An Analysis of "Weak Goals" as an Additional Tool for Evaluating Ice Hockey Goalies

Ryan Savitz, Neumann University

Helen Cooney, Neumann University

Mya McQueen, Neumann University

Abstract

This paper introduces a statistic that may be of potential value in evaluating the value of ice hockey goalies to their teams. This new statistic is the number of weak goals allowed per game. The newly formalized definition of a weak goal was statistically validated to be a good one. The value added of this new statistic, however, was only found to be marginal.

Introduction

The purpose of this paper is to attempt to derive a new a statistic that will help to assess the value of NHL goaltenders to their teams. The art of sports analytics has been omnipresent in baseball in recent years, but has been slower to make its way to the sport of ice hockey. Toward that end, we attempt to add a level of data driven sophistication to the analysis of ice hockey goalies. The statistic that is being analyzed looks at NHL goalies and tests if their team is more likely to lose a game if they let in a "weak" goal. What constitutes a weak goal will be explained and defied by an expert who has played goalie for over 15 years and will also incorporate the opinions of others deemed experts. This paper begins by discussing the history of sabermetrics, and then moves on to our application of sabermetrics in ice hockey.

Literature Review

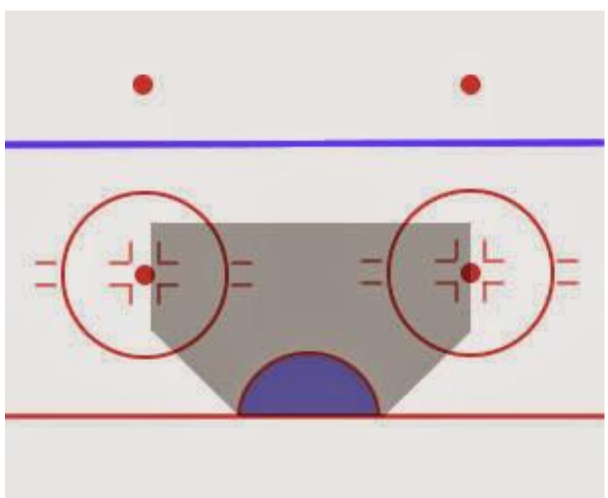
Sabermetrics is the field of study dealing with the statistical analysis of baseball data (Costa, Huber, and Saccoma, pg.1). It is defined by Bill James as, "the search for objective knowledge about baseball." Sabermetrics started to gain popularity in baseball during the 1980s (Society for American Baseball, A guide to Sabermetric research). It uses the numbers to try to place a value on the baseball players based on their performance. Some statistics that are used in baseball are earned run average, slugging percentage, and on base percentage. Each one of these statistics helps to measure the value of the player from which they are calculated.

The current statistics that are used to judge a goaltender currently are, GAA (goals against average), SV% (save percentage), wins, shutouts, QS (quality start), and RBS (really bad start). (NHL.com). Each of these is calculated using extant formulas to analyze data collected during games. Each of the following statistics is defined using the definition given on the NHL official site. GAA is the number of goals let in divided by the number of games played, using minutes played (60 minutes per game), or one whole game, as 1. SV% is the number of saves a goalie makes divided by the number of shots faced. Wins is the number of times a goalie wins a game. Note that in order to earn a win, it means that the player was the goalie in the net when the game winning goal was scored on the opposing goalie. A shutout is when a goalie plays the entire game and does not have any goals against them. Another goalie can not come in and play for a game to be considered a shutout, but if the goalie leaves the net during a delayed penalty and then returns once the is whistle blown, their time off the ice is taken off for their minutes, but they can still record the game as a shutout please explain a little more clearly. A quality start is when a goalie has a higher save percentage than the average, which is 91.7% (Sports-reference.com) or has a save percentage of at least 88.5% and let in two or fewer goals. The final statistic currently used to evaluate goaltenders is, RBS, or really bad starts. In order for a goalie to qualify for this they must have a save percentage lower then 85%. When a goalie has a save percentage under 85%, the chances that their team wins drops to 10%. (sports-reference.com)

