

UConn Men's Hockey Data Analysis

Giovanni Lunetta and Alex Pugh

Outline

1. Abstract
2. Goal/Motivation
3. Methods
4. Data
5. Data Analysis
6. Discussion & Conclusion

Abstract

This study focuses on examining the relationship between power play performance and game outcomes for the UConn Men's Hockey Team across the 2019-2020 to 2021-2022 seasons. By analyzing data from 127 games, we investigate various metrics such as power play efficiency, faceoff success, zone possession percentages, and shot/goal distribution during power plays. This approach aims to provide insights into the strategic aspects of power play that most significantly impact winning games.

Goal/Motivation

The primary goal of this analysis is to determine how different aspects of power play performance correlate with the team's success in winning games. By understanding these relationships, the team can focus on specific areas of improvement during power plays. Additionally, this analysis seeks to provide a comprehensive view of the team's power play strategies and their effectiveness, thereby aiding in tactical decision-making.

Methods

The primary analytical technique employed in this study is logistic regression, chosen for its suitability in modeling binary outcomes, in this case, the win/loss record of the UConn Men's Hockey Team. The logistic regression model will incorporate a range of predictors related to power play performance, including scores on power plays, faceoff success rates, and shot/goal distribution. These predictors are selected based on their potential influence on game outcomes. The model will also account for variables like home advantage and opponent strength. The focus here is on the statistical approach and model specification rather than the dataset specifics, which are detailed in the "Data" section.

Data

The dataset for this analysis includes detailed game metrics from 127 games spanning the 2019-2020 to 2021-2022 seasons. However, 18 games were excluded from the analysis for the following reasons:

- 11 games were removed because there were no powerplay faceoffs taken in the OZ during those games.
- 3 games were removed because there was no possession in the DZ during the powerplay.
- 1 game was removed because there was no possession in the OZ or the NZ during the powerplay.
- 2 games were removed due to data inconsistencies causing the possession percentage in the OZ to exceed 100% of the total elapsed time, which is not possible.

Our justification for removing these columns is as follows:

Our dataset differentiates between a 0% faceoff win rate in the offensive zone (which implies faceoffs occurred in that zone) and a '-' symbol, which indicates that no faceoffs took place in the offensive zone (this is the same for possession columns as well). It was deemed impractical to include these games, as this would distort the analysis of power play effectiveness. This reasoning is the same for the games with no possession in the DZ or OZ/NZ. The two games with possession percentages exceeding 100% were removed because this is not possible, and the data was likely entered incorrectly.

The remaining dataset and data we will use for our analysis is 110 games and 39 variables (or predictors). The variables are as follows:

Note: All variables are specifically during power plays unless otherwise stated.

- **Date** - The date the game was played
- **Opponent** - The opponent UConn played
- **Team_Total_UConn Score** - The total number of goals UConn scored
- **Team_Total_Opponent Score** - The total number of goals the opponent scored

- Home - Whether the game was played at home or away (1 = Home, 0 = Away)
- Win - Whether UConn won the game (1 = Win, 0 = Loss)
- Team_Score_On_PP - Whether UConn scored on the power play (1 = Score, 0 = No Score)
- Team_PP_Faceoffs_won_in_OZ, % - The percentage of faceoffs won in the offensive zone
- Team_PP_Faceoffs_won, % - The percentage of faceoffs won
- Team_Shots_OT_Behind_Net - The number of shots taken behind the net that were on target
- Team_Shots_Behind_Net - The number of shots taken behind the net
- Team_Goals_Behind_Net - The number of goals scored behind the net
- Team_Shots_OT_at_Net - The number of shots taken at the net that were on target
- Team_Shots_at_Net - The number of shots taken at the net
- Team_Goals_at_Net - The number of goals scored at the net
- Team_Shots_OT_Left_Close - The number of shots taken on the left side of the net that were on target
- Team_Shots_Left_Close - The number of shots taken on the left side of the net
- Team_Goals_Left_Close - The number of goals scored on the left side of the net
- Team_Shots_OT_Center_Close - The number of shots taken in the center of the net that were on target
- Team_Shots_Center_Close - The number of shots taken in the center of the net
- Team_Goals_Center_Close - The number of goals scored in the center of the net
- Team_Shots_OT_Right_Close - The number of shots taken on the right side of the net that were on target
- Team_Shots_Right_Close - The number of shots taken on the right side of the net
- Team_Goals_Right_Close - The number of goals scored on the right side of the net
- Team_Shots_OT_Left_Far - The number of shots taken on the left side of the net that were on target
- Team_Shots_Left_Far - The number of shots taken on the left side of the net
- Team_Goals_Left_Far - The number of goals scored on the left side of the net
- Team_Shots_OT_Center_Far - The number of shots taken in the center of the net that were on target
- Team_Shots_Center_Far - The number of shots taken in the center of the net
- Team_Goals_Center_Far - The number of goals scored in the center of the net
- Team_Shots_OT_Right_Far - The number of shots taken on the right side of the net that were on target
- Team_Shots_Right_Far - The number of shots taken on the right side of the net
- Team_Goals_Right_Far - The number of goals scored on the right side of the net
- Team_Shots_OT_Not_in_Offensive_Zone - The number of shots taken that were on target but not in the offensive zone
- Team_Shots_Not_in_Offensive_Zone - The number of shots taken that were not in the offensive zone
- Team_Goals_Not_in_Offensive_Zone - The number of goals scored that were not in the offensive zone

