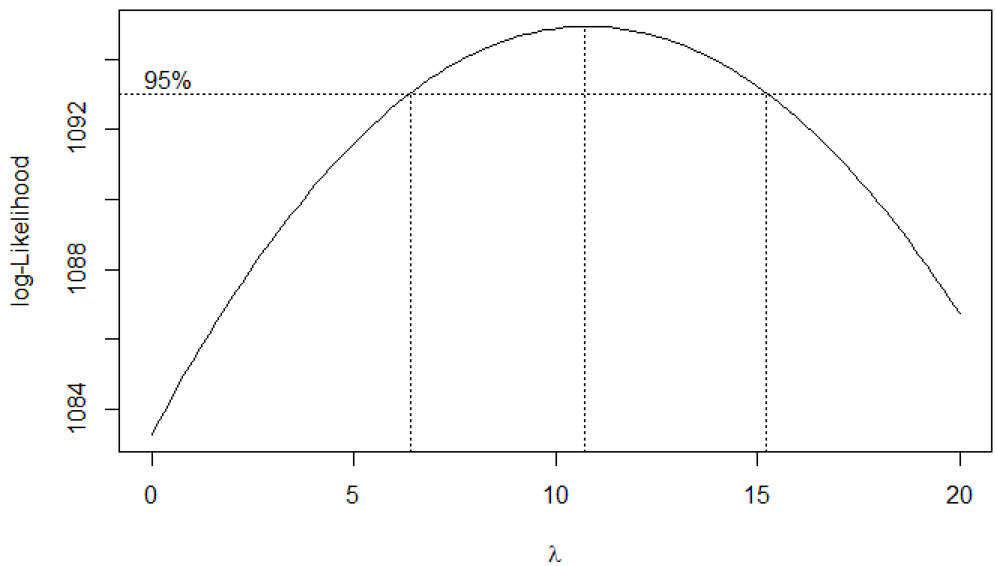


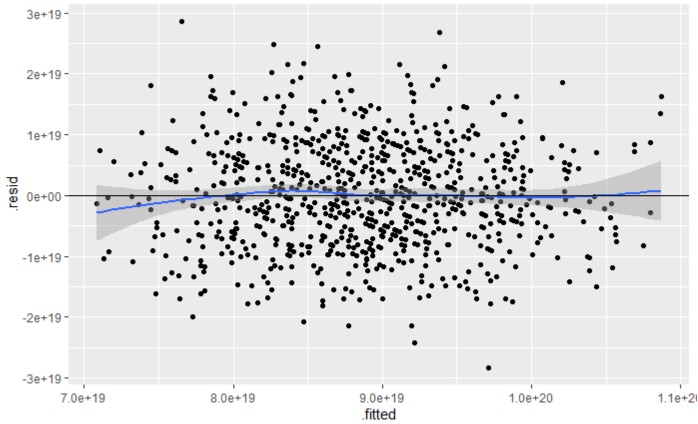
FIGURE 14: Plot of Lambda for Box-Cox transformation

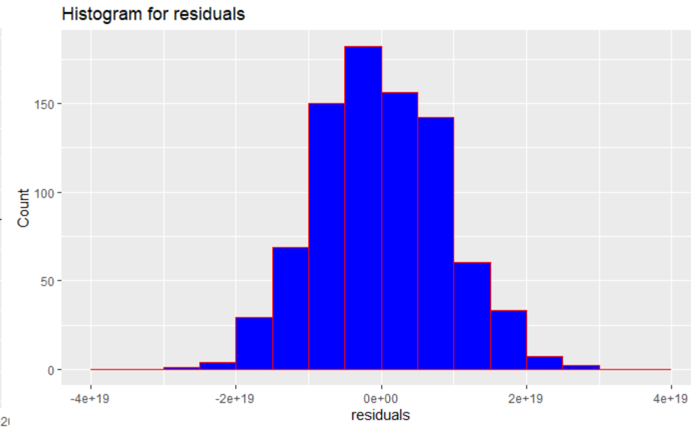
The box-cox transformed model yielded an **improved** Adjusted R-squared value of 0.3921 (see Figure 11). This model was using for checking higher order terms and to run full diagnostics. The Box-Cox transformation did improve the BP value (0.00065 to 0.01857), but not enough to correct for all the Heteroscedasticity. The SP test value was 0.1924, which was reduced (0.8059 to 0.1924), but still > 0.05, and thus meeting our model assumption for normality.

All previous-discussed diagnostic tests were again performed on the final model from this modified dataset, with the following results plotted in Figure 15.

