Who's the Next Winner? Using Conditional Logistic Regression Model to Predict Results for Academy Awards

Yun-Tsz Tsai

March 13, 2025

Introduction

Academy Awards, also known as Oscars, is one of the most prestigious awards in the film industry. It is presented annually in the first quarter of the year. Aside from Academy Awards, there are several other similar awards being awarded, for instance, Golden Globe Awards, which is held at the end of a year. The first Academy Awards was held in 1929, reviwing films made in 1928. Some of its awards go to individuals such as the actors and actresses in a film while some go to a film.

As actors and actresses will be evaluated on their performance in the film they participated, it is reasonale to assume that their peformance will also affect the evaluation of the film. Previous studies have shown that being nominated for other Oscars awards (e.g., best directing) or having won a non-Oscars award can have an impact on whether a film will subsequently win an Oscars. However, not all types of nominations seem to be equal. A study by Pardoe, I., & Simonton, D. K. (2008) shows that if the leading actress has been nominated for the best actress in a leading role, it might lower the possibility of the film winning a best picture.

In this study, I will examine the correlation between having strong acting performance (i.e., either having been nominated for other Oscars awards or already won a non-Oscars) and winning Academy Awards for best picture, using conditional logistic regression model. Three hypotheses are formulated as follows,

1. The total number of nominations will affect the possibility of a film winning best picture.
2. The performance of actors and actresses will affect the possibility of a film winning best picture.

3. Whether or not a file has won a non-Oscars will affect the possibility of a film winning best picture.

In addition, predictions on the results of Oscars for 2025 will be presented using a model that best describes previous data.

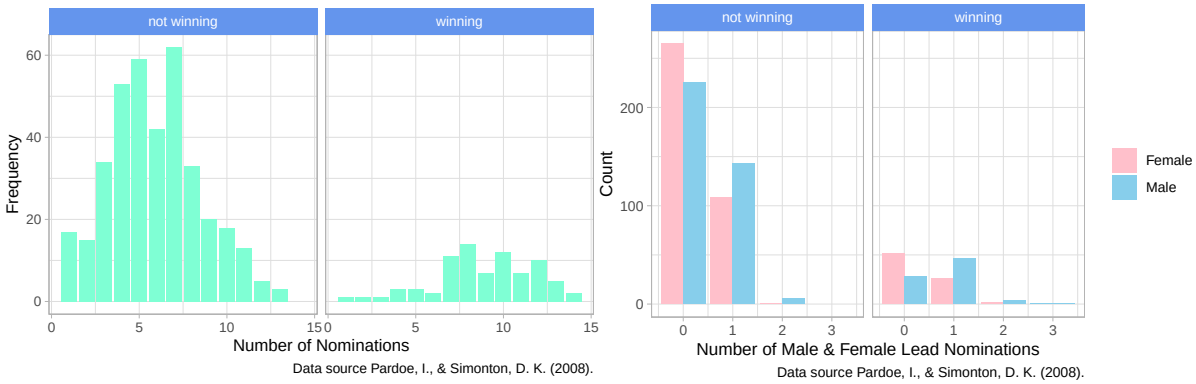
Data

The data being used in this study covers the records from 1928 to 2006. In total, there are 453 observations and for each year, one of the nominations for Oscars best picture will be awarded such honor. The dependent variable of interest is the result of Oscars best picture. The winner will be coded as 1 while those that did not win will be coded as 0.

Based on previous studies and my hypotheses, the independent variables I will examine are 1) the total number of Oscar nominations, 2) being nominated for best directing, 3) the number of best actor in leading role 4) the number of best actress in leading role 5) having won Golden Globe in the previous year. Figure 1 and Table 1 show the summary statistics and distribution of nominations respectively. Except for 1), 3) and 4), all the other variables are binary, taking either 0 (indicates not winning) or 1 (indicates winning).