- **Team_PP_OZ possession, %** - The percentage of time UConn had possession in the offensive zone
- **Team_PP_NZ possession, %** - The percentage of time UConn had possession in the neutral zone
- **Team_PP_DZ possession, %** - The percentage of time UConn had possession in the defensive zone

Analysis

Lets start by getting a look at the data. We can sample 5 of the games:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import statsmodels.api as sm
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split

df = pd.read_csv('/Users/giovanni-lunetta/uconn_masters/hockey/hockey_repo/data/final_data')

pd.set_option('display.max_columns', None)
df.sample(5)
```

	Date	Opponent	Team_Total_UConn_Score	Team_Total_Opponent_Score
30	2022-10-01	Vermont Catamounts	4.0	1.0
64	2021-02-09	UMass-Lowell River Hawks	2.0	3.0
21	2022-11-05	Maine Black Bears	3.0	2.0
70	2021-01-16	New Hampshire Wildcats	8.0	3.0
35	2022-02-26	Northeastern Huskies	2.0	5.0

As well as look at the summary statistics for each variable:

```
df.describe()
```

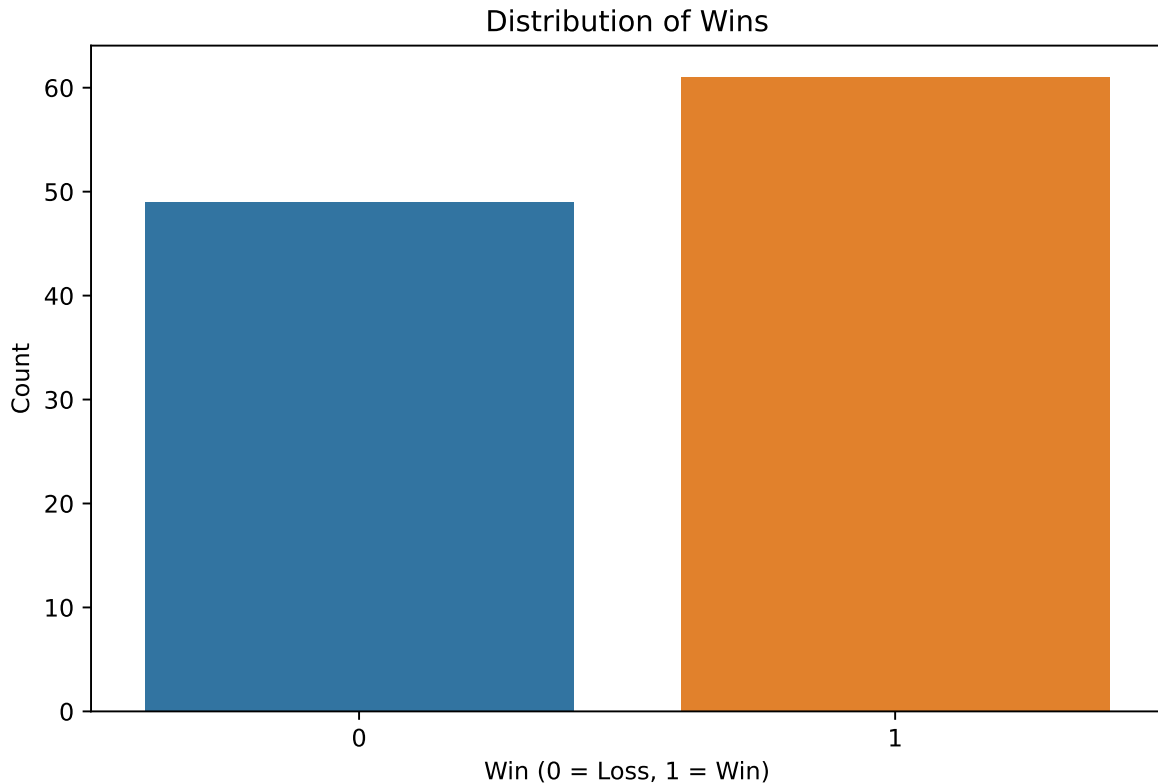
	Team_Total_UConn_Score	Team_Total_Opponent_Score	Home	Win	Team_Score
count	110.000000	110.000000	110.000000	110.000000	110.000000
mean	3.127273	2.827273	0.481818	0.554545	0.400000

	Team_Total_UConn_Score	Team_Total_Opponent_Score	Home	Win	Team_Score
std	1.877482	1.630327	0.501956	0.499291	0.49214
min	0.000000	0.000000	0.000000	0.000000	0.00000
25%	2.000000	1.000000	0.000000	0.000000	0.00000
50%	3.000000	3.000000	0.000000	1.000000	0.00000
75%	4.000000	4.000000	1.000000	1.000000	1.00000
max	8.000000	7.000000	1.000000	1.000000	1.00000

Analysis (cont.)

We can also visualize the distribution of our dependent variable, `win`.

