HOW TO WIN AN OSCAR

the_SQL_was_worse
Alex Ogilvy, Casey Kim, Leah Lam, Mikey Abela

INTRODUCTION & 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 tested the hypothesis that white, straight men born on the American coasts reach their first Oscar win faster than people of other identities.**

DATA

Data Set #1: Kaggle Oscar Winner Demographics
- Contains race, sexual orientation, religion, place of birth for winners of the Academy Awards in the acting and directing categories.

Data set #2: TMDb All Movies Database
- Contains titles, popularity ratings, release dates, revenue, critical vote average, critical vote count, and cast/crew credit lists for all movies ever released.

Our data is positively skewed because it only contains information for Oscar winners and not for any other actors or directors. 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.

METHODOLOGY

We tested two major claims. The initial dependent variable that we focused on in both claims was **"credits to success,"** meaning the number of film credits that individuals accumulated before their first Oscar win. After determining that white men had on average **more film credits overall than other actors (67.34 vs. 52.44)**, we modified our dependent variable. We normalized our credits to success variable by dividing each individual's credits to success by their total number of film credits (providing the proportion of the person's film career spent without an Oscar), creating our **"credit success ratio"** variable.

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.
We bucketed birthplaces into regions (Far West, Rocky Mountain, Southwest, Plains, Great Lakes, Southeast, Mideast, New England and International), and then converted all categorical variables (race, sexuality, birthplace) into dummy variables.

We performed multiple linear regression using our dummy independent variables (race, sexuality, birthplace) with our "credits to success" and then our new "actor success ratio" dependent variables.

Claim 2:
Straight, white men born on the coasts have a statistically significant, lower mean success ratio than others.
In order to test this, 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.

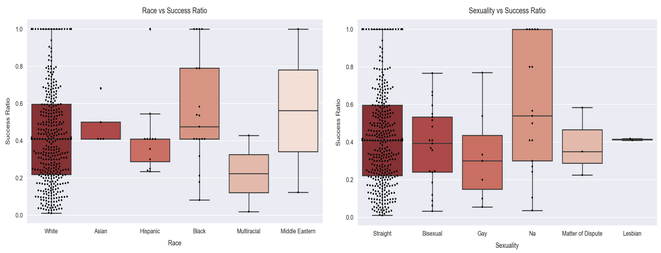


Figure 1. Box plots showing success ratio for different races and sexualities. The high volume of data points in the "white" and "straight" column is indicative of the skew in our data set

RESULTS/ANALYSIS

Claim 1:

Multiple Linear Regression Results on Credits to Success:

	Race: Asian	Race: Black	Race: Hispanic	Race: Middle Eastern	Race: Multiracial	Race: White	Region: Far West	Region: Great Lakes	Region: International	Region: Mideast	Region: New England	Region: Plains	Region: Rocky Mountain	Region: Southeast	Region: Southwest	Sexuality: Bisexual	Sexuality: Gay	Sexuality: Lesbian	Sexuality: Disputed	Sexuality: N/A	Sexuality: Straight
coef	-0.73	7.88	5.07	-4.70	-8.81	7.49	23.92	10.18	3.81	3.96	-0.53	3.52	13.50	9.28	17.07	21.81	-5.08	-14.61	5.69	-3.43	1.81
p-value	0.96	0.42	0.64	0.86	0.73	0.32	0.25	0.62	0.85	0.85	0.98	0.87	0.64	0.66	0.44	0.01	0.63	0.42	0.71	0.70	0.74

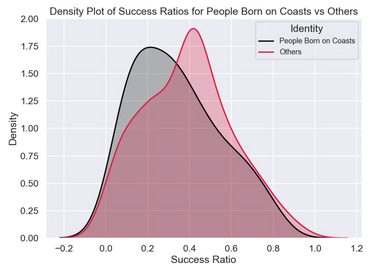
R-Squared: 0.08
Coefficients suggest that being white, straight, and from the Far West/Mideast actually tend to increase credits to success whereas being a lesbian (for example), tends to decrease credits to success. However, with the R-Squared value being so low and few p-values being below 0.05, we did not find any significant independent predictors of a change in credits to success.

