MODULE 7: CRITICAL THINKING

Portfolio Project - Paper

Rachel Thornes

Colorado State University Global

MIS581 Capstone - Business Intelligence and Data Analytics

Dr. Orenthio Goodwin

February 25, 2024

Abstract

This paper seeks to understand the role between goalie performance in the National Hockey League (NHL) and rest days, goalie experience and wins, and goals against and losses in the 2022-2023 regular season. At this time, the NHL has fairly recently begun investing in data analytics and tools such as “The Edge” to seek further understanding of the game, teams, and players for fans, teams, and coaches. Goalies take a lot of the blame for the performance of a team due to how much time they spend on the ice and the fact that they are the last line of defense before a goal can be scored.

Through SAS studio and using linear regressions along with data taken from the NHL website, the relationship the impact of rest days on save percentages, the relationship between goalie experience and wins, and the correlation between goals against and team losses demonstrates the importance of a goalie on a game outcome. This study found no significant relationship between rest days and save percentages but was able to identify a relationship between goalie experience and wins, as well as between goals against and team losses. This study only confirms the importance in the NHL to continue investing in data analytics of teams and exploring goalie performance.

Introduction

Mark Scheifele, current captain for the Winnipeg Jets, stated “To be honest, I don't have much time for analytics. You know, I don't pay any attention to it... I don't even understand them." (*Larkinnov, 2021*). Unfortunately, this is the sad truth of many players and teams in hockey who could be utilizing analytics to make gains. Team analytics have really only been a thing the past ten years when the Toronto Maple leafs hired someone to build an analytics team. In hockey, goalies often play the most important role. Teams often build their defense around their goalies strengths and weaknesses. Goalies take on a lot of responsibility for any goals let into their net. Additionally, goalies are often the only player who spends the entire game on the ice. Therefore, the purpose of this project is to better understand the relationship between goalie’s and their performance. Goalie performance metrics, such as save percentage, goals against average, overtime losses are influenced by various attributes including experience, rest, age, glove hands, and so much more.

Objectives

The objective of this project is to be able to better understand goalie performance in the NHL. The research will explore three areas of goaltending to better understand the objective. It will look at the relationship between rest days and save percentages, goalie experience and wins, and goals against and loses..

Overview

Several studies have demonstrated the importance of data analytics in understanding team performance. However, focusing on goalie performance is still needed. This paper seeks to understand the relationship between goalie performance and goalie attributes like height, weight,

and time in the NHL. Furthermore, it hopes to identify the effect rest days have on goalie performance in better understanding the benefits resting may have on a goalie. This research will do this by designing the questions and hypotheses, better understanding current trends in data, deciding the tools to use, looking at the ethical considerations, and then testing the hypotheses.

Questions and Hypothesis

1. Is there a relationship between rest days and save percentages?
    a. H0 states: There is no relationship between rest days and save percentages
    b. Ha: There is a relationship between rest days and save percentages
2. Is there a relationship between goalie experience and wins?
    a. H0 states: There is no relationship between the goalie's experience and goals against.
    b. Ha: There is a relationship between the goalie's experience and goals against
3. Is there a relationship between goals against and team losses?
    a. H0 states: There is no relationship between the goals against and team wins
    b. Ha: There is a relationship between the goals against and team wins

The reason the goalie metrics were picked for analysis was in relation to trends in fitness. Many athletes, workout influencers, and scientists have pointed out the importance of rest. In fact, most athletic bands have a recovery aspect relating to days off and sleep. Even the Children's Hospital of the King's Daughter states, "Rest days also prevent overtraining. Constantly working out and training without recovery days can be detrimental to an athlete's body."(Gilmartin, 2021). With rest at the forefront of athletic and fitness performance, it would be valuable to find out if rest had the same effect on goalie performance.

The reason goalie experience was picked for analysis was due to the belief that practice makes perfect. As stated in "Outliers" which studies how mastery is made, specifically in a championship youth hockey team, “the idea that excellence at performing complex tasks requires a critical minimum level of practice surfaces again and again in the studies of expertise.” (Gladwell, 2009). This supports that even excellent hockey players would need experience in the NHL to become high level performing goalies in the NHL. Using NHL statistics on goalie’s length in the NHL will allow this to be further explored.

Lastly, goalie loses and goalie goals against was compared to better understanding the relationship between a goalie and their team. Gene Ubriaco stated, “goaltending is 75 percent of the game — unless it's bad goaltending. then it's 100 percent” (Oil, 2023). This will better understand the relationship between bad goaltending, or goals against, and game outcomes.

Literature Review

Hockey has come a long way since the beginning of the NHL in 1917. As the sport has grown, so has the performance and talent of the skaters. With faster shooting pucks and high speed crashes, goalies have had to adjust to the game becoming their own unstoppable force. Currently, data analytics in hockey is only about 10 years old. Some teams have only hired analytics teams in the past 2 years as the idea of data has been embraced by all aspects of the world. Goalies play one of the most important roles on the team as the player with the most ice time, the last man standing between the puck and the net, and as someone who, if they have a terrible game, it will be felt by the whole team. There are multiple factors that go into a team’s performance and ranking in the NHL. These factors can include skater skill, defensive strategy, or goalie save percentages depending on which study you read. The goal of the research is to better understand

how goalie and team performance is currently being evaluated. This is important because it will allow better understanding for where goalie performance evaluation can still be improved on.

Team and player performance is heavily influenced by a variety of factors in the NHL. In a quantitative study examining progression of an NHL team, Crittenden (2008) found that goalie performance had a direct impact on team standings and allowed predictions for which teams would make playoffs. Using goalie score differentials, save percentage, and goals against average over the past few seasons, he was able to cluster map teams into 5 levels of possible standing. In a similar study, expected goals were measured by figuring out the probability of a type of goal to be scored. Naples, Gage, and Nussbaum (2018) found that by increasing save statistics to include a wider range of what defines a "save" created a more reliable way to read goalie performance. Additionally, they looked at scatter plots, correlation maps, probability analysis, and heat maps to better understand the relationship between goalies save percentages and team performance. They concluded that goalie save percentage had less to do with goalie performance, and more to do with team performance. In the third study, Wu, Shen, and Cui (2023) found 5 key performance indicators that could predict a game outcome. They did this using clustering methods comparing teams and game outcomes. They also found that "defensive ability, consistent possession skills and high shooting accuracy" (Wu, 2023) would lead to better overall performance in league standings. These studies all compare team performance using goals to better understand the relationship between the NHL rankings and team performance. They differ in what they identified as the key factors that affect team performance. For example, study 1 used mainly goalie statistics to provide more information on NHL standings, while study 2 found that goalie save percentages were a poor predictor of team performance and more

representative of the team's performance as a whole. In the last study, they found that NHL standings were most affected by skater skill and strategy.