```
# Plotting the distribution of the dependent variable 'Win'
plt.figure(figsize=(8, 5))
sns.countplot(x='Win', data=df)
plt.title('Distribution of Wins')
plt.xlabel('Win (0 = Loss, 1 = Win)')
plt.ylabel('Count')
plt.show()
```



Analysis (cont.)

```
# Removing the duplicate column 'Team_Shots_Behind_Net'
df = df.drop('Team_Shots_Behind_Net', axis=1)

# Checking for columns that have all zeros
columns_all_zeros = df.columns[(df == 0).all()]
columns_all_zeros

# Dropping columns with all zeros
df = df.drop(columns=columns_all_zeros)

# Dropping non-numeric columns for simplicity
df = df.drop(columns=['Date', 'Opponent'])
df = df.drop(columns=['Team_Total_UConn_Score', 'Team_Total_Opponent_Score'])
df = df.drop(columns=['Team_Shots_OT_Behind_Net'])
df = df.drop(columns=['Team_PP_Faceoffs_won_in_OZ_percent'])
```

After some preprocessing we can run the logistic regression model:

```
import statsmodels.formula.api as smf

# Joining all column names with '+' except 'Win', 'Date', and 'Opponent'
formula = 'Win ~ ' + ' + '.join(df.columns.drop('Win'))

# Fitting the logistic regression model using smf.logit
logit_model = smf.logit(formula=formula, data=df).fit()

# Displaying the summary of the model
logit_model_summary = logit_model.summary()
print(logit_model_summary)
```

Warning: Maximum number of iterations has been exceeded.

Current function value: 0.501595

Iterations: 35

Logit Regression Results

