Predicting Bechdel Test Results through Statistical Modeling

Hailey Quintavalle

Table of contents

Abstract

The Bechdel Test is a simple measurement designed to analyze the representation of women in film. To pass, a movie must have two female characters who have a conversation that is not about a man. The present research aims to use genre, release year, movie budget, user ratings, and critics rating scores to predict the probability a movie will pass the test. Logistic regression analysis reveals more recent movies are predicted to have a higher probability of passing the test. For most years, genres such as Romance and Comedy are predicted to have a higher probability of passing, while genres such as Action, Sport, War, and Western are predicted to have a lower probability of passing the test. Although the Bechdel Test has its flaws, the test is a useful metric to bring attention to the roles women hold in film.

Introduction

If you think of your all-time favorite movie, try to think about how many of the main characters were women. In 2023, 35% of speaking roles in movies belonged to women, and only 28% of the top grossing films contained female protagonists (Lauzen, 2024). Just one decade before, in 2013, women occupied 30% of all speaking roles and just 15% of protagonists were women (Lauzen, 2014). This is a small change for a span of 10 years, showcasing the underrepresentation of women in film, and the slow steps towards progress the film industry seems to be taking. As we examine the portrayal of women's roles in film, it is useful to explore various criteria used to measure their representation. One notable measure is the Bechdel Test, unintentionally introduced by American cartoonist Alison Bechdel in 1986, and the primary measure of women in film that is the focus of the present analysis.

The Bechdel Test originated from a comic strip titled "The Rule", a part of her comic *Dykes To Watch Out For*. The comic depicts two women discussing certain criteria necessary for them to watch a movie. The criteria for a movie included: two named women, who talked to each other, and held a conversation with each other that was not about a man. In the early 2000s, the test quickly gained popularity online and has since been used as a common tool for analyzing the role of women in film.

The test has only three basic requirements, meaning that a film can pass with just one line. Because of the simplicity of this tool, many have come forward with fair criticisms of the test. For example, the test does not take into account demographic factors, such as age, the voices of women of color, or those who do not speak English as their first language (O'Meara, 2016). Other flaws of the test include the oversight of conversations that are not directly about a man, but may be indirectly about them, where the conclusion of the test are unclear.

Regardless of its flaws, the Bechdel Test is a straightforward and easy tool that allows people to quickly make base assumptions about the presence of women in film. For this reason, this analysis will be examining the characteristics of thousands of movies in an attempt to use logistic regression analysis to model the likelihood of passing the Bechdel Test.

Read the comic here.

Data

The data used for the present analysis comes from a combination of multiple online sources. A data set containing 10,183 movie titles was available through IMDb Non-Commercial Datasets. Release dates for these movies range from 1874 to 2023. Variables included in this data set can be seen in the preview of the data below.

IMDb Non-Commercial Data set:

tconst	originalTitle	startYear	runtimeMinutes	genres
tt27502426	Les filles d'Olfa	2023	107	Documentary
tt15398776	Oppenheimer	2023	180	Biography,Drama,History
tt15326988	Ghosted	2023	116	Action,Adventure,Comedy
tt8400584	The Perfect Find	2023	99	Comedy,Drama,Romance
tt14230388	Asteroid City	2023	105	Comedy,Drama,Romance
tt18257464	Polite Society	2023	104	Action,Comedy

The code manual for the data from IMDb:

- **tconst**: alphanumeric unique identifier of the title

- **originalTitle**: title of the movie

- **startYear**: release year of the movie

- **runtimeMinutes**: runtime in minutes

- **genres**: up to three genres associated with the title

Other data that was used in this analysis comes from the Bechdel Test Movie List, where users can submit movies with their Bechdel test rating through their online platform. The data set pulled from this website contains 10,251 movies with release dates ranging from 1874 to 2024. For the rating variable, a movie is given a rating from one to three, directly corresponding with the number of requirements of the Bechdel test that it passes. A preview of this data can be shown below.

id	title	imdbid	year	rating
11166	Hunger Games: The Ballad of Songbirds & Snakes, The	10545296	2023	3
11169	Marvels, The	10676048	2023	3
11171	Royal Hotel, The	18363072	2023	3
11172	Nowhere	15789472	2023	3
11173	Leo	15654328	2023	3
11230	Book of Clarence , The	22866358	2023	1