Research on hockey currently looks mostly at team's standing in the NHL. Currently, this has been done using clustering, heat mapping, and scatter plots. So far, researchers have found that team performance is based on the skills of a team including agility, defensive strategy, and save percentages by goalie. Further research about goalie performance alone would allow a narrowed focus on one aspect of team performance in the NHL allowing for an improved chance at ranking higher in the NHL by improving the goalie's ability to keep pucks out of the net. Additionally, by looking at it through regression analysis, this will create a new way to view the data and possibly create a stronger model than some of the other models were able to produce. One possible weakness of this research is that there is not enough data published by the NHL to provide a full understanding. As mentioned in Study 2, the way save percentage is currently measured is not all encompassing of what goalies actually "save" as it only counts on-net shots. This means there is information that could be useful that the dataset leaves out such as shots that are just complete misses, more information on goalie's training program, information on goalie's fear factor from other players in the league, and so much more. Next steps for the NHL is to include more information on their players training programs and include qualitative data on the players to allow for better analysis of their performance.

Research Design

The tool used to perform the analysis will be SAS Studio. SAS Studio is perfect for the project as it is a "major package for powerful statistical analysis and data manipulation" (Cody, 2018) that is used by many workplaces and universities for database manipulation and analytics.

Within SAS, a technique that will be utilized is a regression analysis. A regression analysis "can be used to explore the relationships between variables and make accurate forecasts based on those relationships" (*Chatterjee, 2013*). This will allow the relationship between variables like goalie's days of rest to be compared to their save percentage of that game.

Regression analysis allows for better understanding of the relationship between the dependent and classification variables. A good fit model will be demonstrated by a low p value which represents the relationship between the variables is statistically significant. Additionally, a good fit model will have a high r-squared value. A higher value, number closer to one, represents that more strength in the linear relationship between the two variables. Next, the printed graphs will be used to better understand the model. A good fit model will be represented in the observed by predicted graph by having observations that are randomly scattered but closely following the line. The fit diagnostic graphs will also be used to demonstrate a good fit model. For example, the predicted vs residual graph will demonstrate a good fit model if the observations are randomly scattered. A quantile quantile plot will demonstrate a good fit model if the observations closely follow the line. The residual by percent plot will demonstrate a good fit if the errors closely follow the distribution line. Anything opposing a good fit model such as a high p value or low r-squared value may be representative of a bad fit model validating the null hypothesis.

The Dataset

The dataset includes 49 variables. The dataset was created using three datasets given by the NHL on goalie advanced summary, goalie bio, and goalie days of rest. Repeat variables were deleted. The dataset is arranged in alphabetical order based on the goalie's name. It includes only

goalies who played the 2023-2024 regular season and included only statistics from the 2023-2024 regular season. All variables are stored as either a character or a number.

This dataset was picked because it has the scope to answer the question, “What attributes contribute to a goaltender's performance?” Data should be judged on its “availability, usability, reliability, relevance, and presentation quality” (*Cai, 2015*). This data is easily available as it is public on the NHL website. It is updated after every NHL game with current statistics and goes back as far as 1917 although the statistics look quite different. The data is usable. It is clearly defined, checkable, and in understood formats of sports statistics. The sata is reliable. The dataset was created by the NHL meaning that the information is accurate. It includes unbiased statistics of all goalies, regardless of their play time or status, to ensure all games are represented. All goalies are tracked in the same way. The dataset is consistent and readable. After being downloaded from the website, uploaded as a csv, and then stored in Sas, the formats of all variables remain the same. The dataset is relevant. All data relates back to the goalies of the 2022-2023 NHL season. It contains a wide range of variables that allow for exploration of goalie performance such as goalie demographics, game outcomes, and goalie game play statistics. The model will answer the question with this dataset by demonstrating if there is a relationship between some of these variables and goalie performance.

Data Dictionary

Alphabetic List of Variables and Attributes				
#	Variable	Type	Len	Description

12	0 Days Rest	Num	8	Games played 0 days rest
17	0 Days Sv%	Char	5	Save percentage 0 days rest
13	1 Days Rest	Num	8	Games played 1 days rest
18	1 Days Sv%	Char	5	Save percentage 1 days rest
33	1st Season	Num	8	First season of game type
14	2 Days Rest	Num	8	Games played 2 days rest
19	2 Days Sv%	Char	5	Save percentage 3 days rest
15	3 Days Rest	Num	8	Games played 3 days rest
20	3 Days Sv%	Char	5	Save percentage 3 days rest
16	4+ Days Rest	Num	8	Games played 4+ days rest
21	4+ Days Sv%	Char	5	Save percentage 4+ days rest
24	Birth City	Char	16	Where the goalie was born
37	CG	Num	8	Complete games when started
39	CG%	Char	4	Complete game percentage

26	Ctry	Char	3	Country
23	DOB	Num	8	Date of birth.
30	Draft Yr	Char	4	Year drafted
43	GA	Num	8	Goals against
45	GAA	Num	8	Goals against average
42	GF	Num	8	Goals for
44	GFA	Num	8	Goals for average
5	GP	Num	8	Games played
6	GS	Num	8	Games started
40	GS > .900	Num	8	games started with > .900 save percentage
41	GS > .900 %	Char	4	Percentage of games started with > .900 save percentage
34	HOF	Char	1	In hall of fame
28	Ht	Num	8	Height

38	IG	Num	8	Incomplete games when started
8	L	Num	8	losses
27	Ntnlty	Char	3	Nationality
10	OT	Num	8	Overtime Losses
32	Overall	Char	3	Player draft overall pick number
1	Player	Char	19	Player name
47	ROL	Num	8	Regulation losses plus overtime losses
46	ROW	Num	8	Regulation wins plus overtime wins
31	Round	Char	2	Player draft round number
4	S/C	Char	1	Skater shoots/goalie catches
25	S/P	Char	2	Player birth stat/providence
48	SA/60	Num	8	Shots against per 60 minutes
36	SO	Num	8	Shutouts.
2	Season	Num	8	Season

11	Sv%	Char	5	Save Percentage
9	T	Char	2	Ties
49	TOI	Num	8	Time on ice
3	Team	Char	3	Teams played for
7	W	Num	8	Wins
29	Wt	Num	8	Loses

Limitations and Ethical Considerations

This research is limited by the data available. The goalies have different training programs and schedules that would provide much more information on what is affecting their performance. Additionally, there is much more data that individual teams keep private in understanding their own team performance taken by their video coaches and data analytics teams. One piece of information that would have been potentially eye opening is individual game stats per goalie as that may have provided a more in depth understanding of their performance.