1. **Linearity:** Yes
2. **Independence:** Not time series data, N/A
3. **Equal Variance Assumption:** Heteroscedasticity evident. Breusch-Pagan test, P-value = 0.0164 (improved, still not significant to not reject the null hypothesis)
4. **Normality of Error Terms:** Yes, Shapiro-Wilk test, P-value = 0.2879 (reduced, but still significant to not reject the null hypothesis)
5. **Multicollinearity:** None

The Box-Cox transformed model output and coefficients are shown in Figure 16.






```
        studentized Breusch-Pagan test

data:  bcmodel_goalies
BP = 47.011, df = 29, p-value = 0.01857

        Shapiro-Wilk normality test

data:  residuals(bcmodel_goalies)
W = 0.99734, p-value = 0.1924
```

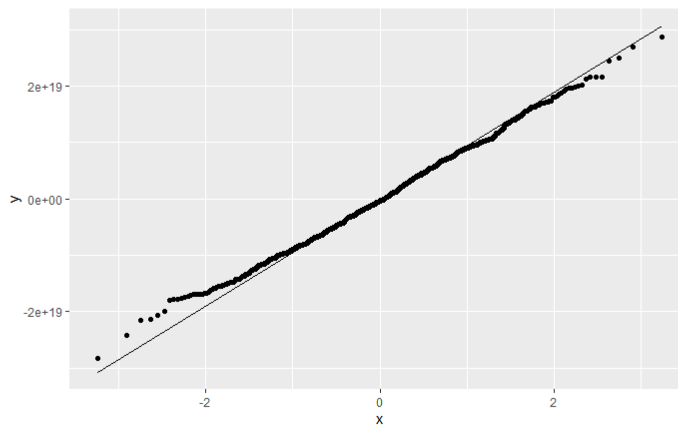


FIGURE 15: Model Assumption plots on interactive model after Box-Cox transformation

```
Call:
lm(formula = (((SavePct^bestlambda_goalie) - 1)/bestlambda_goalie) ~
    Shutouts_y + factor(Era) + factor(Country) + factor(DraftCatagory) +
        YIM + WeightPounds + TOI + factor(DraftCatagory):factor(Era) +
        factor(DraftCatagory):YIM + factor(Country):factor(Era) +
        factor(Country):WeightPounds + TOI:Shutouts_y, data = goalies_no_outliers)

Residuals:
       Min         1Q     Median         3Q        Max
-2.832e+19 -6.433e+18 -3.560e+17  6.340e+18  2.864e+19

Coefficients: (2 not defined because of singularities)
                                           Estimate Std. Error t value Pr(>|t|)
(Intercept)                               5.135e+19  6.575e+18   7.810 1.78e-14 ***
Shutouts_y                                5.321e+18  5.593e+17   9.513  < 2e-16 ***
factor(Era)2                              2.772e+18  1.259e+18   2.201 0.028013 *
factor(Era)3                              8.123e+18  1.401e+18   5.798 9.66e-09 ***
factor(Country)FIN                        2.528e+19  1.822e+19   1.388 0.165640
factor(Country)OTHER                      1.929e+19  1.191e+19   1.619 0.105813
factor(Country)SWE                        5.557e+19  2.352e+19   2.362 0.018396 *
factor(Country)USA                        2.221e+19  1.157e+19   1.920 0.055265 .
factor(DraftCatagory)late                -5.305e+17  1.795e+18  -0.296 0.767620
factor(DraftCatagory)middle               4.820e+17  1.768e+18   0.273 0.785159
YIM                                      -2.780e+17  3.792e+17  -0.733 0.463658
WeightPounds                              9.912e+16  3.275e+16   3.027 0.002547 **
TOI                                       2.261e+15  6.443e+14   3.509 0.000475 ***
factor(Era)2:factor(DraftCatagory)late   -4.850e+18  1.838e+18  -2.639 0.008484 **
factor(Era)3:factor(DraftCatagory)late   -3.647e+18  1.968e+18  -1.853 0.064296 .
factor(Era)2:factor(DraftCatagory)middle -7.173e+18  2.281e+18  -3.145 0.001721 **
factor(Era)3:factor(DraftCatagory)middle -1.784e+18  2.086e+18  -0.855 0.392688
factor(DraftCatagory)late:YIM             8.237e+17  4.689e+17   1.757 0.079361 .
factor(DraftCatagory)middle:YIM           4.686e+17  5.831e+17   0.804 0.421818
factor(Era)2:factor(Country)FIN           1.626e+18  2.691e+18   0.604 0.545757
factor(Era)3:factor(Country)FIN                  NA         NA      NA       NA
factor(Era)2:factor(Country)OTHER         1.618e+18  2.157e+18   0.750 0.453435
factor(Era)3:factor(Country)OTHER         2.001e+17  2.260e+18   0.089 0.929463
factor(Era)2:factor(Country)SWE          -3.915e+18  3.823e+18  -1.024 0.306125
factor(Era)3:factor(Country)SWE                  NA         NA      NA       NA
factor(Era)2:factor(Country)USA           8.565e+18  2.198e+18   3.897 0.000105 ***
factor(Era)3:factor(Country)USA           4.768e+18  2.223e+18   2.145 0.032240 *
factor(Country)FIN:WeightPounds          -1.255e+17  9.125e+16  -1.375 0.169386
factor(Country)OTHER:WeightPounds        -9.423e+16  5.938e+16  -1.587 0.112891
factor(Country)SWE:WeightPounds          -2.609e+17  1.149e+17  -2.271 0.023384 *
factor(Country)USA:WeightPounds          -1.242e+17  5.921e+16  -2.098 0.036243 *
Shutouts_y:TOI                           -8.490e+14  1.668e+14  -5.091 4.44e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.952e+18 on 805 degrees of freedom
  (123 observations deleted due to missingness)
Multiple R-squared:  0.4131,	Adjusted R-squared:  0.3919
F-statistic: 19.54 on 29 and 805 DF,  p-value: < 2.2e-16
```