The code manual for the Bechdel Test Movie List data:

- **id**: ID number

- **title**: title of the movie

- **imdbid**: IMDb number ID

- **year**: release year

- **rating**: Bechdel Test rating (0-3)

The final data set that was used in this analysis comes from the [TidyTuesday](#) social data project through GitHub. This data set contains 1,794 movies released from 1970 up to 2013.

year	imdb	title	test	clean_test	binary	budget
2013	tt0770828	Man of Steel	ok-disagree	ok	PASS	2.25e+08
2013	tt1821549	Nebraska	ok-disagree	ok	PASS	1.20e+07
2013	tt1670345	Now You See Me	notalk-disagree	notalk	FAIL	7.50e+07

domgross	intgross	code	budget_2013	domgross_2013	intgross_2013
291045518	687999518	2013PASS	2.25e+08	291045518	687999518
17482517	17482517	2013PASS	1.20e+07	17482517	17482517
117723989	351723989	2013FAIL	7.50e+07	117723989	351723989

period_code	decade_code	imdb_id	response	rated	language	runtime
1	1	0770828	TRUE	PG-13	English	143 min
1	1	1821549	TRUE	R	English, Spanish	115 min
1	1	1670345	TRUE	PG-13	English	115 min

plot
A young itinerant worker is forced to confront his secret extrastellar origin when Earth is invaded by members of his own race.
An aging, booze-addled father makes the trip from Montana to Nebraska with his estranged son in order to claim a million-dollar Mega Sweepstakes Marketing prize.
An FBI agent and an Interpol detective track a team of illusionists who pull off bank heists during their performances and reward their audiences with the money.

writer
David S. Goyer (screenplay), David S. Goyer (story), Christopher Nolan (story), Jerry Siegel (Superman created by), Joe Shuster (Superman created by)
Bob Nelson
Ed Solomon (screenplay), Boaz Yakin (screenplay), Edward Ricourt (screenplay), Boaz Yakin (story), Edward Ricourt (story)

country	metascore	imdb_rating	director	released	imdb_votes
USA, Canada, UK	55	7.4	Zack Snyder	14 Jun 2013	359556
USA	86	7.9	Alexander Payne	24 Jan 2014	33503
France, USA	50	7.3	Louis Leterrier	31 May 2013	280199

actors	genre	awards
Henry Cavill, Amy Adams, Michael Shannon, Diane Lane	Action, Adventure, Fantasy	4 wins & 16 nominations.
Bruce Dern, Will Forte, June Squibb, Bob Odenkirk	Adventure, Drama	Nominated for 6 Oscars. Another 26 wins & 61 nominations.
Jesse Eisenberg, Mark Ruffalo, Woody Harrelson, Isla Fisher	Crime, Mystery, Thriller	1 win & 3 nominations.

The code manual for the TidyTuesday data:

- **year**: release year

- **imdb**: IMDb id

- **title**: title of movie

- **test**: Bechdel Test outcome

- **clean_test**: Bechdel Test outcome (cleaned)