Not all goalies are created equally. For example, many of the goalies on this list do not even belong to a team but were simply pulled in for a game, or even part of game, as a third or fourth string goalie. When some goalies have over 4,000 minutes spent on the ice, the goalies with only 20-50 minutes have not had the same opportunity to demonstrate their success, or lack thereof. Additionally, there is no control in shots. Goalies will play different players, different

teams, take different types of shots leading to the possibility of different statistics. Additionally, as a team sport, the goalies statistics can't be left fully accounted to the goalies success or lack of success. "Hockey is a unique sport in the sense that you need each and every guy helping each other and pulling in the same direction to be successful," said Wayne Gretzky, one of the all time greatest hockey players. A goalie who is appearing bad statistically has the opportunity to be great with a team that is built around his goalie's strengths and weaknesses. The statistics may be less a sign of the goalies performance and more a sign of the team's performance.

One thing the statistics do a good job of is measuring fairly and accurately. Being done using over 20 cameras, infrared tracking technology, and machine learning (EDGE, 2023), the statistics are all generated without the error of human eye or human bias. The edge has tracking devices in all players' jerseys and all NHL pucks to get the most updated information from every game. Additionally, it covers all games and all goalies making everything that can be tracked, tracked.

Findings

Testing Question 1

There is no relationship between rest days and goals let in. First, a summary statistic was completed to see if there was any pattern in days rest to save percentages through means. Wins were also included to see if they followed a similar trend. This can be seen in Figure 1. The mean increased slightly between 0-2 days rest but decreased for 3 days rest between rising slightly with 4+ days rest. Wins followed a very similar pattern. This represented that there could be some correlation as there was a clear increase between 0-2 days rest. However, it pointed out as unlikely due to how close all the numbers were.

Figure 1

Summary Statistics

Rest Days	N Obs	Variable	Mean	Std Dev	Minimum	Maximum	N
0	23	Sv% W	0.8656522 0.2608696	0.1961601 0.4489778	0 0	0.9640000 1.0000000	23 23
1	530	Sv% W	0.8898072 0.4660377	0.0832613 0.4993165	0.2500000 0	1.0000000 1.0000000	529 530
2	339	Sv% W	0.9002684 0.5250737	0.0741305 0.5001091	0.5000000 0	1.0000000 1.0000000	339 339
3	260	Sv% W	0.8850654 0.4038462	0.0982824 0.4916137	0 0	1.0000000 1.0000000	260 260
4	696	Sv% W	0.8985661 0.4770115	0.0702871 0.4998305	0.5000000 0	1.0000000 1.0000000	696 696

Note. Save percent means are all within 4% of each other. They do increase slightly between 0-2 days of rest before decreasing again. Wins follow a similar pattern.

A linear regression demonstrated that there is no relationship between rest days and goals let in. Save percent average was set as the dependent variable. Days rest was set as the continuous variable and listed as 0, 1, 2, 3, or 4+. In figure 2, the analysis of variance is printed. The p value is 0.097 which is high. A p value of less than 0.05 represents a good fit model. Additionally, the r-squared value is also small. The r-squared value is 0.0015 meaning that days of rest can explain about 0.15% of the variability in save percentages. In figure 3, the observed by predicted is printed. The observations are not real close to the line and are not randomly scattered representing a bad fit model. In Figure 4, the fit diagnostics are printed. In the Quantile-Quantile plot, the observations steeply curve off the line on both the left and the right side of the plot representing a bad fit. Additionally, the distribution of the residual by percent plot

shows left-skewed distribution. Therefore the negative hypothesis must be accepted and the alternative hypothesis must be rejected. There is no relationship between rest days and save percentages.

Figure 2

Analysis of Variance for Save Percentage by days of Rest

Model: MODEL1
Dependent Variable: Sv%

Number of Observations Read	1848
Number of Observations Used	1847
Number of Observations with Missing Values	1

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.01848	0.01848	2.76	0.0969
Error	1845	12.35745	0.00670		
Corrected Total	1846	12.37593			

Root MSE	0.08184	R-Square	0.0015
Dependent Mean	0.89406	Adj R-Sq	0.0010
Coeff Var	9.15377		