FIGURE 16: Box-Cox model output and coefficients

PREDICTION

Using our Box-Cox transformed interactive model, we tested how the model would perform using an input dataset. In practice, this would be what a theoretical NHL team executive might undertake, if they wanted to evaluate an existing NHL goaltender, for which there was relevant data that conformed to the input data used to make our final interactive model. As our predictive model incorporates an Era factor variable (our representation of potential variance in different NHL time periods), the input data would need to confirm to one of the existing "Eras" within our source data (as described in the Data section). In practice, an executive might undertake this analysis to determine the fair market value of a potential goaltender candidate to be offered a free-agent contract, or for determination of relative value in trade negotiations with another NHL team.

In addition, we wanted to test if the Box-Cox transformation of the model resulted in actual improved accuracy in the model, since it did not fully remove the concern about our Heteroscedasticity assumption (but it did improve it). An important note to make when using the model to determine a goaltenders save percentage is to only predictors which are bound to the data used to create the model. Therefore, the bounds of predictors as limited to the following ranges:

- Shutouts = 0 to 13 [integer]
- YIM = 0 to 8 [integer]
- Weight = 165 to 238 [integer]
- TOI = 1012 to 4696 [integer]
- Era = "1", "2", "3" [factor]
- Country = "CAN", "USA", "FIN", "SWE", "OTHER" [factor]
- Draft Category = "late", "early", "middle", " NA " [factor]

A random player from the dataset was selected to test the model's predictive performance is shown below.

Input variables from randomly selected player/row:

- Shutouts = 7
- YIM = 0
- Weight = 210
- TOI = 3542
- Factor(Era) = 1
- Factor(Country) = USA
- Factor(Draft Category) = early

$$\widehat{SVPCT}_i = 89.5670 + 0.5305Shutouts - 0.0333YIM - 0.0032Weight + 0.00026TOI - 0.00009Shutouts * TOI$$

Model Output (Box-Cox transformed final interactive model):

- Fit = 91.3694
- Lwr = 89.5725
- Upr = 92.9234

Model Output (Final interactive model):

- Fit = 91.3315
- Lwr = 89.5785
- Upr = 93.0845

Actual player values:

- Tom Barasso, 1997-1998 season, Pittsburgh
- Actual season SV% = 92.2

In the predict test, the Box-Cox transformed interactive model predicted a SavePct of 91.4% (range 89.6% – 92.9%), while the final interactive model predicted 91.3% (range 89.6% – 93.1%), demonstrating that both models output a statistically accurate result. However, the Box-Cox transformed interactive model performed closer to the actual value and had a tighter confidence interval range.

As the response variable of our model is a proportion, to better understand its output we feed every row of data from the parent data set into a predict function. This was to ensure that with its own data, the model was bound to the percent range of 0-100%. Below is a comparison of the values predicted by the model, and the actual values for Save Percentage in our data set, shown in Figures 17 and 18.