- **binary**: binary Pass or Fail of the Bechdel Test

- **budget**: budget as of release year

- **domgross**: domestic gross in release year

- **intgross**: international gross in release year

- **code**: code for movie ###

- **budget_2013**: budget normalized to 2013

- **domgross_2013**: domestic gross normalized to 2013

- **intgross_2013**: international gross normalized to 2013

- **period_code**: period code ###

- **decade_code**: decade code
- **imdb_id**: IMDb ID ###
- **plot**: plot summary of the movie
- **rated**: content rating of movie
- **response**: ??? ###
- **language**: language of movie
- **country**: country movie is produced in
- **writer**: writer of the film
- **metascore**: Metascore (critic) rating (0-100)
- **imdb_rating**: IMDb (user) rating (0-10)
- **director**: Director(s) of movie
- **released**: released date
- **actors**: main actors in movie
- **genre**: genre
- **awards**: awards won
- **runtime**: runtime in minutes
- **type**: type of film ###
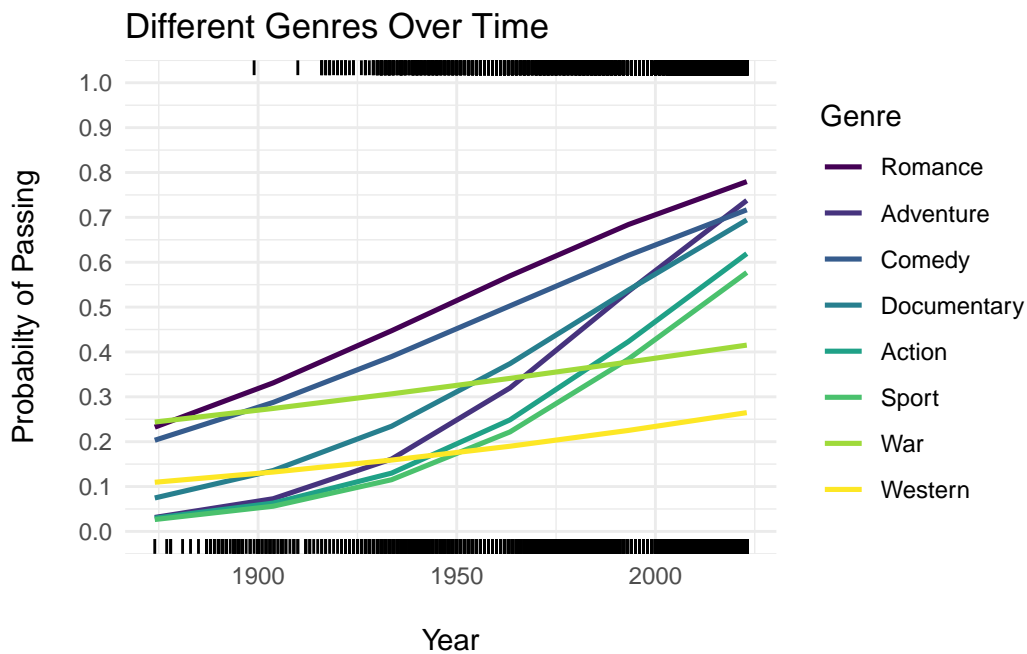- **imdb_votes**: number of IMDb votes

Exploratory Analysis

Modeling

$$logit(\pi) = \beta_0 + \beta_1 \cdot \text{Budget} + \beta_2 \cdot \text{Year} + \beta_3 \cdot \text{IMDb Rating}$$

Full model with all genres and all interactions:

[insert model] ??why wont latex work here??

```
list3 |> filter(values == 1, genre %in%
                c("Sport", "Comedy", "Romance", "Adventure", "Action", "Documentary", "War"
  mutate(genre = fct_reorder2(genre, .x = year, .y = .pi)) |>
  ggplot(aes(x = year, y = .pi)) +
  geom_line(aes(color = genre), linewidth = 0.9) +
  geom_rug(data = movies_pass, aes(x = year, y = as.numeric(binary)), sides = "t") +   ## alp
  geom_rug(data = movies_fail, aes(x = year, y = as.numeric(binary)), sides = "b") +
  scale_y_continuous(breaks = seq(0, 1, by = 0.1), limits = c(0, 1)) +
  theme_minimal(base_size = 11) +
  labs(x = "\nYear",
       y = "Probabilty of Passing\n",
       color = "Genre",
       title = "Different Genres Over Time") +
  scale_color_viridis_d()
```


Conclusion

References

1. Alison Bechdel, 1986. *Dykes to watch out for*. Firebrand Books.

2. Lauzen, M. (2024). It's a Man's (Celluloid) World: Portrayals of female characters in the top grossing U.S. Films of 2023. *Center for the Study of Women in Television & Film.* https://womenintvfilm.sdsu.edu/its-a-mans-celluloid-world-portrayals-of-female-characters-in-the-top-grossing-u-s-films-of-2023/

3. Lauzen, M. (2014). It's a Man's (Celluloid) World: On-Screen Representations of Female Characters in the Top 100 Films of 2013. *Center for the Study of Women in Television & Film.* https://womenintvfilm.sdsu.edu/files/2013_It's_a_Man's_World_Report.pdf