DCM assignment 4

Yiqiu Wang

2023-03-08

Task 1

The Academy Awards, better known as the Oscars, are awards for artistic and technical merit for the global film industry. The awards are regarded by many as the most prestigious, significant awards in the entertainment industry in the United States. The Oscars has a total of 23 awrad catagories in which the Best Picture Award is one of the most prestigious honors in the film industry. The process of selecting the winner of the Best Picture Award involves a two-stage voting process. In the first stage, members of the Academy nominate films that they believe should be considered for the award. In the second stage, all members of the Academy are given the opportunity to vote on the final nominees and select the winner.

In this study, we are interested in which nominations could be the best predictor of the best picture award. Between the director, female lead, and male lead, we believe that the director should be the th best predictor. A director is responsible for various things such as interpretation of the script, communication and collaboration while the actors and actresses perform under the guide of the director. So if a director is given a nomination, the quality of the film is recognized.

Oscar awarding question could be seen as a discrete choice problem. Every year, members of the Academy decide which movies should be given the nomination and then vote for the final winner given all movies of the year. Each movies has different attributes such as the quality of costom and the performances of actors/actresses. Discrete choice problems can be modeled using mathematical and statistical methods, such as utility theory, which assumes that individuals have a utility or value for each alternative(movie in our case) based on the attributes associated with it. This utility can be estimated through surveys or experiments, and used to predict which alternative an individual is most likely to choose given a specific set of attributes, that is, which movie is most likely to win the best picture award.

Apart from the director, female lead and male lead, we also expect the quality of editing(whether the movie is given the editing nomination), screenplay(whether the movie is given the screenplay nomination), the total number of nominations the movie has as well as Producer's Guild of America(PGA) to play a role in predicting which movie is most likely to win. High quality of edition should have correlation with the performance of the director and high quality of performance are given by actors and actresses who follow the instruction of the director. More nominations indicates the movie is recognized and highly praised from various aspects. So the number of nominations should have a positive effect on the chance of the movie winning the award. PGA has long been awarding its honors to the year's most distinguished producing effort.

In Table 1, we could see the descriptive statistics on 7 nominations which we believe would help us predict the winner of the best picture award more accurately,split by whether the picture won the Academy Award or not. 17.47% of all nominated movies won the best picture award.

The average of having the director nomination equals to 0.62. For Non-winner movies, the index is lower than average but for the winner movies. But for winner movies, the average index(0.95) show that it is almost certain that the directors of the winner movies must have the nomination. This to some extent goes inline with our expectation that the director should be the best predictor of the the best picture award. In general, the actors and actresses of the winner movies were more recognized than those of non-winner

Table 1: Descriptive Statistics for Oscar winner Movies of the best picture award(1928-2006)

		Mean	SD	Min	Max	N	Percent
All nominated movies							
	Director Nomination	0.62	0.49	0.00	1.00	458	100.00
	Lead Actor Nomination	0.46	0.55	0.00	3.00	458	100.00
	Lead Actress Nomination	0.31	0.48	0.00	2.00	458	100.00
	Editing Nomination	0.49	0.50	0.00	1.00	458	100.00
	Screenplay Nomination	0.74	0.51	0.00	2.00	458	100.00
	Producer's Guild of America win	0.13	0.34	0.00	1.00	458	100.00
	Number of Nominations	6.42	2.88	1.00	14.00	458	100.00
Award Non-winner Movies							
Choice　0	Director Nomination	0.55	0.50	0.00	1.00	378	82.53
	Lead Actor Nomination	0.41	0.52	0.00	2.00	378	82.53
	Lead Actress Nomination	0.29	0.46	0.00	2.00	378	82.53
	Editing Nomination	0.42	0.49	0.00	1.00	378	82.53
	Screenplay Nomination	0.70	0.52	0.00	2.00	378	82.53
	Producer's Guild of America win	0.04	0.20	0.00	1.00	378	82.53
	Number of Nominations	5.90	2.61	1.00	13.00	378	82.53
Award winner Movies							
1	Director Nomination	0.95	0.22	0.00	1.00	80	17.47
	Lead Actor Nomination	0.71	0.62	0.00	3.00	80	17.47
	Lead Actress Nomination	0.38	0.54	0.00	2.00	80	17.47
	Editing Nomination	0.80	0.40	0.00	1.00	80	17.47
	Screenplay Nomination	0.92	0.35	0.00	2.00	80	17.47
	Producer's Guild of America win	0.55	0.50	0.00	1.00	80	17.47
	Number of Nominations	8.90	2.81	1.00	14.00	80	17.47

Based on Oscar nominees and winners from 1928 to 2006

Choice is a dummy variable which is coded 1 if the movie won the best picture award
and 0 otherwise.

movies(having more nominations). But the average indices of the lead actor and actress nominations for winner movies(0.71 and 0.38) is not so high as that of the director nomination. The indices are almost too low to say this nomination would actually have correlation the award.