```
=====
Dep. Variable:          Win    No. Observations:          110
Model:                  Logit    Df Residuals:              80
Method:                  MLE     Df Model:                29
Date:                   Mon, 11 Dec 2023    Pseudo R-squ.:          0.2701
Time:                   13:46:19    Log-Likelihood:         -55.175
converged:               False    LL-Null:                -75.590
Covariance Type:         nonrobust    LLR p-value:            0.07126
=====
```

	coef	std err	z	P> z	[0.025
Intercept	-4.0347	1.819	-2.218	0.027	-7.600
Home	0.2455	0.538	0.456	0.648	-0.809
Team_Score_On_PP	-0.7143	0.877	-0.815	0.415	-2.433
Team_PP_Faceoffs_won_percent	0.0324	0.015	2.197	0.028	0.004
Team_Shots_OT_at_Net	1.0630	0.687	1.548	0.122	-0.283
Team_Shots_at_Net	-0.4428	0.550	-0.804	0.421	-1.522
Team_Goals_at_Net	0.2230	0.665	0.335	0.737	-1.080
Team_Shots_OT_Left_Close	0.7856	0.478	1.643	0.100	-0.151
Team_Shots_Left_Close	-0.4516	0.383	-1.180	0.238	-1.202
Team_Goals_Left_Close	1.7281	1.727	1.001	0.317	-1.657
Team_Shots_OT_Center_Close	-0.1511	0.345	-0.438	0.661	-0.827
Team_Shots_Center_Close	-0.2721	0.216	-1.258	0.208	-0.696
Team_Goals_Center_Close	1.3111	0.847	1.548	0.122	-0.349

Team_Shots_OT_Right_Close	0.5649	0.457	1.236	0.216	-0.331
Team_Shots_Right_Close	0.0174	0.369	0.047	0.962	-0.707
Team_Goals_Right_Close	-35.4667	5.56e+07	-6.38e-07	1.000	-1.09e+08
Team_Shots_OT_Left_Far	0.2178	0.760	0.286	0.775	-1.272
Team_Shots_Left_Far	0.0757	0.472	0.160	0.873	-0.850
Team_Goals_Left_Far	36.5677	6.71e+07	5.45e-07	1.000	-1.32e+08
Team_Shots_OT_Center_Far	-0.0457	0.446	-0.102	0.918	-0.920
Team_Shots_Center_Far	-0.2383	0.212	-1.122	0.262	-0.655
Team_Goals_Center_Far	25.9586	4.65e+04	0.001	1.000	-9.1e+04
Team_Shots_OT_Right_Far	0.3775	0.706	0.535	0.593	-1.007
Team_Shots_Right_Far	-0.3988	0.457	-0.872	0.383	-1.295
Team_Goals_Right_Far	-37.0806	5.56e+07	-6.67e-07	1.000	-1.09e+08
Team_Shots_OT_Not_in_Offensive_Zone	16.8973	2493.072	0.007	0.995	-4869.435
Team_Shots_Not_in_Offensive_Zone	-15.1358	2493.072	-0.006	0.995	-4901.467
Team_PP_OZ_possession_percent	0.0408	0.020	2.046	0.041	0.002
Team_PP_NZ_possession_percent	-0.1577	0.083	-1.890	0.059	-0.321
Team_PP_DZ_possession_percent	0.1228	0.064	1.930	0.054	-0.002

=====

```
/Users/giovanni-lunetta/Library/Python/3.10/lib/python/site-packages/statsmodels/base/model.py
warnings.warn("Maximum Likelihood optimization failed to "
```

Analysis (cont.)

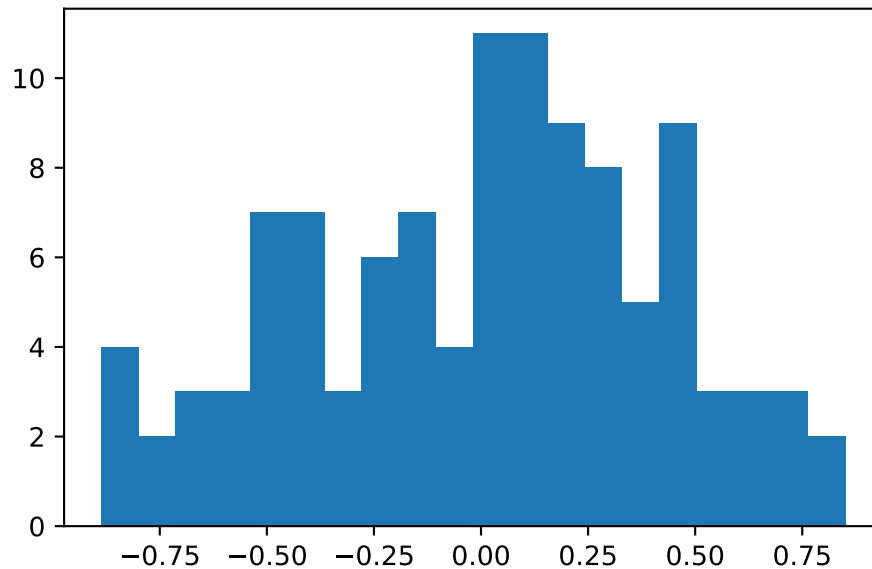
Before commenting on the results, we can confirm that the underlying assumptions are not violated by checking residuals and multicollinearity. Because games deemed to be “outliers” can still be useful in our analysis, we will not remove any data points. We also do not check for independence as this is satisfied by the nature of the data.