The current statistics that are used in hockey currently to evaluate goalies are certainly of some value, but the team that a goalie is on can easily affect a goalie's statistics (NHL.com). Additionally, the statistics can be misleading due to the number of minutes the goalie played throughout the whole season. A goalie who only plays three games could have a better save percentage and goals against then a goalie who plays in 50 games, but clearly the goalie who is playing 50 games is deemed as better, even though the statistics may lead one to think otherwise. The statistic of weak goal would be most closely related to the statistic of quality start but, unlike quality starts, it would solely depend on the goalie's play and would eliminate any confounding effects related to the quality of the team they are on. For instance, a better team may eliminate high scoring chances, which allows the goaltender to make more saves from lower risk areas causing them to have a higher save percentage and a lower goals against average. These statistics are viewed as purely goalie statistics, but they can also be influenced by the team the goalie is on.

The weak goal can be compared to the error in baseball, as it's a play that should have been made, but was not. An error in baseball is defined by MLB, (major league baseball) as, "A fielder is given an error if, in the judgment of the official scorer, he fails to convert an out on a play that an average fielder should have made. Fielders can also be given errors if they make a poor play that allows one or more runners to advance on the bases. A batter does not necessarily need to reach base for a fielder to be given an error. If he drops a foul ball that extends an at-bat, that fielder can also be assessed an error." (MLB.com) This is a very close in nature to the definition of a "weak goal" in ice hockey that we will propose. Formally, we define a weak goal as follows:

A weak goal is a shot that a goaltender should normally stop but doesn't. A weak goal is a judgement call, but it must not have been deflected on the way in, the goalie must be able to see the shot, the goalie must be set and not on the move, and any shot that comes outside of the high scoring area (diagram 1) if the first three conditions are met. There must be at least one defender back, so no breakaways, and the goalie must not have been interfered with by either someone on the other team or their own. In order for a goal to be deemed weak in the high scoring area, the shot must go through the goalie, meaning they got a piece of it but let it get through or around their equipment.



(Diagram 1) High Scoring Area

Data Analysis

To confirm the definition of a weak goal as previously stated, Cronbach's alpha was calculated. According to the Institute for digital research and education (https://idre.ucla.edu), Cronbach's alpha is a test used to measure the internal reliability of survey instruments. In this particular case, however, the test is used to test the consistency of the understanding of the definition of a weak goal across independent subjects. 10 people were used in this part of the study and they were each asked to read our definition of the weak goal and use it to evaluate 10 NHL games from the 2016-2017 season. The ten participants all have some prior knowledge about the game of hockey. We believe that this is appropriate, since, if our definition is every formally applied, it will most likely be applied by hockey referees, team statisticians, and others with a knowledge of the game.

Based on the aforementioned data collection, Cronbach's alpha was calculated to be 0.991, which is considered very strong. There is no definitive cut-off for the Cronbach's alpha that makes it significant, but most people tend to agree that above a .7 is good. (https://idre.ucla.edu/) This high of a value means that the definition of a weak goal is extremely consistent across subjects and, therefore, that one can accurately state whether a goal is weak or not. (https://idre.ucla.edu/)

Since we have now confirmed the statistical validity of our definition, we next evaluate the value of this new statistic. In order to do so, we randomly selected 74 NHL ice hockey games in the 2017-2018 regular season. Each game of the season was assigned a number and then a random number generator was used to select 74 games. The highlights of these games, including all goals scored in the game, were watched. The current statistics used for NHL goalies were recorded as was the number of weak goals allowed in each game. Once all the games were watched and all the data were collected, the data were entered into SPSS, a statistical software package.

Using SPSS, three logistic regressions were performed. In each case win/loss was the dependent variable. The 3 sets of independent variables used were: (1) just weak goals, (2) just SVP, (3) both weak goals and SVP. The first model we considered involves both the SVP and Weak variables. The logistic regression run for this model was:

- $Y = B_0 + B_1*SVP + B2*Weak$,

Where $Y = 0$ if a team loses and 1 if they win, SVP is the goalie's save percentage, Weak is the number of weak goals allowed, and B_0,B_1, and B_2 are the parameter estimates.

