ACE 261 Research Paper

Lisa Gunderson

May 8th, 2019

Zach Karr

# The Effect of Age on an NHL Goalie's Save Percentage

**Introduction-**

When exploring which research topic I should pursue, I knew I wanted to do something NHL related. I am a huge hockey fan. As a result, I have the benefit of already having a deep understanding of the topic, while also choosing a question which would keep me entertained throughout the project. I ended up choosing a question based on goalie save percentage, as it fit the project guidelines perfectly and is one of the most important stats in any hockey league. In the NHL, a goalie's save percentage is one of the most widely used numbers in measuring the effectiveness of a goalie's play. This is measured by taking the amount of shots a goalie has saved and dividing that number by the total amount of shots faced, in order to get a percentage representing how many shots that goalie saves on average. For example, if a NHL goalie were to face 24 shots in a game, and he saves 22 of them, his save percentage would be 22/24= **.**917 for the game. Furthermore, you can use a goalie prospect's sv% numbers to help evaluate whether he is ready to come in and replace a struggling or veteran goaltender.

There are many different elements that can affect a goalies save percentage. Elements such as how good a team's defense is in front of him, to his current mental

state. However, in this study I want to examine the effects a goalie's age has on their save percentage in particular. As a result, my research question is as follows- *Is there a significant relationship between a goalie's age and his performance (as measured by sv%)?* My prediction is that there will be significant correlation between the two variables. In addition, I would predict that a goalie's peak performance would be from 24-33 years on average, with a steady decrease in save percentage into their mid-late thirties.

**Steps 1 and 2-**

I will using an existing data set collected from Hockey Reference, (www.hockey-reference.com), in order to evaluate my research question. My population of interest will be starting goalies from the 2000-2001 NHL regular season to the latest 2018-2019 season. I have to cut off any data before that, as the introduction of new equipment and goaltending strategies such as "the butterfly technique", has drastically increased the performance of goaltenders in recent years. I also omitted any players who played a sample size of under 20 games, as these players are not starters and the results would create unrealistic outliers when examining the average goalies performance in the long run.

This test will use two variables:

-Age: independent variable

-Sv%: dependent variable

I will measure the two by comparing the two corresponding numbers and running them through a simple linear regression test. This will be an observational study, and my

sample size is a total of 1,406 different goaltender performances since the 2000-2001 season. I transfered this data from hockey reference into a microsoft excel sheet, and used a wide range of analysis tools within the program to calculate and study my data.

**Step 3/Regression Analysis-**

I estimate the following linear regression model:

**Sv% = $\alpha$ + $B_1$age**

Figure 1 is a table that shows the descriptive statistics for both of the variables I used in my regression analysis. According to my findings, the mean age for a goaltender in my population pool is 28.53, and the median is 28 years of age. The mean sv% was .9067, with the median save percentage being .909. Lastly, the standard deviation for age was found to be 4.4178 and the standard deviation for sv% was found to be .0169. These findings fit my predictions as I estimated the prime years for a goalie to generally be between 24-33 years of age, and both the median and mean values for age were right in the middle of that predicted range. In addition, the mean goalie sv% of .9067 also makes sense to me. Typically, a save percentage below .9000 is seen as being poor, and a sv% above .9200 is seen as being dominant, (⅘ of the last 5 vezina trophy winners for best goaltender had a sv% above .930). So getting an average between the two ends is what I would have expected. 65% of the data is within 4.417 standard deviations of the mean.

Figure 2 displays the data in the form of a histogram. This histogram is in the shape of a bell shaped curve and therefore follows a normal distribution. As a result, the distribution satisfies the assumption of normality of errors.

Figure 3 displays the data in the form of a scatter diagram. The diagram shows a negative linear relationship between the variables. We can also tell that the relationship is significant, due to a calculated p-value of .0031, which is well below .05. Another way I know that this is significant at the 5% level would be that the value 0 is not included in the 95% confidence interval, which ranges from -.000502 to -.0001019. This is what I expected to see from my experience watching hockey my entire life. Age should have an impact on an athlete's performance, and as a player gets older, they usually start to decline by their mid thirties. The significant, negative linear relationship displayed in the scatter diagram displays this assumption. The calculated slope suggests that for every single increase in age, save percentage decreases by .0003. The sum of squares shows the total variation shown by the model, which would be a total of .0023. The sum of squares for the residual was .3997, which is the amount of variation not shown by the model. The age residual plot is fairly homoscedastic which implies there are constant variances between. My R^2 value was calculated as .0062. This value is very far away from 1, meaning the goodness-of-fit for my data is very low.

**Hypothesis Test-**

In order to test for significance of the coefficient on my independent variable, I first had to calculate my t-Value. As shown in figure 5, I calculated my t-value to be -2.9608. Using a .05 confidence level, I used a t-Chart to figure out the significance. The t-chart showed that for my t-value to be significant at a .05 confidence level, the absolute of that t-value would have to be greater than 1.96:

t-value: 2.9608 > 1.96, showing great significance in the coefficient of my independent variable.