```
import matplotlib.pyplot as plt

residuals = logit_model.resid_generalized
plt.hist(residuals, bins=20)

(array([ 4.,  2.,  3.,  3.,  7.,  7.,  3.,  6.,  7.,  4., 11., 11.,  9.,
        8.,  5.,  9.,  3.,  3.,  3.,  2.]),
 array([-0.88656364, -0.79971038, -0.71285711, -0.62600385, -0.53915058,
        -0.45229731, -0.36544405, -0.27859078, -0.19173752, -0.10488425,
        -0.01803098,  0.06882228,  0.15567555,  0.24252881,  0.32938208,
        0.41623535,  0.50308861,  0.58994188,  0.67679514,  0.76364841,
        0.85050168]),
```


<BarContainer object of 20 artists>)



```
from statsmodels.stats.outliers_influence import variance_inflation_factor

# Selecting only the independent variables
X = df.drop('Win', axis=1)

# Calculating VIF for each feature
vif_data = pd.DataFrame()
vif_data['feature'] = X.columns
vif_data['VIF'] = [variance_inflation_factor(X.values, i) for i in range(len(X.columns))]

vif_data.sort_values(by='VIF', ascending=False)

df = df.drop(columns=['Team_Shots_OT_at_Net'])
```

The residuals are normally distributed, and the only variable of concern in terms of multicollinearity is `Team_Shots_OT_at_Net` which we removed. This variable is intuitively highly correlated with `Team_Shots_at_Net` as a very high proportion of shots at the net are on target. This is confirmed by the VIF values, which were both near 20 before dropping this variable.

We also saw that the model was not able to converge, which is likely due to the fact that some variables have very large coefficients, most likely due to their high variability where the model is unable to converge to a solution. We drop these columns as well.

```
df = df.drop(columns=['Team_Goals_Right_Close', 'Team_Goals_Left_Far', 'Team_Goals_Center_
```

Analysis (cont.)

Now we can rerun the model and look at performance metrics:

```
import statsmodels.formula.api as smf

# Joining all column names with '+' except 'Win', 'Date', and 'Opponent'
formula = 'Win ~ ' + ' + '.join(df.columns.drop('Win'))

# Fitting the logistic regression model using smf.logit
logit_model = smf.logit(formula=formula, data=df).fit()

# Displaying the summary of the model
logit_model_summary = logit_model.summary()
print(logit_model_summary)
```

Optimization terminated successfully.

Current function value: 0.576667

Iterations 6

Logit Regression Results

=====						
Dep. Variable:	Win	No. Observations:	110			
Model:	Logit	Df Residuals:	87			
Method:	MLE	Df Model:	22			
Date:	Mon, 11 Dec 2023	Pseudo R-squ.:	0.1608			
Time:	13:46:19	Log-Likelihood:	-63.433			
converged:	True	LL-Null:	-75.590			
Covariance Type:	nonrobust	LLR p-value:	0.3310			
=====						
	coef	std err	z	P> z	[0.025	0

Intercept	-3.8386	1.679	-2.286	0.022	-7.130	-0.547
Home	0.4161	0.487	0.854	0.393	-0.539	1.371
Team_Score_On_PP	-0.6707	0.734	-0.914	0.361	-2.110	0.769
Team_PP_Faceoffs_won_percent	0.0320	0.014	2.363	0.018	0.005	0.059
Team_Shots_at_Net	0.4058	0.209	1.945	0.052	-0.003	0.814
Team_Goals_at_Net	0.4165	0.588	0.708	0.479	-0.736	1.569
Team_Shots_OT_Left_Close	0.5352	0.400	1.338	0.181	-0.249	1.319

Team_Shots_Left_Close	-0.2931	0.323	-0.908	0.364	-0.926	0
Team_Goals_Left_Close	1.3515	1.635	0.827	0.408	-1.853	4
Team_Shots_OT_Center_Close	-0.0171	0.328	-0.052	0.958	-0.660	0
Team_Shots_Center_Close	-0.2185	0.206	-1.063	0.288	-0.621	0
Team_Goals_Center_Close	1.0175	0.741	1.373	0.170	-0.435	2
Team_Shots_OT_Right_Close	0.4856	0.408	1.192	0.233	-0.313	1
Team_Shots_Right_Close	-0.1238	0.341	-0.363	0.717	-0.792	0
Team_Shots_OT_Left_Far	-0.0198	0.638	-0.031	0.975	-1.271	1
Team_Shots_Left_Far	0.2053	0.417	0.493	0.622	-0.612	1
Team_Shots_OT_Center_Far	0.0941	0.353	0.266	0.790	-0.598	0
Team_Shots_Center_Far	-0.2046	0.189	-1.085	0.278	-0.574	0
Team_Shots_OT_Right_Far	0.2789	0.629	0.443	0.658	-0.955	1
Team_Shots_Right_Far	-0.3100	0.429	-0.722	0.470	-1.151	0
Team_PP_OZ_possession_percent	0.0385	0.019	2.060	0.039	0.002	0
Team_PP_NZ_possession_percent	-0.1078	0.071	-1.512	0.131	-0.248	0
Team_PP_DZ_possession_percent	0.0662	0.053	1.259	0.208	-0.037	0

=====

