Gender Distribution in Movie Roles

Vijay Ravuri and Kelly Bodwin
California Polytechnic University Statistics Department

Background

How has the gender distribution in movie roles changed over the past 50 years? Is there any relationship between the gender distribution of a movie and how successful a movie is? To answer these questions I gathered data on the top 50 grossing movies in every year from 1969-2018 from IMDB along with cast information for each movie.

Data

IMDB provides a wide range of information on movies including genre, ratings, and cast lists.

Cast information was used to calculate two measures of gender distribution:

1. **G-SCORE:** The percentage of credited roles played by a man, weighted by how highly-billed that role was
2. Percentage of uncredited roles played by a man, unweighted.

Title	Year	% Cr Male	G Score
The Breakfast Club	1985	61.54	74.73
The Blind Side	2009	65.08	53.52
300	2006	65.06	79.03
Mad Max: Fury Road	2015	72.41	60.26

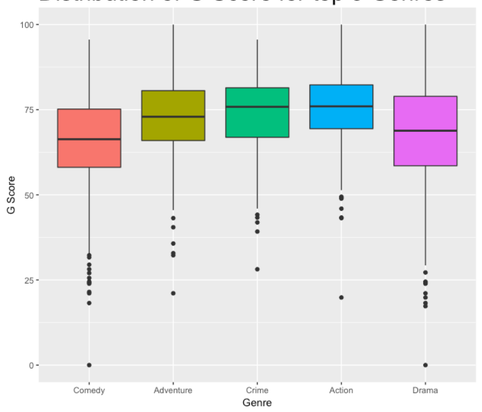
Results

Genre and G-Score

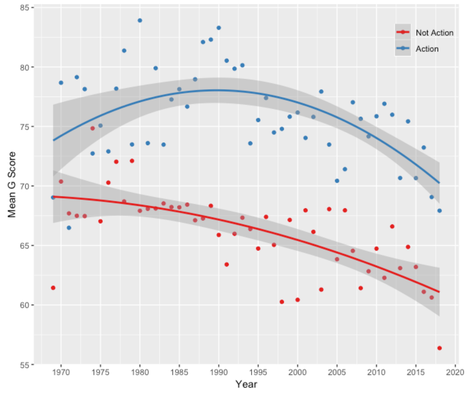
Genre to genre differences are the largest differences in gender distribution by a large margin.

Some genres shows a clear interaction with time. In particular, the **Action** genre is rapidly becoming more gender balanced.

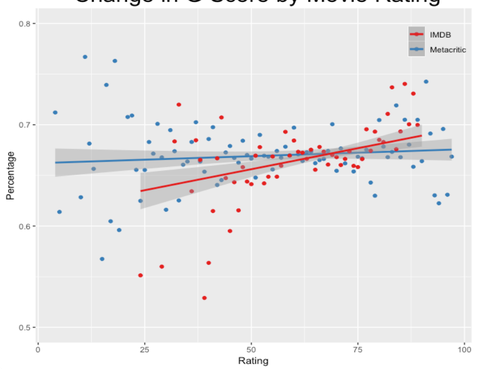
Distribution of G Score for top 5 Genres



Mean G Score over Time



Change in G Score by Movie Rating



Ratings and G-Score

IMDB ratings seem to trend higher for movies with **less female roles by percentage**.

This trend does not hold for **Metacritic ratings.**

LASSO Model

Models were built for the following response variables using *Multivariate LASSO Regression,* similar response variables were paired.

Response	R^2 %	Notable Predictors
G Score	33.32%	Year, Action, Sport, Director Gender
% Uncredited Male	09.23%	Year, War, Romance, Director Gender
Votes	62.69%	G Score, Fantasy, Comedy, MPAA Rating
Gross	12.16%	G Score, Horror, Fantasy, Adventure
IMDB Score	21.42%	G Score, Comedy, Horror, Biography
Metacritic Score	14.56%	Animation, Drama, Horror, Director Gender

(Note: Director Gender was coded with Female as the baseline)

To highlight some key findings, **Credited roles** (**G Score**) were much more explainable than **uncredited** roles, though **year was a large factor in both**. **Popularity on IMDB** was surprisingly explainable while **gross** was not easily explained. For the measures of quality (IMDB and Metacritic Scores), **IMDB** was more explainable and, more importantly, was significantly associated with G Score, **Metacritic scores** were less explainable and did not show the same significant association.

Acknowledgments

I would like to thank Dr. Kelly Bodwin for her assistance, insight, and advising on this research project. Additionally, I would like to thank the Cal Poly Statistics Department for funding this presentation.