Multiple Linear Regression Results on Success Ratio:

	Race: Asian	Race: Black	Race: Hispanic	Race: Middle Eastern	Race: Multiracial	Race: White	Region: Far West	Region: Great Lakes	Region: International	Region: Mideast	Region: New England	Region: Plains	Region: Rocky Mountain	Region: Southeast	Region: Southwest	Sexuality: Bisexual	Sexuality: Gay	Sexuality: Lesbian	Sexuality: Disputed	Sexuality: N/A	Sexuality: Straight
coef	0.075	0.105	-0.003	0.528	-0.427	0.027	0.042	0.027	0.047	-0.049	0.005	0.022	-0.003	-0.024	-0.038	0.015	-0.018	0.038	0.053	0.124	0.094
p-value	0.625	0.267	0.976	0.039	0.082	0.707	0.832	0.894	0.814	0.806	0.981	0.912	0.991	0.905	0.859	0.842	0.860	0.828	0.721	0.153	0.077

R-Squared: 0.044
Coefficients still suggest that being white, straight and from the Far West tends to increase the success ratio (but being from the Mideast tends to decrease it). Being multiracial is the single identity that tends to decrease the ratio. This time, the R-Squared value was even lower, and p-values remained high. So again, we did not find any significant independent predictors of a change in credits to success.
In all, 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:



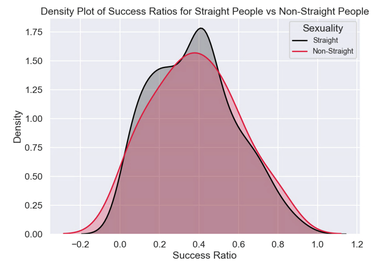
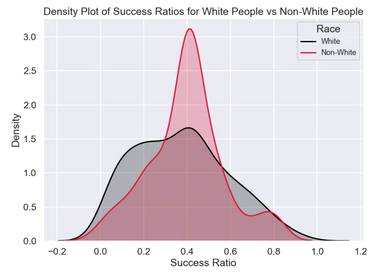
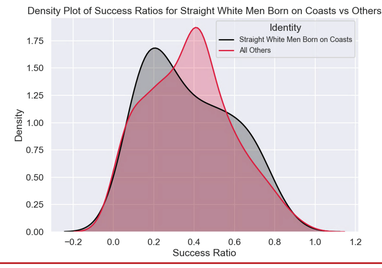


Figure 2. Density plots of success ratio for privileged versus non-privileged identities





RESULTS CONTINUED

t-Test Results:

	t-statistic	p value	Statistically 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
Straight vs Non-Straight	-0.00010	0.99992	No
Coastal vs Non-Coastal	-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

Mean Success Percentages:

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%

CHALLENGES & FUTURE WORK

Data Collection and Cleaning Issues: Our data set includes a confidence value for each demographic, indicating the level of confidence in the designation of a specific identity. 92% of these values are 1 (indicating 100% confidence). However, it is possible that the 8% of reports on identities could have impacted our findings. Names with special characters such as accented letters were discarded, but these represented <1% of our initial data set.
Further Research: Our multiple linear regression models suggest that there may be fewer overall roles available to people of color to begin with. Further research would need to be conducted on the amount of roles that are written specifically for white, straight males vs. others. Furthermore, comparing these same metrics between different award shows may reveal consistent biases within the entire film industry and provide more sufficient data to explore our hypothesis.

CONCLUSION

Claim 1: We did not find significant evidence to support our claim that identifying as Straight and White and being born on the American coasts tended to decrease the credits to success and success ratio variables. In both of our performed multiple linear regressions, the R-Squared value was quite low and while some interesting relationships emerged (ie. Multiracial tending to decrease the success ratio the most of any single identity), few p-values were below our alpha threshold of 0.05, and the predicted models did not fit our data very well.
Claim 2: We did not find sufficient support to reject the null hypothesis that Straight White men born on the American coasts have the same mean success ratio as others. The only statistically significant result of our t test allowed us to reject the null hypothesis that Men and Women have the same mean success ratio, but it supported that women had a lower mean (not our alternative hypothesis that men did).
Summary: We cannot accept our hypothesis that straight white men born on the American coasts reach their first Oscar after a fewer number of film credits than others.

SIGNIFICANCE

While our analysis does not necessarily support #OscarsSoWhite specifically in terms of credits to success and credit success ratio, it certainly does not reject or combat the movement in any way. The mere fact that this data is as skewed toward the straight and white identities as it is indicates that the Academy Awards have historically favored a certain model of "success" in the film industry. Although we did not find statistical support for the claim that straight white men from the coasts reach this recognition faster than individuals who hold other identities, we hope that the points brought up by this study encourage further research and that welcome anomalies (such as Parasite's historic win for Best Picture this year - the first foreign language film to win) continue to arise at the Oscars.