**Conclusion:**

This study examined and answered the proposed question- *Is there a significant relationship between a goalie's age and his performance (as measured by sv%)?* The results of the study were largely in line with what I had originally predicted. Age was found to have a significant impact on a goalies performance, and save% did seem to decrease more drastically into a player's 30's on average. If I were to go back and do this study again there really is not anything I would have changed. I am satisfied with my data set and findings. Perhaps I would challenge myself by running a multi-regression to see how sv% effects win% also. Overall, this was a very interesting research assignment that let me get a deeper understanding of a sport and activity I very much enjoy both playing and watching.

Figure 1:

| | Age | SV% |
|---|---|---|
| Mean | 28.5309609 | 0.90671388 |
| Standard Err | 0.1178606 | 0.0004516 |
| Median | 28 | 0.909 |
| Mode | 27 | 0.908 |
| Standard Dev | 4.41780772 | 0.01692734 |
| Sample Vari | 19.5170251 | 0.00028653 |
| Kurtosis | -0.2603498 | 3.62216795 |
| Skewness | 0.38142076 | -1.1827501 |
| Range | 25 | 0.151 |
| Minimum | 18 | 0.809 |
| Maximum | 43 | 0.96 |
| Sum | 40086 | 1273.933 |
| Count | 1405 | 1405 |

Figure 2:

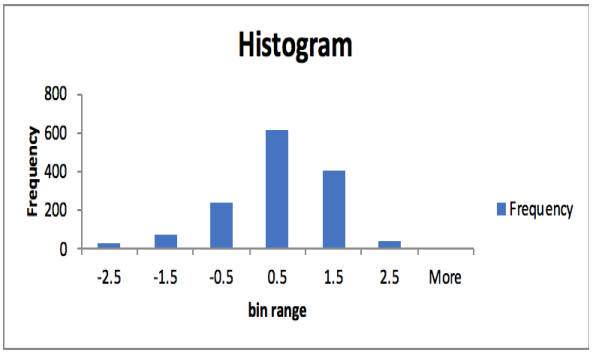


Figure 3:

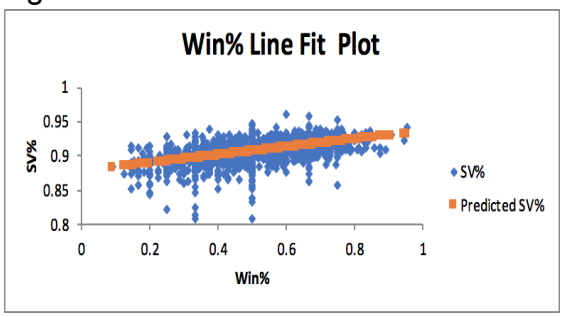


Figure 4:

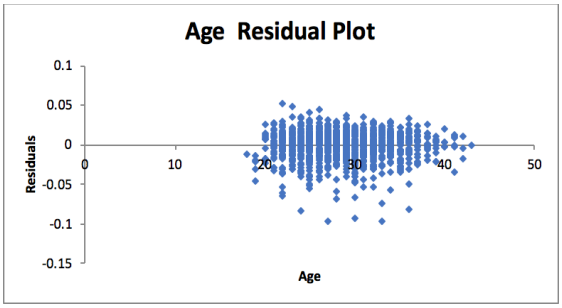


Figure 5:

| | Coefficients | tandard Erro | t Stat | P-value | Lower 95% | Upper 95% | Lower 95.0% | Upper 95.0% |
|---|---|---|---|---|---|---|---|---|
| Intercept | 0.91532833 | 0.00294414 | 310.898531 | 0 | 0.90955295 | 0.92110372 | 0.90955295 | 0.92110372 |
| Age | -0.0003019 | 0.00010198 | -2.9608121 | 0.00311984 | -0.000502 | -0.0001019 | -0.000502 | -0.0001019 |

| ANOVA | | | | | |
|---|---|---|---|---|---|
| | df | SS | MS | F | Significance F |
| Regression | 1 | 0.00249806 | 0.00249806 | 8.76640816 | 0.00311984 |
| Residual | 1403 | 0.39979692 | 0.00028496 | | |
| Total | 1404 | 0.40229498 | | | |

| Regression Statistics | |
|---|---|
| Multiple R | 0.07880058 |
| R Square | 0.00620953 |
| Adjusted R S | 0.0055012 |
| Standard Err | 0.01688072 |
| Observations | 1405 |

# Works Cited

"2018-19 NHL Goalie Statistics." *Hockey*,

www.hockey-reference.com/leagues/NHL_2019_goalies.html.