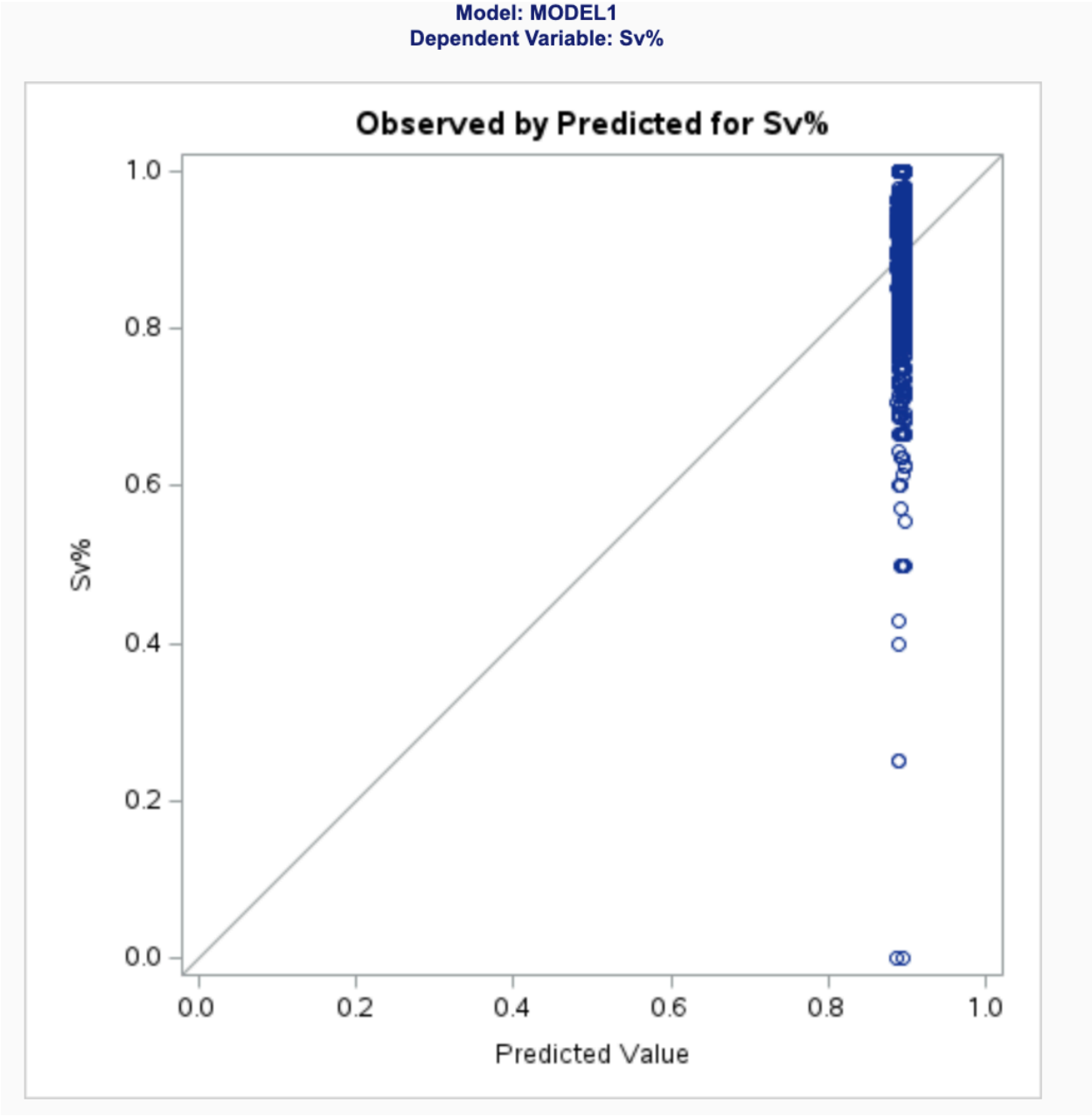
Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.88769	0.00428	207.27	<.0001
Rest Days	1	0.00247	0.00149	1.66	0.0969

Note. The p value is greater than 0.05 representing a bad fit.

Figure 3

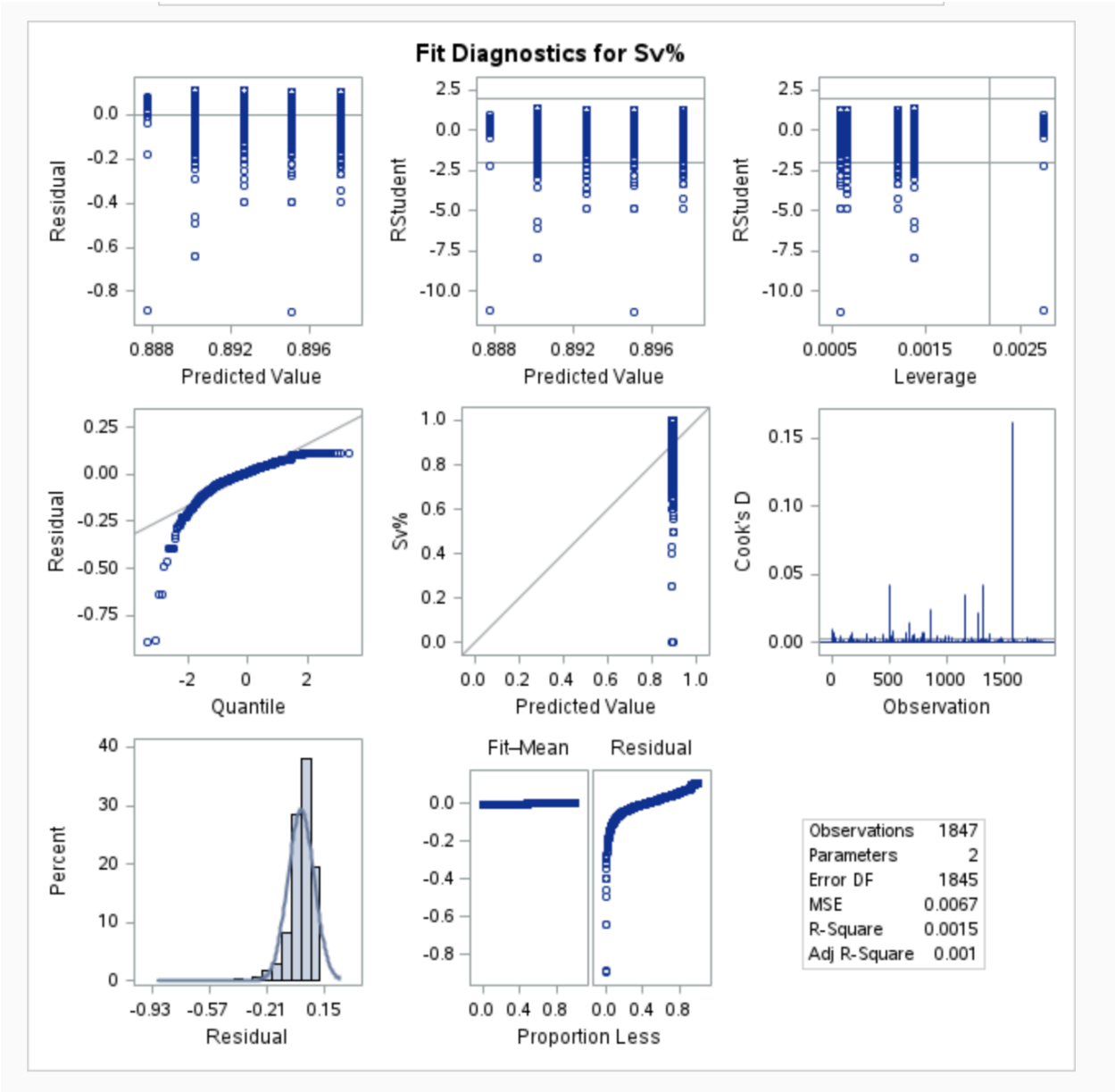
Observed by Predicted for Save Percentage



Note. The observations do not follow the line representing a bad fit.

Figure 4

Fit Diagnostics for Save Percentage



Note. The quantile quantile plot has a steep curve off to the left and right representing a bad fit.