The Nagelkerke's R^2 for this model (Table 1) is 0.463. While this statistic is not as straight forward to interpret as the R^2 value in ordinary least squares regression (Allison 1998), it does give us some idea of the explanatory power of this model (on a scale from 0 to 1).

The Homer and Lemeshow goodness of fit statistic explains how well the model fits the data and if the model is a good fit. Based on the data in table 2 we can see its value of 0.647 is not nearly statistically significant, which means we do not reject the null hypothesis, which means that the model is a good fit.

Also in table 3, the classification model explains what our model has predicted versus what actually happened. For example, when we have .00 (lose) as both the real and predicted value (top left of table), this means that we correctly predicted 26 losses.

In Table 4, The SVP variable does appear to have a statistically significant relationship with winning. The p-value is < 0.05, so we would reject the null hypothesis of no relationship. This model also shows that, after SVP is accounted for, there does not appear to be any relationship between weak goals and winning.

Model Summary

Cox & Snell R^2	Nagelkerke R^2
0.347	0.463

Table 1

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	6.000	8	.647

Table 2

Classification Table[a]

			Predicted		
			Win		Percentage Correct
			.00	1.00	
Step 1	Win	.00	26	8	76.5
		1.00	8	29	78.4
	Overall Percentage				77.5

Table 3

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	SVP	35.711	9.272	14.833	1	.000	3228318113129173.000
	WEAK	-.094	.439	.046	1	.831	.911
	Constant	-32.528	8.540	14.507	1	.000	.000

Table 4

The next model we considered involves only the SVP variable. The logistic regression run for this model was:

$Y = B_0 + B_1X_1$,

Where Y = 0 if a team loses and 1 if they win,

X_1 is the save percentage for the goalie of interest in the game, and B_0 and B_1 are the parameter estimates.

Based on the "variables in the equation" table (Table 5), we can see that the significance of the variable SVP is 0.000. Since this is lower than the cut-off significance level of 0.05, we do have strong evidence that a statistically significant relationship exists between the SVP and a team's likelihood of winning.

Related to this, if we look at the "classification table" (Table 6), we see that this model correctly predicts 76.5 % of the losses and 78.4% of the wins. Thus, this model does quite a good job of predicting wins when they actually occurred.

The "Homer and Lemeshow" goodness of fit test above tests (Table 7) has a p-value of 0.832. This indicates that, as with the first model, this model is also a good fit to the data.

Looking at the "Model Summary" table (Table 8), it can be seen that Nagelkerke's R^2 is equal to 0.462, which is only slightly lower than the value of this statistic from the first model. This means that we do not lose much explanatory power by eliminating the weak goals variable.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	SVP	36.147	9.056	15.934	1	.000	4995962283870102.000
	Constant	-32.965	8.304	15.759	1	.000	.000

Table 5

Classification Table[a]

Observed			Predicted: Win .00	Predicted: Win 1.00	Predicted: Percentage Correct
Step 1	Win	.00	26	8	76.5
		1.00	8	29	78.4
	Overall Percentage				77.5

Table 6

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	4.266	8	.832

Table 7

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	68.116[a]	.346	.462

Table 8

The final model we will consider involves only the "Weak" variable. The logistic regression run for this model was:

- $Y = B_0 + B_1 * Weak$

Where Y = 0 if a team loses and 1 if they win,

X_1 is the number of weak goals allowed by the team of interest in their game, and B_0 and B_1 are the parameter estimates.

The Nagelkerke's R^2 (Table 9) is seen to be just 0.076. This implies that, alone, the number of weak goals does not predict much of the variability in whether or not a team wins.

The Homer and Lemeshow statistic, again, explains how well the model fits the data and if the model is a good fit. Based on Table 10, the statistic's value is 0.937, which is quite high, which means that the model including only number of weak goals is a reasonably good fit to the data, despite this model's relatively low explanatory power.