```

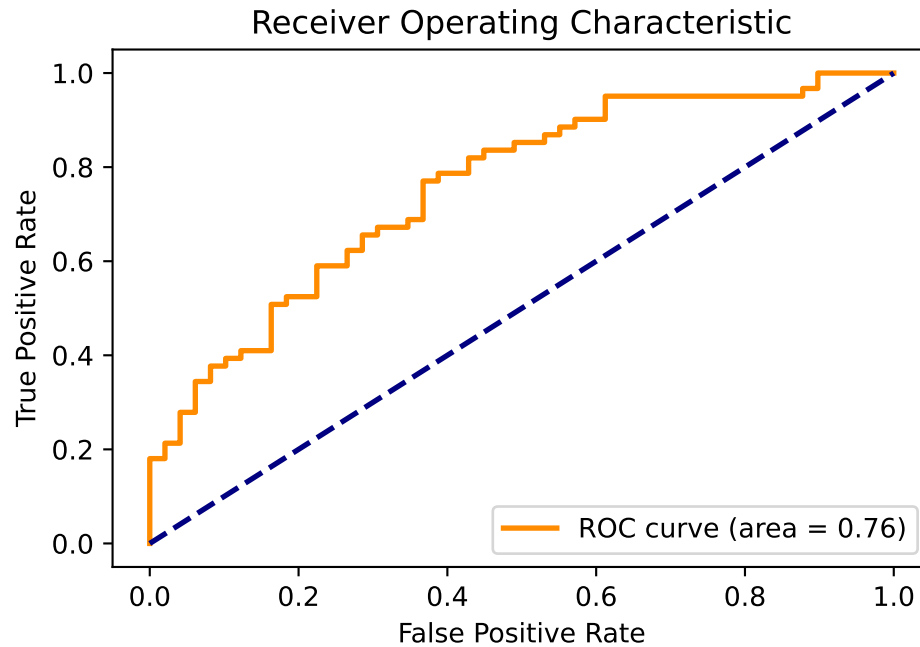
from sklearn.metrics import roc_curve, auc

predictions = logit_model.predict(df)

fpr, tpr, thresholds = roc_curve(df['Win'], predictions)
roc_auc = auc(fpr, tpr)

plt.figure()
plt.plot(fpr, tpr, color='darkorange', lw=2, label='ROC curve (area = %0.2f)' % roc_auc)
plt.plot([0, 1], [0, 1], color='navy', lw=2, linestyle='--')
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('Receiver Operating Characteristic')
plt.legend(loc="lower right")
plt.show()