Testing Question 2

A linear regression demonstrated that there is a relationship between goalie experience and average goals against. Goals against was set as the dependent variable. Goalies 1st season was set as the continuous variable. In figure 5, the analysis of variance is printed. The p value is 0.0006 which is low. A p value of less than 0.05 represents a good fit model. This p value represents a good fit. The r-squared value is 0.1131 meaning that the start season of the goalie can explain about 11.31% of the variability in average goals against. In figure 6, the observed by predicted is printed. The observations are relatively randomly scattered representing a good fit model but do not closely follow the line which causes some concern. In Figure 7, the fit diagnostics are printed. In the Quantile-Quantile plot, the observations closely follow the line representing a good fit. Additionally, the distribution of the residual by percent plot follows the normal distribution although slightly right skewed. The model is a good fit. Therefore the null hypothesis must be accepted and the alternative hypothesis must be rejected. The null hypothesis can be rejected and the alternative hypothesis can be accepted. There is a relationship between goalie experience and goals against.

Figure 5

Analysis of Variance for Goals Against

Model: MODEL1
Dependent Variable: GA

Number of Observations Read	100
Number of Observations Used	100

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	31440	31440	12.50	0.0006
Error	98	246515	2515.45628		
Corrected Total	99	277955			

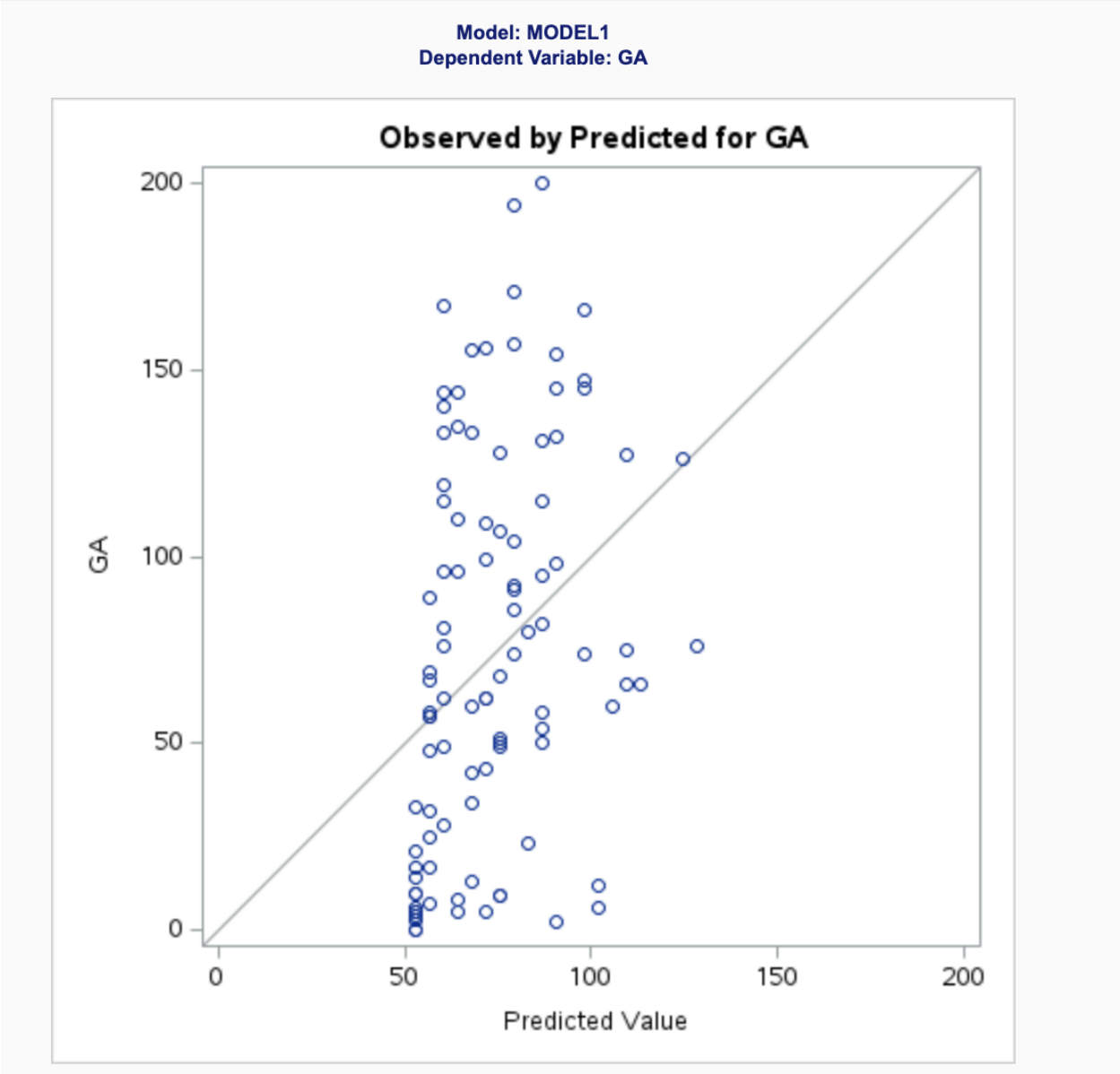
Root MSE	50.15432	R-Square	0.1131
Dependent Mean	73.43000	Adj R-Sq	0.1041
Coeff Var	68.30223		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	7689.20435	2154.18888	3.57	0.0006
1st Season	1	-0.00037762	0.00010681	-3.54	0.0006

Note. The p value is low representing a good fit model.