The classification table (Table 11) explains what our model has predicted versus what actually happened. From this table, we can see that this model correctly predicts 41.2% of the losses and 78.4% of the wins. Thus, this model does quite a good job of predicting wins when they actually occurred. Unfortunately, it does not do such a good job of predicting losses (many of the actual losses were predicted to have been wins). This means that this model predicts that a team will win more often than it really does.

Table 12 shows the significance of the variable in this equation (number of weak goals). Based on the "variables in the equation" table, we can see that the significance of the variable WEAK is 0.054. Since this is greater than the cut-off significance level of 0.05, we do not have strong evidence that a statistically significant relationship exists between the number of weak goals allowed and a team's likelihood of winning. That said, the significance of WEAK is just barely over 0.05, so there is some suggestion that a relationship might exist between the variables. An additional study, with a larger sample size, might be able to discern this relationship with more accuracy. That said, the exp(B) value of Weak in this table is 0.493. This means that, for each weak goal allowed, a team's likelihood of winning is just 49.3% of what it once was (Pampel 2000). In other words, a team's chance of winning a game is roughly cut in half for each weak goal allowed.

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	94.112[a]	.057	.076

Table 9

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	.006	1	.937

Table 10

Classification Table[a]

			Predicted		
			Win		Percentage Correct
	Observed		.00	1.00	
Step 1	Win	.00	14	20	41.2
		1.00	8	29	78.4
	Overall Percentage				60.6

Table 11

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	WEAK	-.708	.366	3.728	1	.054	.493
	Constant	.383	.283	1.830	1	.176	1.467

Table 12

The overarching results of these regressions was that weak goals is nearly, but not quite, statistically significant on its own (p-value from logistic regression = 0.054). while SVP is highly significant on its own (p-value from logistic regression < 0.001). When SVP is already in the logistic regression model, weak goals only adds very small bit of value, as evidenced by Nagelkerke's R^2 increasing by just 0.001 from the model with SVP only to the model with both SVP and weak goals. Weak goals also tended to overpredict wins. By this, we mean that a low number of weak goals allowed makes it seems like a win is highly likely, when, in reality, it is not that certain. This leads us to the conclusion that the number of weak goals allowed does not does not give a complete indication if their team will win or not.

Conclusions

In summary, it can be seen that this newly defined version of the "weak goal" statistic is statistically valid, based on the Cronbach's alpha value that was calculated. This means that, in the future, ice hockey fans, players, and score keepers can use this definition as a consistent, clear, and statistically valid ice hockey metric.

Unfortunately, it was shown that "weak goals" did not add much explanatory power to the currently used statistic of save percentage (when attempting to predict whether or not a team wins). On its own, however, weak goals was found to be nearly statistically significant, which implies that this may be yet another statistic that can be used to evaluate the value of ice hockey goalies.

Two exciting areas for future research came from this work. First, an additional study, with a larger sample size, might be able to discern the relationship between winning and weak goals with more accuracy. Second, the fact that the statistic of "weak goals," which was the first statistic thought of and evaluated by the authors, was nearly statistically significant, gives hope to the possibility that, with future research, an even more powerful new statistic might be derived. Some possibilities for new statistics include variants of the weak goal statistic as well as terms consisting of an interaction between currently used statistics and the weak goal statistic.

References

Allison, P.D. (1998). Multiple Regression: A Primer. Sage, Thousand Oaks, CA.

https://idre.ucla.edu/

NHL Schedule | ESPN. (n.d.). Retrieved from http://www.espn.com/nhl/schedule

NHL Stats, History, Scores, & Records | Hockey-Reference.com. (n.d.). Retrieved from https://www.hockey-reference.com/

NHL.com - Stats. (n.d.). Retrieved from http://www.nhl.com/stats/

Pample, F.C. (2000). Logistic Regression: A Primer. Sage, Thousand Oaks, CA.

What does Cronbach's alpha mean? | SPSS FAQ. (n.d.). Retrieved from https://stats.idre.ucla.edu/spss/faq/what-does-cronbachs-alpha-mean/