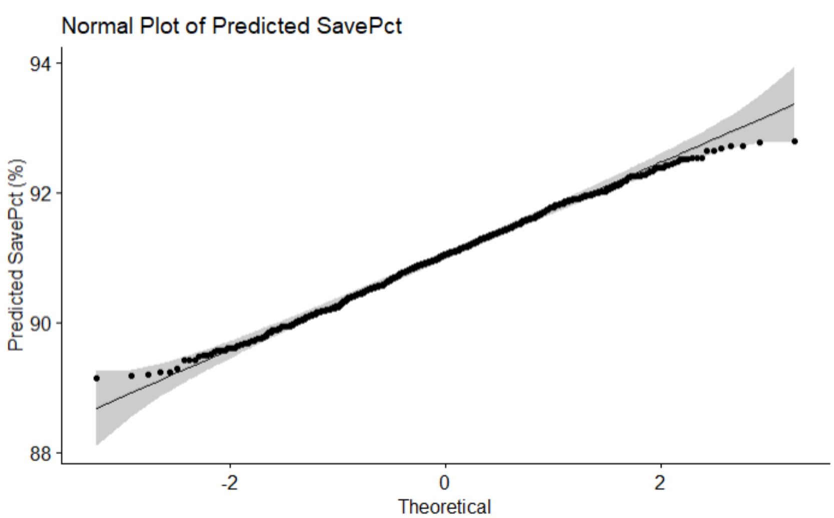


FIGURE 17: Distribution of Predicted SavePct values to the theoretical from the Box-Cox transformed model

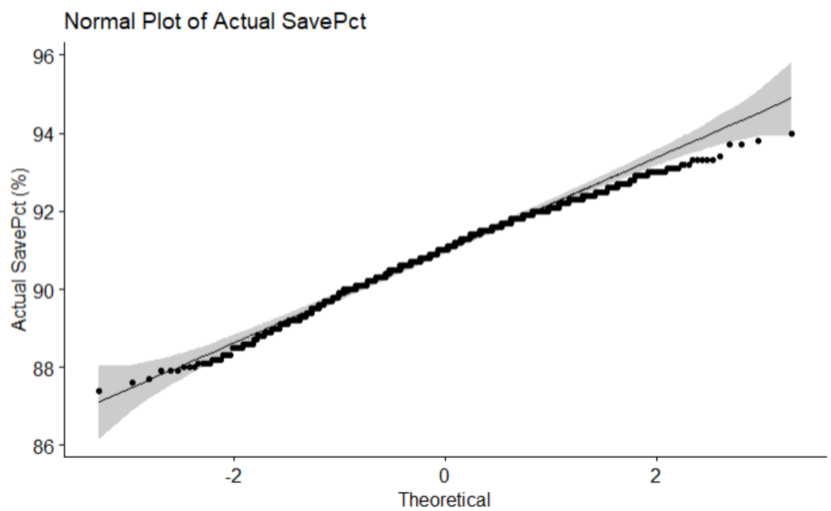


FIGURE 18: Distribution of Actual SavePct values to the theoretical from the Box-Cox transformed model

Referring to differences between Figure 16 and 17, the distribution of both sets is quite similar, especially around the middle, but also the patterns at either tail. This observation is confirmed with the favstats() outputs in Figures 19 and 20, where we see a lower standard deviation in the predicted values. The obvious concern of fitting a model to a proportional response variable is that it could model SavePct values beyond the practical ratio limits of 0 to 1. We have not observed this in our model predictions. This is not to say that our model will never produce values that are outside of the theoretical limits of the response variable, or that the accuracy and precision of our model is not potentially compromised by attempting to fit a proportion. However, the decision to proceed with our model was made with the caveat that it could potentially be improved with the introduction of more advanced modeling tools.

min <dbl>	Q1 <dbl>	median <dbl>	Q3 <dbl>	max <dbl>	mean <dbl>	sd <dbl>	n <int>
89.1506	90.53795	91.05323	91.51176	92.78435	91.02485	0.70686	862

FIGURE 19: Output of favstats from Predicted SavePct Values

min <dbl>	Q1 <dbl>	median <dbl>	Q3 <dbl>	max <dbl>	mean <dbl>	sd <dbl>	n <int>
87.4	90.2	91	91.8	94	90.96985	1.13065	985

FIGURE 20: Output of favstats from Actual SavePct Values

CONCLUSION

Our final interactive model included the following significant predictors, in addition to several interactions between the below-described predictors.