Figure 6

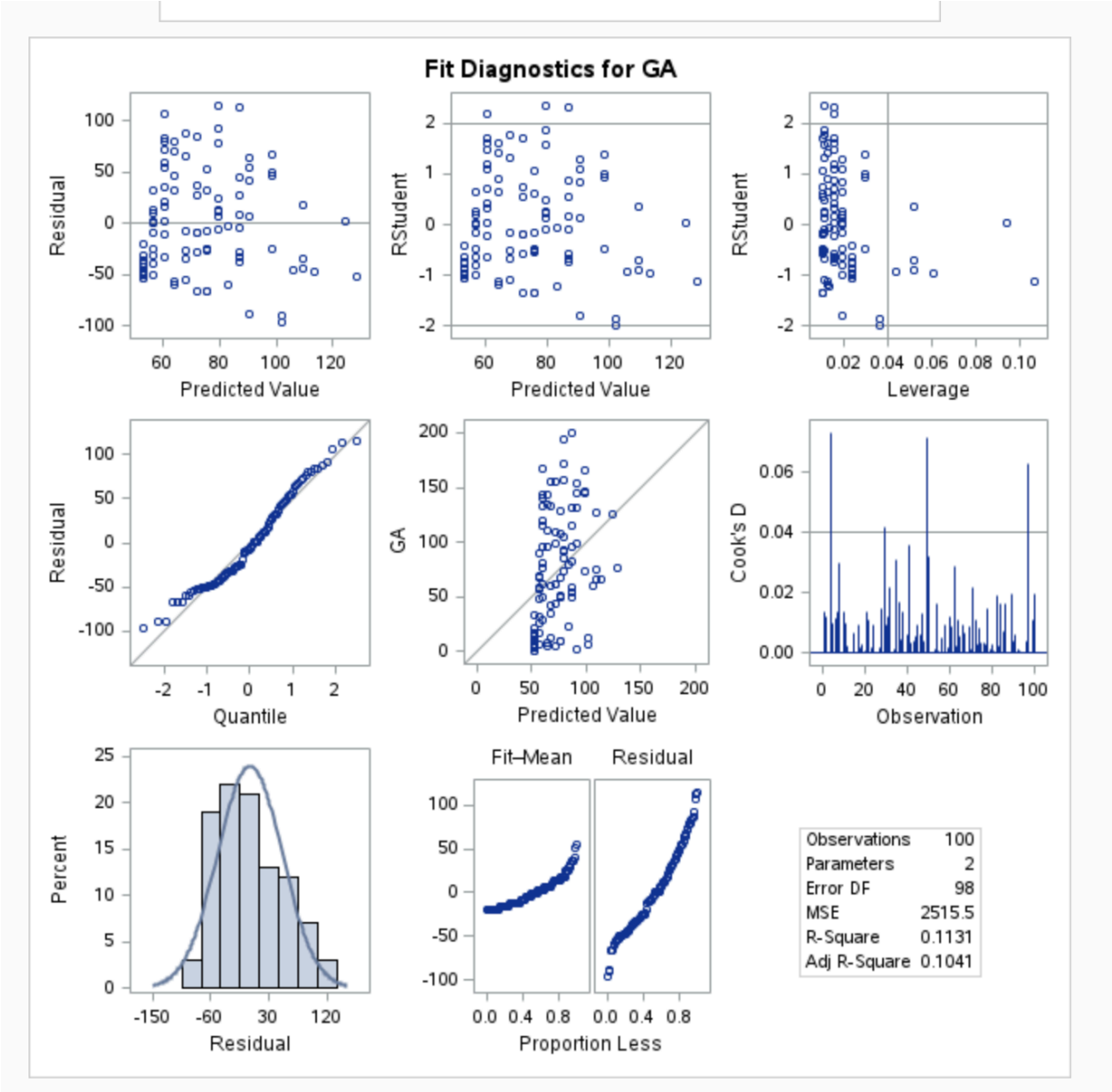
Observed by Predicted for Goals Against



Note. The observations are scattered fairly randomly but do not closely follow the line.

Figure 7

Fit Diagnostics for Goals Against



Note. The quantile-quantile plot closely follows the line representing a good fit.

Testing Question 3

A linear regression demonstrated that there is a relationship between losses and goals against. Losses was set as the dependent variable. Goals against was set as the continuous variable. In figure 8, the analysis of variance is printed. The p value is <0.0001 which is low. A p

value of less than 0.05 represents a good fit model. The lower the value, the better the fit. This p value represents a good fit. Additionally, the r-squared value backs this up. The r-squared value is 0.88 meaning that the goals against can explain about 88% of the variability in losses by a team. In figure 9, the observed by predicted is printed. The observations are randomly scattered and follow the line representing a good fit model. In Figure 10, the fit diagnostics are printed. In the Quantile-Quantile plot, the observations closely follow the line representing a good fit. The distribution mostly follows the normal distribution although there is one high spike in the middle possibly representing outliers. The model is a good fit. Therefore the null hypothesis must be accepted and the alternative hypothesis must be rejected. The null hypothesis can be rejected and the alternative hypothesis can be accepted. There is a relationship between losses and goals against.

Figure 8

Analysis of Variance for Losses

Model: MODEL1
Dependent Variable: L

Number of Observations Read	99
Number of Observations Used	99

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	4861.03661	4861.03661	711.21	<.0001
Error	97	662.98359	6.83488		
Corrected Total	98	5524.02020			

Root MSE	2.61436	R-Square	0.8800
Dependent Mean	9.58586	Adj R-Sq	0.8787
Coeff Var	27.27310		