If we take a closer look at the data, we will notice that the average total number of nominations for best picture winners is 50% higher than those who did not win. Among those who won, almost all of them (95%) the director had won the award for best directing compared to only 5% for those who won without having received best directing award. Besides, several of the movies that won the best picture have previously won a Golden Globe.

The distribution of the number of best leading role for both actors and actresses are not very skewed between the winning movies and movies that did not win. However, the distribution of the total number of nominations is more noticeably different between the winning movies and movies that did not win.



(a) Total Nominations

(b) Male and Female Actor Lead Nomination

Figure 1. Distributions of Variables by Winning or Not

Table 1. Descriptive Data on Variables

Variable	**Overall** N = 453[1]	**Winning**	
		0 N = 374[1]	**1** N = 79[1]
Oscar Nominations	6 (4, 8)	6 (4, 7)	9 (7, 11)
Director Nomination	281 (62%)	206 (55%)	75 (95%)
Best Actor in Leading Role			
0	253 (56%)	225 (60%)	28 (35%)
1	189 (42%)	143 (38%)	46 (58%)
2	10 (2.2%)	6 (1.6%)	4 (5.1%)
3	1 (0.2%)	0 (0%)	1 (1.3%)
Best Actress in Leading Role			
0	316 (70%)	265 (71%)	51 (65%)
1	134 (30%)	108 (29%)	26 (33%)
2	3 (0.7%)	1 (0.3%)	2 (2.5%)
Gloden Globe Drama	61 (13%)	27 (7.2%)	34 (43%)
Gloden Globe Musical/Comedy	35 (7.7%)	25 (6.7%)	10 (13%)

[1]Median (Q1, Q3); n (%)

0 = Did not win Oscars

1 = Won Oscars

Data source Pardoe, I., & Simonton, D. K. (2008)

Method

The model being used in this study is conditional logistic regression model. This model can be considered a specification of a general model and the mathematical formulation is presented as follows.

$$Pr(y_{ij} = 1) = \frac{e^{Z_{ij}\alpha_{ij}}}{\sum_{k=1}^{K_i} e^{Z_{ik}\alpha_{ik}}}$$

The idea of employing this model specification is that nominations of each year will be considered a choice set, and the choice (i.e., the winner in this case) will be made based on their attributes (i.e., other Oscars nominations and winning other non-Oscars) Z_{ij} represents alternative specific attributes and α represents individual specific coefficient that may vary either from decision maker or from alternatives.

Analysis

There are a total of three models. Table 2 illustrates the results from the models. The coefficients have been exponentiated and confident interval are shown.

In the first model, I treat the effects of best actor and actress in leading roles equal, and in the second model, the effects are considered different. Previous honor from Golden Globe is included in the third model and the effects of Golden Globe Drama and Golden Globe Musical/Comedy are considered identical as a movie cannot belong to both of the two categories.

Goodness of model fit is presented in Table 3. As Model 3 performs the best, I will focus mainly on the interpretation of Model 3.

The effects of the total number of nomination and best directing nomination remain significant throughout the three different model specifications while the effects of best actor and best actress in leading roles are not significant. It is however, worth noticing that when treating the effects of best actor and best actress in leading roles separately, the odds of best actress in leading role is lower than 1, which suggests that it might potentially negatively influence the chance of winning best picture for a movie, and this is aligned with previous studies.

In addition to the total number of nominations and best directing nomination, Golden Globe appears to have a positive impact on winning Oscars. The odds 3.19 suggests that if a movie had won Golden Globe, the odds of it winning Oscars best picture will increase 219%.