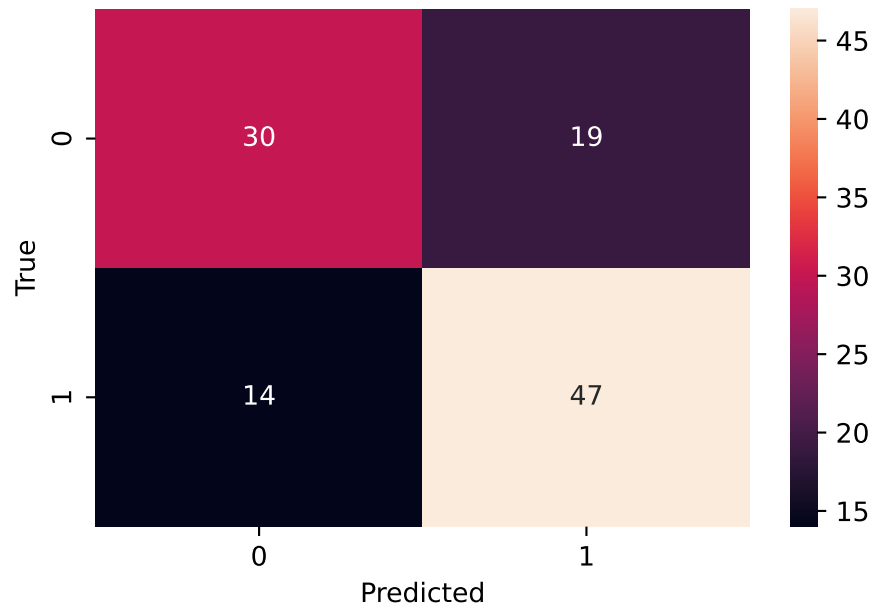
```



```
from sklearn.metrics import confusion_matrix
import seaborn as sns

# Convert probabilities to 0/1 predictions based on a chosen threshold (e.g., 0.5)
pred_labels = predictions > 0.5
cm = confusion_matrix(df['Win'], pred_labels)

sns.heatmap(cm, annot=True)
plt.xlabel('Predicted')
plt.ylabel('True')
plt.show()
```



```
import numpy as np
import statsmodels.api as sm

# Assuming logit_model is your fitted logistic regression model
# And X_test is your test features dataset

predicted_probabilities = logit_model.predict(X)
predicted_labels = np.where(predicted_probabilities > 0.5, 1, 0) # Using 0.5 as the threshold

from sklearn.metrics import precision_score, recall_score, f1_score, accuracy_score

precision = precision_score(df['Win'], predicted_labels)
recall = recall_score(df['Win'], predicted_labels)
f1 = f1_score(df['Win'], predicted_labels)
accuracy = accuracy_score(df['Win'], predicted_labels)