The average index of editing nomination(0.42) of non-winner movies is lower than that of all nominated movies(0.49) as well as that of winner movies(0.80). Winner movies' editing was more recognized.

The average indices of screenplay nomination of both non-winner movies(0.70) and winner-movies(0.92) are low considering the maximum number of this nomination could reach 2. Maybe this nomination is not a very good predictor for the award.

About half of winner movies(0.55) also won the Producer's Guild of America while the number of non-winner movies winning the PGA is scarce(only 0.04 of all non-winner movies). So PGA could also be a strong predictor.

As we expect, winner movies had more nominations than non-winner movies.

Table 2: Results of conditional logit models estimating how various nominations affect the likelihood of a movie winning the best picture award
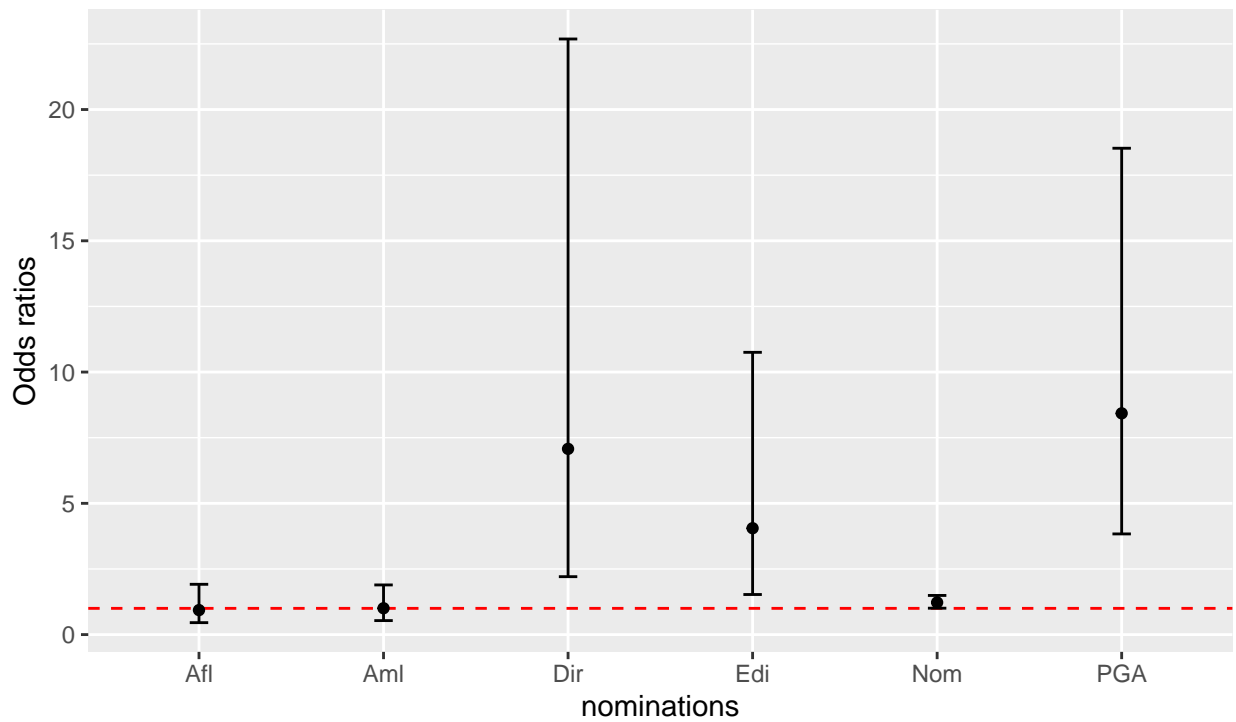
	Model 1	Model 2
Director Nomination	10.812***	7.075***
	(5.733)	(4.206)
Lead Actor Nomination	1.874**	1.004
	(0.431)	(0.325)
Lead Actress Nomination	1.445	0.933
	(0.394)	(0.342)
Editing Nomination		4.052**
		(2.018)
Producer's Guild of America win		8.428***
		(3.387)
Number of Nominations		1.225*
		(0.123)
Num.Obs.	458	458
AIC	223.5	143.9
BIC	235.9	168.7
RMSE	0.35	0.28