- Shutouts
- YIM (Years in Minors)
- Weight
- TOI (Time on Ice)
- Factor(Era) – 3 levels
- Factor(Country) – 5 levels
- Factor(DraftCategory) – 3 levels
- Shutouts * TOI (interaction)
- Era * DraftCategory (interaction)
- DraftCategory * YIM (interaction)
- Era * County (interaction)
- Country * Weight (interaction)

As described above, we were unable to remove the heteroscedasticity from our final model, but we were able to increase the Breusch-Pagan test P-value considerably by removing influential outliers and running a Box-Cox transformation. Therefore, we can conclude that we have a minor amount of heteroskedasticity in our data. With this, comes the risk of incorrect parameter values and artificially small P-values. The model did pass all other assumption requirements (Normality of residuals, Linearity, Independence, and multicollinearity)

When testing our, we inserted predictor values from our data set, and compared the fitted values to the actual save percentages for those goaltenders. We found that our model performed well, all test cases we examined produced a confidence interval that contained the actual save percentage value. As well, the distribution of both sets of values are similar.

In the most basic sense, if a team were able to more reliably predict which goaltenders might perform better (higher save percentage), they could statistically improve the odds of achieving a higher win percentage in each season. Our overall objective of modeling and predicting goaltender effectiveness was accomplished, as we produced an interactive model with an adjusted R-squared value of ~40%. Our model only uses data related to the goaltenders themselves, so could theoretically be used to predict a given goaltender's expected performance on any NHL team, irrespective of the team that they were a member of. An example of how this model could be used in a real-world context would be in selecting a goaltender from free agency where the goaltender candidates have at least one previous season of data to feed into the model, provided that season is consistent with the current input dataset used to build this model.

DISCUSSION

A few opportunities for future work were identified. Since the SavePct response variable is a proportion, more advanced statistical methods could be more effective for modeling the data, to avoid the prediction of impossible save percentages (negative values or values over 100%). We could also examine some of the other potential response variables (shutouts, games won, etc.) but because of the fixed number of games in an NHL season, most reasonable response variables are constrained and proportional in nature (i.e. games won translates directly to win percentage). Another future work item could be layering in additional data for NHL teams and non-goaltender players, which could provide greater predictive power. On the same note, controlling for the effects of a team's win percentage and defenceman skill on goaltender's save percentage could make the model more useful for predicting goaltender success in isolation.

The predictive capability of the current model is principally constrained by the 'Era' categorical variable that was introduced into the model to remove the time-effect, and thus represents different time periods in the NHL that share various traits. Currently, any input into the model for prediction requires an 'Era' of 1, 2, or 3, which is defined by our dataset (and our factor segmentation of that variable) as a given subset of seasons. Work effort in the future to improve the model could treat the season/time variable in different way than our 'Era' approach, to allow for prediction before or after our current dataset. A different approach to the season/time variable would still have to account for the fundamental changes in how the game has evolved through the years, in ways that influence the relative save percentage of a goaltender, specifically when comparing different years to each other. This could involve the inclusion of different variables about team or the season, or even incorporate non-linear relationship between modeled variables. If this was successful, the improved model might then be able to have predictive utility for current and future seasons. It also may have predictive utility with respect to junior (draft) prospect, or international goaltenders who lack data specific to the NHL, but are engaged competitive hockey seasons that generate goaltender and performance based data of an equivalent nature.

REFERENCES

Goldman, S., & O'Connor, C. (2022, June 22). *How do NHL analytics measure player and team value? Explaining the key advanced stats*. https://theathletic.com/3377555/2022/06/22/nhl-analytics-player-team-value/

Gross, N., (2013, June 27). *When Is the Best Time to Pick a Goaltender in the NHL Draft?* https://bleacherreport.com/articles/1685899-when-is-the-best-time-to-pick-a-goaltender-in-the-nhl-draft

NHL.com, (n.d.). *Statistics*. https://www.nhl.com/stats/

NHL.com, (2022, June 21). *NHL Vezina Trophy Winners*. https://www.nhl.com/news/nhl-vezina-trophy-winners-complete-list/c-287773436

Randjelovic, D., (2020, June 3). *11 Most Profitable Sports Leagues – Their Value Will Surprise You*. https://apsportseditors.org/others/most-profitable-sports-leagues/

Statistica, (2022, October). *Countries by number of registered ice hockey players in 2021/22*. https://www.statista.com/statistics/282349/number-of-registered-ice-hockey-by-country/

Woodley, K., (2014, December 18). *Unmasked: Analytics provide new evaluation tools.* https://www.nhl.com/news/unmasked-analytics-provide-new-evaluation-tools/c-744483

Zirhen, J., (2021, April 8). *Adjusting How We Evaluate and Analyze Goaltenders.* https://thehockeywriters.com/evaluating-analyzing-goaltenders-changes/