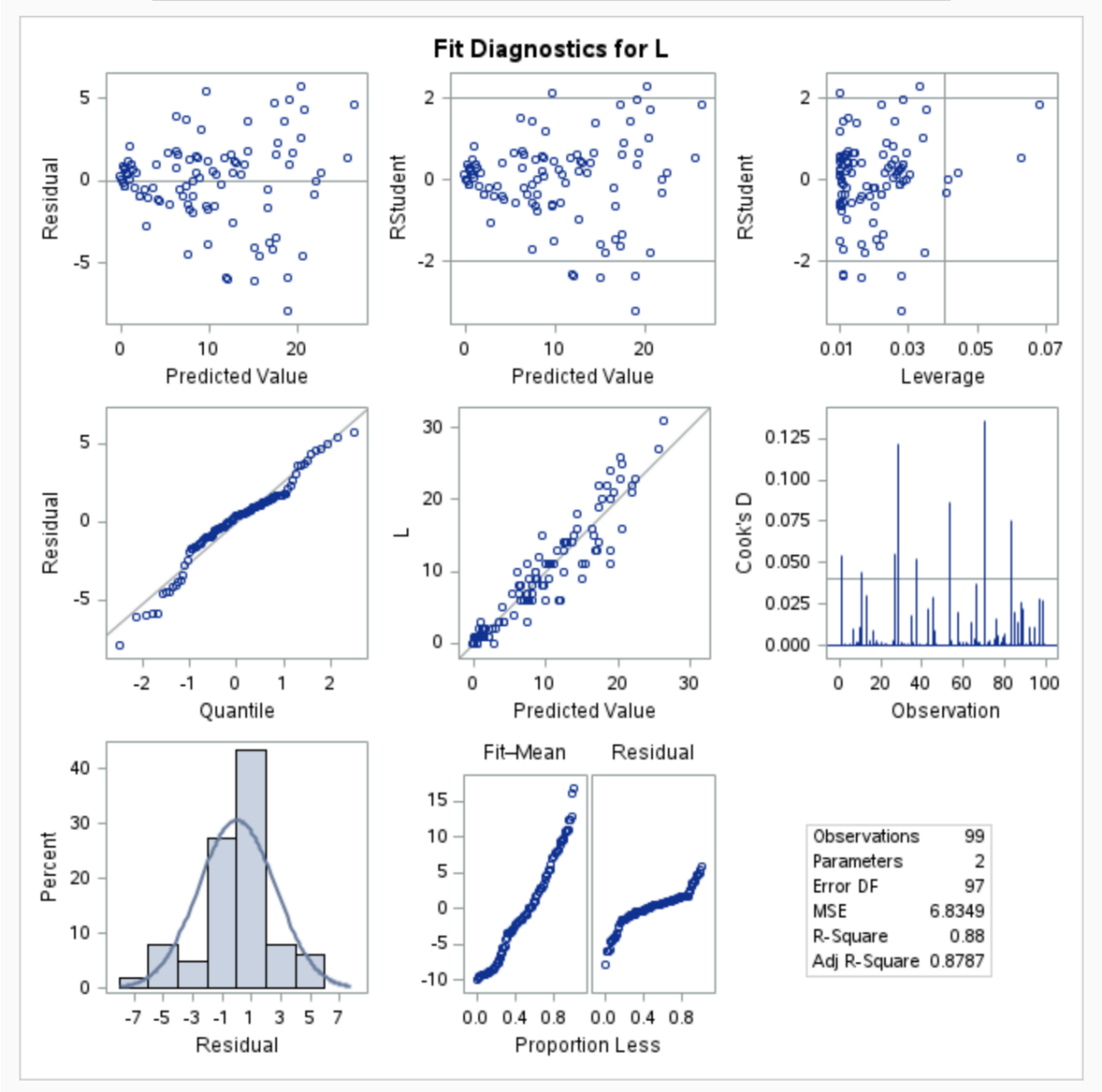
Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-0.32048	0.45500	-0.70	0.4829
GA	1	0.13356	0.00501	26.67	<.0001

Note. The p value is low representing a good fit and the r-squared value is high.

Figure 9

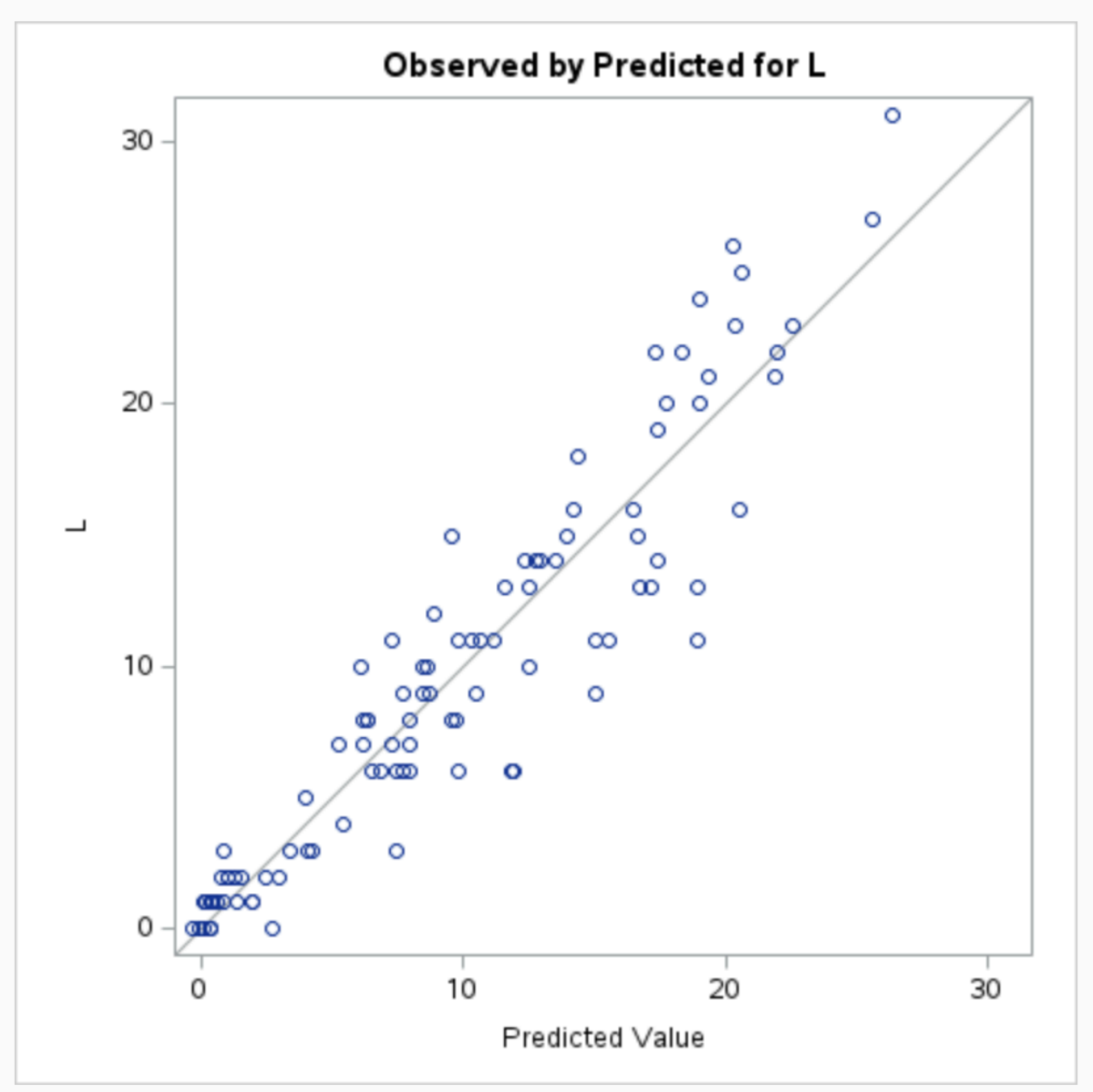
Fit Diagnostics for Losses



Note. The quantile-quantile plot closely follows the line representing a good fit.

Figure 10

Observed by Predicted for Losses



Note. The observations closely follow the line representing a good fit.