Note:

Based on Oscar nominees and winners from 1928 to 2006

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Graph 1: A conditional logit model on how various nominations affect the likelihood of a movie winning the Oscar award(Best Picture)

Based on Oscar nominees and winners from 1928 to 2006

Table 3: Model fit statistics

	Model 1	Model 2
Loglik	-108.751	-65.971
DF(compared to Model 1)	/	3
Chisq(compared to Model 1)	/	85.558
Pr(>Chisq)(compared to Model 1)	/	< 2.2e-16

We ran two conditional logit models here. Model 1 is the basic model which we only included the nominations of the director, lead male and lead female and in Model 2 we also included the nominations of editing, Producer's Guild of America and the number of nominations. The right part of Table 2 and Graph 1 shows the results(odds ratio) of Model 2.

Between director, male lead and female lead, the director nomination is the only variable that has a significant effect($p < 0.001$) on the likelihood of winning the best picture award. After the director is given a nomination, the odds of winning the award are increased by a factor of 7.075. Having an editing nomination significantly($p < 0.01$) increases the odds of winning the award by a factor of 4.052. Winning Producer's Guild of America(PGA) significantly($p < 0.001$) increases the odds of winning the award by a factor of 8.428. The positive effect is even greater than that of the director nomination. Every additional nomination significantly($p < 0.05$) increases the odds of winning the award by a factor of 1.225.

Both AIC and BIC of Model 2 are lower than those of Model 1(the basic model) which suggests a better model fit. This is also true when we ran a likelihood ratio test(Table 3) on Model 1 and Model 2. The log likelihood increases obviously when we move from the basic model(Model 1) to our Model 2. The Chisq is great and the p-value which is less than 2.2e-16(<0.001) is significant when we compare Model 2 to Model 1. So Model 2 should be the preferred model.

Based on the results, winning Producer's Guild of America and having a director nomination are actually very good predictors of whether the movie will win the best picture award. Having more nominations and the editing nomination also make a movie more likely to win though their positive effect are not so strong as those of winning Producer's Guild of America and having a director nomination. It does become clear that the director is the key to winning the best picture award rather than the male lead and female lead.

Task 2

Based on the results of Task 1, we expect our preferred model to be Model 2 which includes Director Nomination, Lead Actor Nomination, Lead Actress Nomination, Editing Nomination, Producer's Guild of America win, Number of Nomination. The Director Nomination and Producer's Guild of America have been proved to be strong predictor of winning the best picture award.

In Table 4, we ran two other models. In Alternative 1, we dropped Producer's Guild of America and Number of Nominations from Model 2. In Alternative 2, we added Screenplay Nomination to Model 2. Model 2 has the lowest AIC and BIC among all 4 models. So it has the best model fit.

In Alternative 1, Lead Actor Nomination has a significant($p < 0.05$) effect on the likelihood of a movie winning the prize. With this nomination, the odds of winning the award are increased by a factor of 1.836. The positive effect of having a editing nomination is stronger than that in Model 2.

After adding Screenplay Nomination to our Model 2, neither the number of nominations nor having the screenplay nomination has a significant effect on the likelihood of a movie winning the award in Alternative 2.

We also ran likelihood ratio tests to compare Model 2 to the two alternatives. The results are shown in Table 5. When we compare Model 2 to Alterantive 1, the log likelihood increases. The Chisq is also great and the p-value($6.119e - 13$,< 0.001) shows Model 2 is indeed a preferred model than Alternative 1. When

Table 4: Results of conditional logit models estimating how various nominations affect the likelihood of a movie winning the best picture award