print(f"Precision: {precision}")
print(f"Recall: {recall}")
print(f"F1-Score: {f1}")
print(f"Accuracy: {accuracy}")
```

Precision: 0.7121212121212122

Recall: 0.7704918032786885
F1-Score: 0.7401574803149605
Accuracy: 0.7

Analysis of Results

Now we can comment on the results:

Intercept: The model's intercept is significant, indicating that even when all predictor variables are at zero, the log odds of winning vs. losing are significantly different from zero.

Home ($p = 0.393$): This variable is not statistically significant, suggesting that playing at home does not significantly impact the likelihood of winning in this model.

Team_Score_On_PP ($p = 0.361$): Scoring on the power play is not a significant predictor of winning games, according to the model. This might seem counterintuitive, but it could indicate that other aspects of power play performance are more critical.

Team_PP_Faceoffs_won_percent ($p = 0.018$): This variable is significant, showing that winning a higher percentage of faceoffs during power plays is positively correlated with winning games.

Team_Shots_at_Net ($p = 0.052$): Though borderline, this variable is not statistically significant at the 0.05 level. It suggests that the number of shots at the net during power plays doesn't have a statistically significant effect on winning games.

Team_Goals_at_Net ($p = 0.479$): The number of goals scored at the net during power plays is not a significant predictor of winning games.

Team_Shots_OT_Left_Close ($p = 0.181$): Shots on target from the left close range are not a significant predictor of winning.

Team_Shots_Left_Close ($p = 0.364$): Similarly, the total number of shots from the left close range does not significantly impact the likelihood of winning.

Team_Goals_Left_Close ($p = 0.408$): The number of goals scored from the left close range is not a significant predictor.

Team_Shots_OT_Center_Close ($p = 0.958$): Shots on target from the center close range do not significantly impact winning chances.

Team_Shots_Center_Close ($p = 0.288$): The total number of shots from the center close range is not a significant predictor.

Team_Goals_Center_Close ($p = 0.170$): Goals scored from the center close range also do not significantly affect the chances of winning.

Team_Shots_OT_Right_Close ($p = 0.233$): Shots on target from the right close range are not a significant predictor.

Team_Shots_Right_Close ($p = 0.717$): The total number of shots from the right close range does not significantly influence game outcomes.

Team_Shots_OT_Left_Far ($p = 0.975$): Shots on target from the left far range are not a significant predictor.

Team_Shots_Left_Far ($p = 0.622$): The total number of shots from the left far range does not significantly impact the likelihood of winning.

Team_Shots_OT_Center_Far ($p = 0.790$): Shots on target from the center far range are not significant predictors of winning games.

Team_Shots_Center_Far ($p = 0.278$): The total number of shots from the center far range is not a significant predictor.

Team_Shots_OT_Right_Far ($p = 0.658$): Shots on target from the right far range do not significantly impact winning chances.

Team_Shots_Right_Far ($p = 0.470$): The total number of shots from the right far range is not a significant predictor.

Team_PP_OZ_possession_percent ($p = 0.039$): The percentage of time possessing the puck in the offensive zone during power plays is significant, suggesting a positive impact on winning games.

Team_PP_NZ_possession_percent ($p = 0.131$): Possession in the neutral zone during power plays is not a significant predictor of winning.

Team_PP_DZ_possession_percent ($p = 0.208$): Similarly, possession in the defensive zone during power plays does not significantly influence game outcomes.

In summary, the most significant factors at the 0.05 alpha level are winning faceoffs during power plays and maintaining possession in the offensive zone. Many of the shot-related variables, both in terms of attempts and goals, did not emerge as significant predictors in this model. This analysis provides a nuanced understanding of what aspects of power play performance are most crucial to game outcomes for the UConn Men's Hockey Team.

Additionally, the model's ROC curve area of 0.76 suggests a good discriminatory ability between winning and losing outcomes. Also, Precision, Recall, and F1-Score values (0.71, 0.77, and 0.74 respectively) indicate a balanced and effective model in predicting game outcomes.

Discussion & Conclusion

Discussion

Interpretation of Key Findings

Our logistic regression analysis reveals intriguing insights into the dynamics of power play performance and their impact on game outcomes for the UConn Men's Hockey Team. The significance of certain variables over others provides a nuanced understanding of strategic elements in hockey.

1. Significance of Faceoffs (Team_PP_Faceoffs_won_percent):

- The statistical significance of faceoff wins during power plays underscores the tactical advantage of gaining initial control of the puck. Winning faceoffs sets the tone for the subsequent play, allowing for immediate offensive pressure and potentially disorganizing the opponent's defense.

2. Offensive Zone Possession (Team_PP_OZ_possession_percent):

- The positive correlation between offensive zone possession during power plays and winning outcomes suggests the importance of sustained offensive pressure. It highlights the need for effective puck handling and strategic positioning to maintain control in the offensive zone.

3. Non-significance of Shot and Goal Metrics:

- Interestingly, most variables related to shots and goals, whether from close or far ranges, did not significantly predict winning outcomes. This might indicate that the sheer number of shots or goals during power plays isn't as crucial as the quality and strategic execution of these shots. It also suggests that other factors, like defensive strategies or goaltending, play a significant role in determining the success of power plays.

4. Home Advantage (Home):

- The lack of statistical significance for the 'Home' variable suggests that the home-ice advantage, often considered an influential factor in sports, may not be as impactful in determining the outcomes of the games within the scope of this analysis. This could be due to a well-adapted team performance irrespective of the venue or other balancing factors like travel fatigue not being captured in the model.

Limitations

While our analysis provides valuable insights, it is crucial to acknowledge its limitations:

- **Data Scope:** Our analysis is confined to three seasons of data. Longer-term trends or changes in team composition and strategy over time are not captured.
- **Unmeasured Factors:** Intangible elements like team morale, player injuries, psychological factors, and the quality of opposition strategies are not quantified in our dataset but can significantly influence game outcomes.
- **Model Constraints:** The logistic regression model, while robust, may not capture the complex interactions and non-linear relationships that can exist in sports data.

Recommendations

Based on our findings, we recommend:

1. **Strategic Emphasis on Faceoffs and Offensive Zone Control:** Coaches and players should focus on strategies and training drills that enhance faceoff success and control in the offensive zone.
2. **Quality Over Quantity in Shot Taking:** Rethink the approach to shooting, emphasizing strategic and well-positioned shots over mere volume. Location of shot attempts are not likely to be as significant of a factor as the quality of the shot.

Conclusion

Our logistic regression analysis has shed light on the complex interplay of factors that contribute to successful outcomes in hockey, particularly during power plays. The significant role of faceoff wins and offensive zone possession during power plays emerged as key strategic elements. Contrary to conventional wisdom, the number of shots or goals during power plays was not predictive of winning outcomes, suggesting a need to focus on the quality and execution of offensive strategies.

This analysis serves as a foundational piece for the UConn Men's Hockey Team in formulating data-driven strategies. It encourages a nuanced approach to understanding the multifaceted nature of hockey and the variables that influence game outcomes. Future analyses could expand on this work by incorporating a broader dataset, considering additional variables, and employing more complex statistical models to capture the dynamic nature of the sport.