.

How to Win an Oscar

Mikey Abela, Casey Kim, Leah Lam, Alex Ogilvy

mabela, ckim60, llam3, aogilvy

Hypothesis

In recent years, the film industry and award shows have faced criticism for an apparent lack of representation of people from different backgrounds, as seen with the trending hashtag *#OscarsSoWhite* in 2015. We sought out to investigate the difference regarding how long it took the Academy of Motion Picture Arts and Sciences to award actors and directors holding different identities. Specifically, we wanted to test the following hypothesis: white, straight men from the coasts reach their first Oscar win faster than people of other identities.

Data

Our first data set, found on Kaggle, contains all of the demographic information actors and directors that have won Oscars. Our second data set contains information about all movies ever released, collected using the TMDb API. This data is clean aside from some special characters in the cast and crew lists. Our data is, of course, very positively skewed because it only contains information for individuals who have received Oscars and not for any other actors. However, we scoped our project to account for this, choosing to research the treatment of different groups of people by the Academy of Motion Picture Arts instead of by the entire film industry.

Findings

Claim 1:

The following identities tend to decrease credits to success the most: white (race), straight (sexuality), far west/mideast (birthplace), meaning people with one of these identities were a part of fewer films before winning their first Oscar.

Definition. Credits to Success - the number of credited filmwork until an individual's first Oscar win.

We performed multiple linear regression using our categorical, demographic data (region of birth, race, and sexuality) as independent variables and credits to success as the dependent variable.

	coef	P>|t|
const	6.1992	0.709
race_"Asian"	−0.7353	0.963
race_"Black"	7.8785	0.419
race_"Hispanic"	5.0712	0.635
race_"Middle Eastern"	−4.6951	0.858
race_"Multiracial"	−8.8132	0.726
race_"White"	7.4930	0.315
region_far west	23.9219	0.247
region_great lakes	10.1781	0.622
region_international	3.8062	0.851
region_mideast	3.9619	0.846
region_new england	−0.5257	0.980
region_plains	3.5247	0.866
region_rocky mountain	13.5000	0.635
region_southeast	9.2750	0.657
region_southwest	17.0658	0.437
sexuality_"Bisexual"	21.8126	0.006
sexuality_"Gay"	−5.0777	0.634
sexuality_"Lesbian"	−14.6141	0.421
sexuality_"Matter of Dispute"	5.6965	0.705
sexuality_"Na"	−3.4260	0.700
sexuality_"Straight"	1.8078	0.739

(a) Multiple Linear Regression on the Credits to Success, R-Squared: 0.08

	coef	P>|t|
const	0.3054	0.060
race_asian	0.0747	0.625
race_black	0.1054	0.267
race_hispanic	−0.0032	0.976
race_middle eastern	0.5280	0.039
race_multiracial	−0.4267	0.082
race_white	0.0273	0.707
region_far west	0.0428	0.832
region_great lakes	0.0269	0.894
region_international	0.0466	0.814
region_mideast	−0.0489	0.806
region_new england	0.0049	0.981
region_plains	0.0223	0.912
region_rocky mountain	−0.0031	0.991
region_southeast	−0.0243	0.905
region_southwest	−0.0380	0.859
sexuality_bisexual	0.0154	0.842
sexuality_gay	−0.0184	0.860
sexuality_lesbian	0.0384	0.828
sexuality_matter of dispute	0.0525	0.721
sexuality_na	0.1238	0.153
sexuality_straight	0.0937	0.077

(b) Multiple Linear Regression on the Success Ratio, R-Squared: 0.044

Seeing that "White," "Straight,", "Far West" and "Mideast" had highly positive coefficients seemed strange to us, as this suggested that these identities tended to increase the credits to success metric, whereas we hypothesized that they would decrease it. We realized that White, Straight Men from the coasts had a higher number of credits overall, with 67.37 average total credits as compared to 52.44 average total credits for others. Consequently, we normalized our credits to success metric and calculated a Credit Success Ratio:

$$\text{Credit Success Ratio} = \frac{\text{film credits to first Oscar win}}{\text{total film credits in career}}$$

Now, our dependent variable represents the percentage/portion of an individual's career that was spent without an Oscar. We performed Multiple Linear Regression again with the same independent variables and this new dependent variable

These results still did not support our hypothesis, showing for example that the "Gay", "Mideast" and "Multiracial" identities tended to decrease the ratio most. Regardless, the low R Squared values and high p values for both regressions suggested to us that these predicted models did not fit our data very well and thus were not strong means with which to support or reject our hypothesis.

Claim 2:

Straight, white men born on the coasts have a statistically significant, lower mean actor success ratio than others.

We performed t tests on the ratios of straight, white men from the coasts to all others and on each privileged identity independently compared to all non-privileged identities.

	Mean Credit Success Ratio
White/male/straight/coasts vs. Other	43.31% vs. 44.56%
Race: White vs. Non-white	43.95% vs. 48.76%
Gender: Male vs. Female	46.96% vs. 40.61%
Sexual Orientation: Straight vs. Non-straight	44.36% vs. 44.36%
Region: Coasts vs. Non-coasts	42.08% vs. 45.43%

	t statistic	p value	Significant?
White/male/straight vs. other	-0.35329	0.72461	No
Race: White vs. Non-white	-1.07474	0.28796	No
Gender: Male vs. Female	2.43732	0.01522	Yes
Sexual Orientation: Straight vs. Non-straight	-0.00010	0.99992	No
Region: Coasts vs. Non-coasts	-1.18903	0.23545	No

We found that the only statistically significant difference in means was for Male vs Female, and it was actually a difference that refuted our claim, as the females had a lower mean than the males.