	Basic Model	Alternative 1	Model 2	Alternative 2
Director Nomination	10.812***	9.743***	7.075***	6.243**
	(5.733)	(5.311)	(4.206)	(3.778)
Lead Actor Nomination	1.874**	1.836*	1.004	0.952
	(0.431)	(0.443)	(0.325)	(0.310)
Lead Actress Nomination	1.445	1.501	0.933	0.978
	(0.394)	(0.440)	(0.342)	(0.361)
Editing Nomination		5.736***	4.052**	4.689**
		(2.170)	(2.018)	(2.447)
Producer's Guild of America win			8.428***	8.550***
			(3.387)	(3.489)
Number of Nominations			1.225*	1.191+
			(0.123)	(0.123)
Screenplay Nomination				1.730
				(0.791)
Num.Obs.	458	458	458	458
AIC	223.5	196.2	143.9	144.5
BIC	235.9	212.7	168.7	173.4
RMSE	0.35	0.34	0.28	0.28

Note:

Based on Oscar nominees and winners from 1928 to 2006

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Table 5: Model fit statistics on three models

	Alternative 1	Preferred Model	Alternative 2
Loglik	-94.094	-65.971	-65.234
DF(compared to former model)	/	2	1
Chisq(compared to former model)	/	56.245	1.4744
Pr(>Chisq)(compared to former model)	/	6.119e-13	0.2247

we compare Alternative 2 to Model 2, the log likelihood hardly changes. The Chisq is small and the p-value which equals to 0.2247 is insignificant. Although Alternative 2 has one more variable than Model 2, it does not have a better model fit. So we should use Model 2(our preferred model) instead of the two alternatives.

Table 6: Model prediction statistics(1928-2006)

Model	Corrrect Prediction Rate
Basic Model	0.2025316
Model 2	0.6898734
Alternative 1	0.4303797
Alternative 2	0.6835443

Table 6 shows the correct prediction rates of the 4 models. Model 2, our preferred model, has the highest correct prediction rate(about 69%) among all models.

In Table 7, we recorded the nominations of the movies competing for the best picture award in Oscar 2023 that would be used to make prediction for the winner movie. Dir is the Director Nomination, Aml

Table 7: Movies competing for the best picture in Oscars 2023

Year	id	Movie	Dir	Aml	Afl	Edi	PGA	Nom
2022	1	All Quiet on the Western Front	0	0	0	0	0	9
2022	2	Avatar: The Way of Water	0	0	0	0	0	4
2022	3	The Banshees of Inisherin	1	1	0	1	0	9
2022	4	Elvis	0	1	0	1	0	8
2022	5	Everything Everywhere All at Once	1	0	1	1	1	11
2022	6	The Fabelmans	1	0	1	0	0	7
2022	7	Tár	1	0	1	1	0	6
2022	8	Top Gun: Maverick	0	0	0	1	0	6
2022	9	Triangle of Sadness	1	0	0	0	0	3
2022	10	Women Talking	0	0	0	0	0	2

is the Lead Actor Nomination, Afl is the Lead Actress Nomination, Edi is the Editing Nomination , PGA stands for winning Producer's Guild of America, Nom is the total number of nominations. Before we start our predictions using our preferred model(Model 2), we can see that Everything Everywhere All at Once has the most number of nominations. It also won the Producer's Guild of America which we know from the results of Model 2 is a very strong predictor. The director is also given a nomination. Then The Banshees of Inisherin, though did not win the Producer's Guild of America, has the second most number of nominations and a nomination for the director. These two movies are very likely to win the best picture award.

Table 8: Prediction of the best-picture winner movie of Oscars 2023

Year	id	Movie Name	Predicted Probability of Winning
2022	1	All Quiet on the Western Front	0.0107
2022	2	Avatar: The Way of Water	0.0022
2022	3	The Banshees of Inisherin	0.1168
2022	4	Elvis	0.0163
2022	5	Everything Everywhere All at Once	0.7314
2022	6	The Fabelmans	0.0290
2022	7	Tár	0.0764
2022	8	Top Gun: Maverick	0.0105
2022	9	Triangle of Sadness	0.0060
2022	10	Women Talking	0.0008

Based on the results in Table 8, it seem that the Movie Everything Everywhere All at Once has the highest probability(73.14%) of winning the best picture award in 2023 among all 10 movies. The Banshees of Inisherin is the second most likely to win the award with a probability equals to 0.1168. But apart from Everything Everywhere All at Once, the probability of other movies winning the award are all very low.

In conclusion, we expect Everything Everywhere All at Once to the winner of the best picture award in 2023.