Conclusions

In conclusion, the research was able to further the understanding of goalie performance in the NHL and various variables such as rest days, goalie experience, and goals against. Through quantitative analysis, using linear regressions, the impact of rest days on goalie performance, the

impact of goalie experience on goalie performance, and the impact of goals against on team wins were able to be answered by providing information on whether or not the models were a good fit. This was done using p values, r-squared values, and the printed plots to better understand if there was a relationship or not between the variables.

1. Is there a relationship between rest days and save percentages?

The null hypothesis was accepted and the alternative was rejected. There is no relationship between days of rest and save percentages. Due to a high p value and low r-squared value amongst plots that did not turn out to be representative of a strong relationship between the variables, there is little to no relationship between the variables. This means that while rest days may improve the common athlete or fitness guru, days off from a game do not improve the goalies overall performance. The results highlighted the need for further understanding of what rest is as rest days appeared to have no relationship to save percentages. Perhaps due to rest days being defined by goalies not playing in official games, they are not true rest days due to training schedules, quality rest, and traveling.

2. Is there a relationship between goalie experience and wins?

The alternative hypothesis was accepted and the null was rejected. There is a relationship between goalie experience and wins. Due to a low p value and a quantile quantile plot where the observations closely follow the line, the alternative hypothesis was accepted demonstrating that there is a relationship between goalie experience and wins. However, with a low r-squared value, only some of the team wins can be explained by goalie experience meaning this is worth further investigation.

3. Is there a relationship between goals against and team losses?

The alternative hypothesis was accepted and the null was rejected. There is a relationship between goals against and team losses. The p value was very low and the r-squared value was high meaning the model was a good fit. This proves what hockey player Gene Ubriaco stated, “goaltending is 75 percent of the game — unless it's bad goaltending. then it's 100 percent” (Oil, 2023). If goals are let in, the game is lost. A good goalie, or at least one that can keep pucks out of the net, has an effect on a team's ability to win or lose a game. So by improving goalie performance and stopping goals, the game should be easier to win. Therefore, coaches should think wisely about how to build around their goalies strengths and weaknesses to keep pucks out of the net.

In summary, goalies play a huge role in hockey. What they let into the net affects the team's ability to win a game and their experience on the ice affects their overall performance. The more data collected and the more data analyzed, then the more that can be understood about a goalie’s ability to perform. This will lead to a better team performance allowing for an improved game. By building a defense around a goalie and a training schedule around their weaknesses, improvements to a teams ability to win, and in turn, their ranking in the NHL will improve. It is important that the NHL continues to take steps like investing in data collection, data technology, and data analysis to grow the sport of hockey.

Recommendations

The quality of data should be improved to provide more information on rest. To better understand if there is a relationship between rest and performance, athletes diets, sleep, and strain could be measured to help understand the relationship between these variables and recovery. This could be done simply by requiring all athletes to wear athletic bands like a whoop

band. Additionally, more information on goalie experience would provide a deeper understanding of the effect on goalie wins. For example, when did the goalie start skating, playing hockey, and playing competitive hockey before even being drafted to the NHL that might affect their overall performance. Additionally, do any factors cause goalies to stop performing despite experience such as age, burnout, or injury. More data of a different quality would only allow for deeper understanding of goalie performance.

The quality of data should be improved to provide more information on goalie experience in relation to team wins. To better understand the relationship between goalie experience and wins, it may be worth bringing in more variables such as offensive strength, shots on net, and more in a multilinear regression to understand the role a goalie plays in the grand scheme of a win. Additionally, it would be worth further investigating what experience means in the NHL. Just because a goalie has been around since 2015, he may have played less games due to being a third string goalie than a newly brought on goalie who is played as a first string goalie. Additionally, who knows how long those goalies have played over a lifetime of hockey and at what level. Furthermore, goalie age over experience in the NHL relationship to wins should be looked into to see if it is less about overall experience and more related to age of the goalie. This will help the NHL better understand how goalie performance is affected by life experience.

Another next step would be to validate the findings with other hockey leagues or seasons. By testing the models against data from previous seasons, this would demonstrate a good fit and further prove the hypothesis beyond this one season. Additionally, testing it in other leagues would allow this better understanding in how to build strong goalies over a development program versus just in the NHL. For example, the AHL (American Hockey League) and ECHL (East Coast Hockey League) which are semi-professional leagues that often feed into the NHL.

They could also provide valuable data on better predicting goalie performance as many NHL goalies come up through those feeder leagues.

References:

Colorado Avalanche on the Forbes the business of hockey list. (n.d.). Forbes. https://www.forbes.com/teams/colorado-avalanche/?sh=21b20be559e9

Cai, L., & Zhu, Y. (2015). The challenges of data quality and data quality assessment in the big data era. Data science journal, 14, 2-2.

Chatterjee, S., & Simonoff, J. S. (2013). *Handbook of regression analysis (Vol. 5).* John Wiley & Sons.

Cody, R. (2018). *An introduction to SAS university edition.* SAS Institute.

Crittenden, C. Analyzing NHL Goalie Stats.

EDGE.NHL.com. *(2023). NHL Edge Overview.* https://edge.nhl.com/en/home

Gilmartin, K. (2021, October 1). *Rest and Active Recovery: An Important Step for All Athletes*. Children's Hospital of the King's Daughter. https://www.chkd.org/blog/rest-and-active-recovery--an-important-step-for-all-athletes/#:~:text=Rest%20days%20allow%20the%20body's,that%20can't%20repair%20itself.

Gladwell, M. (2009). *Outliers*. Back Bay Books.

Larkinnov, M. (2021, November 25). *SECRET FORMULA: ANALYTICS IN THE NHL. The Hockey News.* Retrieved January 28, 2024, from https://thehockeynews.com/news/secret-formula-analytics-in-the-nhl

Naples, M., Gage, L., & Nussbaum, A. (2018). Goalie analytics: Statistical evaluation of context-specific goalie performance measures in the National Hockey League. SMU Data Science Review, 1(2), 12.

NHL records. (n.d.). https://records.nhl.com/history

Oil In Goal. (2023, February 23). *Quantifying the impact of goaltending on the Oilers*. Quantifying the Impact of Goaltending on the Oilers. https://oilingoal.substack.com/p/quantifying-the-impact-of-goaltending

Williams, W. B. H., & Williams, D. A. (1998). Performance indices for multivariate ice hockey statistics.

Wu, J., Shen, Y., Qian, D., & Cui, Y. (2023). Effect of performance determinants on nhl final goal difference. Revista Brasileira de Medicina do Esporte, 30, e2023_0091.