Table 2. Odds from Conditional Logistic Regression Models

Model Specifications			
	(1)	(2)	(3)
Nomination	1.54***	1.56***	1.49***
	[1.32, 1.80]	[1.33, 1.82]	[1.26, 1.75]
Director	5.59**	5.18**	4.94**
	[1.88, 16.59]	[1.73, 15.51]	[1.60, 15.18]
Actor & Actress	1.12		
	[0.72, 1.72]		
Actor		1.29	1.20
		[0.78, 2.12]	[0.71, 2.05]
Actress		0.87	0.79
		[0.47, 1.62]	[0.41, 1.51]
Golden Globe			3.19***
			[1.69, 6.00]
Num.Obs.	453	453	453
AIC	179.1	179.8	168.1
BIC	191.4	196.3	188.7

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Source: Estimated based on author's calculations.
Reference: Pardoe, I., & Simonton, D. K. (2008)

Table 3. Likelihood Ratio Test on Models

Model Comparisions

	Model 1	Model 2	Model 3
Chisquare	NA	1.26	13.72***
P-value	NA	0.26	0.00

*** $P < 0.01$
The Likelihood ratio test for Model is not applicable as it is the most basic model in this analysis.
Figures based on author's calculation
Reference: Pardoe, I., & Simonton, D. K. (2008)

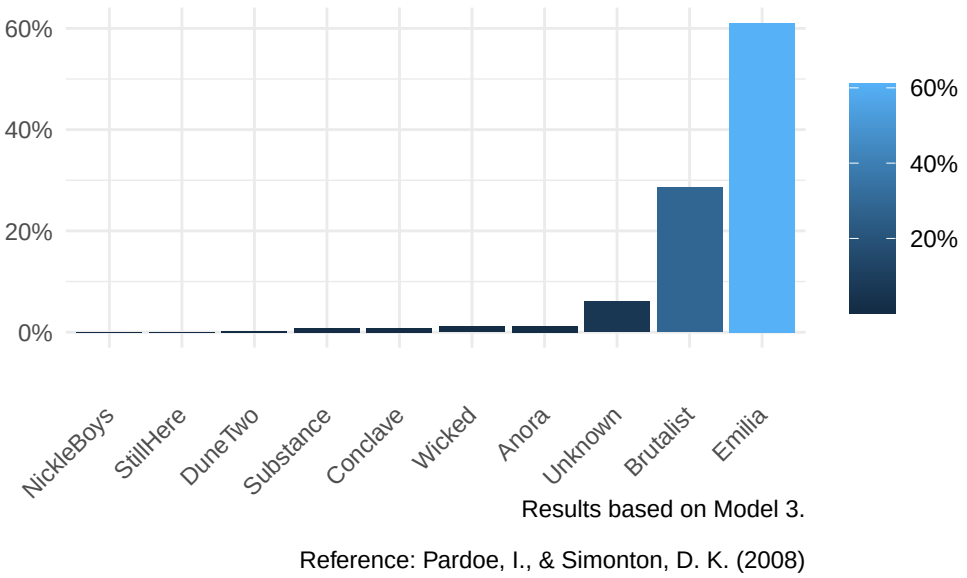


Figure 2. Predicted Probability of Winning Oscars Best Picture in 2025

Results

Results obtained from the models show that the total number of number of Oscars nomination, being nominated for best directing and having won Golden Globe are some of the major factors that will positively affect the odds of a filming winning Oscars best picture.

Furthermore, I tried to make predictions on the result of Oscars in 2025 using Model 3. It is predicted that Emilia Pérez (with a probability of more than 60%) would have won Oscars best picture for this year, while in reality it is Anora that won the prize. In fact, the predicted probability for Anora is rather low (see Figure 2).

Conclusions

Results from the models show that the process of deciding the winner for Oscars best picture might be more complicated and cannot be fully explained by the limited number of variables included in the model. However, the surprisingly high values for the estimates for the total number of nominations, best directing nomination and having won Golden Globe suggest that the process of choosing a winner in the film industry might be more intertwined than expected.

Reference

Pardoe, I., & Simonton, D. K. (2008). Applying discrete choice models to predict Academy Award winners. Journal of the Royal Statistical Society: Series A (Statistics in Society), 171(2), 375–394. https://doi.org/10.1111/j.1467-985X.2007.00518.x