Bayesian Analysis of the Proportion of Movies That Pass the Bechdel Test

Sam Edds, Olivia Hackworth, Katherine Wilkinson
University of Michigan
Department of Statistics

April 29, 2019

Abstract

The authors use hierarchical modeling and hierarchical regression to show that both genre and time influence whether a movie passes the Bechdel test. For any genre, newer movies have higher odds of passing the Bechdel test than older movies. Additionally, time matters more for some genres than others. All results have the caveat that the data are crowd-sourced, with limitations discussed.

Contributions

All authors contributed to the data cleaning and exploration. Katherine modeled the proportion of movies that pass the Bechdel test by both decade and genre. Olivia and Sam worked together to run the regression model and interpret the results. All authors wrote the report collaboratively.

1 INTRODUCTION

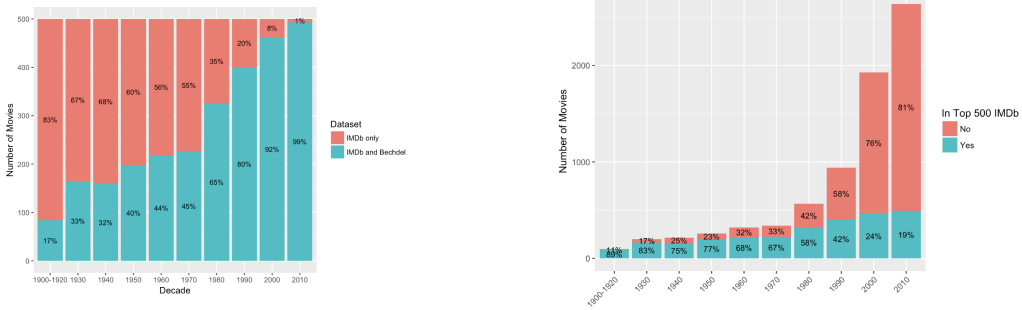
The representation of women in fiction, particularly movies, has become a topic of interest in modern day discourse. Our main goal in this analysis is to determine how this representation has changed over time and how it differs by types of movies (genres). We use the Bechdel test as a measure of representation. The Bechdel test was created in 1985 by cartoonist Alison Bechdel and has three criteria that a movie must meet in order to pass:

1. The movie has at least 2 named women in it,
2. who talk to each other,
3. about something other than a man.

Our exploratory analysis indicates that the proportion of movies that pass the Bechdel test may have a linear trend, and that genre may have an influence. We first model the proportion of movies by decade and then by genre that pass the Bechdel test to see if proportions differ by time and movie type. Furthermore, we fit hierarchical regression model to determine whether there is a trend in passing the Bechdel test due to time for each genre.

2 DATA

The primary data we used is crowd-sourced and obtained at bechdeltest.com. To gather more variables such as Year and Genre, we merged the Bechdel data set with data pulled from IMDb.com based on the IMDb identification. We only examine movies that have at least one genre attached, also removing duplicate entries, and movies with other missing variables. We also choose to ignore information about the genre "news" because it includes just four observations which leads to instability. Altogether, these data cleaning choices remove fewer than twenty observations. Our final working data set has 7,497 unique movies.



(a) % of top 500 IMDb movies in working dataset

(b) % of working dataset caputred in top 500 IMDb movies

Figure 1: How much of our dataset is captured in top 500 IMDb Movies by Decade

Our data set is crowd-sourced and therefore not a random sample, so we compare our results to mainstream movies in order to determine if our data set captures most mainstream movies. We compare against the top 500 movies in each decade which are determined as those movies on IMDb that have the top 500 number of user rating votes. This method captures not just the top rated movies, but also movies that are seen more frequently by the general public, regardless of how well they are subsequently rated. From Figure 1(a), we can see that as decade increases, our data set captures more of the top 500 movies. Additionally, Figure 1(b), shows that the data set includes more than just the top 500 movies, as the portions shaded in pink

represent non-top 500 movies. This exploration allows us to move forward feeling more confident that despite the non-random generation of the data set, it includes a diverse selection of movies including many popular ones.

At first glance, the data set contains more movies that pass the Bechdel test than that fail. Figure 2 illustrates the difference while also showing the reasons that movies fail. The portion shaded in pink represents the movies that failed because they did not contain two or more named women, the green portion represents movies where the women did not talk to each other, and the blue portion represents movies where the women did not discuss a subject other than a man.

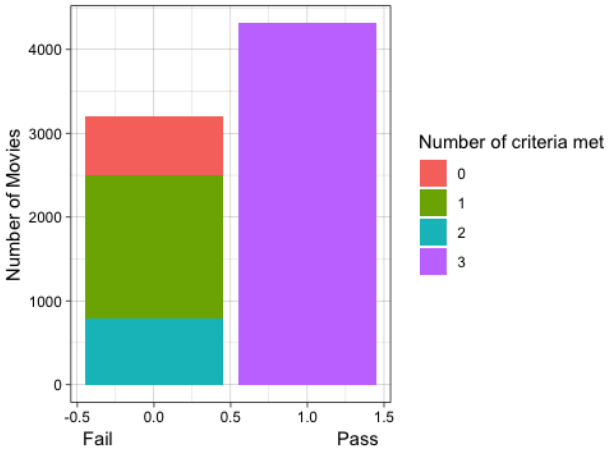


Figure 2: Proportion of movies that passed the Bechdel test

Figure 3 shows how the data set breaks dow in terms of genre. Drama and comedy are the largest genres, while film-noir, westerns and documentaries are the smallest. Figure 4 illustrates the proportion of movies that pass the Bechdel test by genre. This plot is broken into three size chunks where the purple plot includes genres with 0 to 499 movies, the blue plot includes genres with 500 to 999 movies and the green plot includes genres with 1000 or more movies. Overall, this plot highlights the fact that there seems to be some variability across genre for the proportion of movies that pass the Bechdel test.

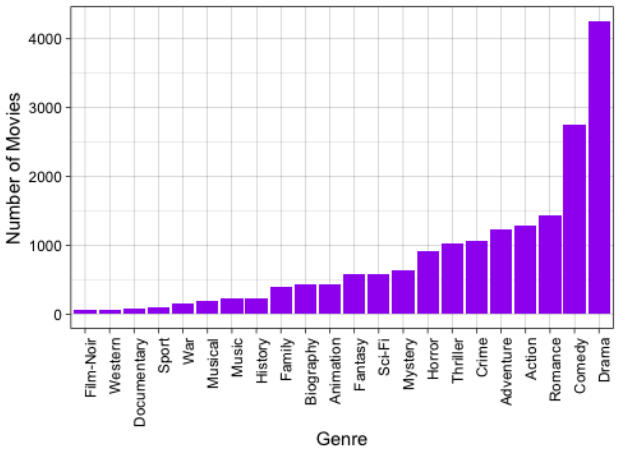


Figure 3: Counts of movies by genre

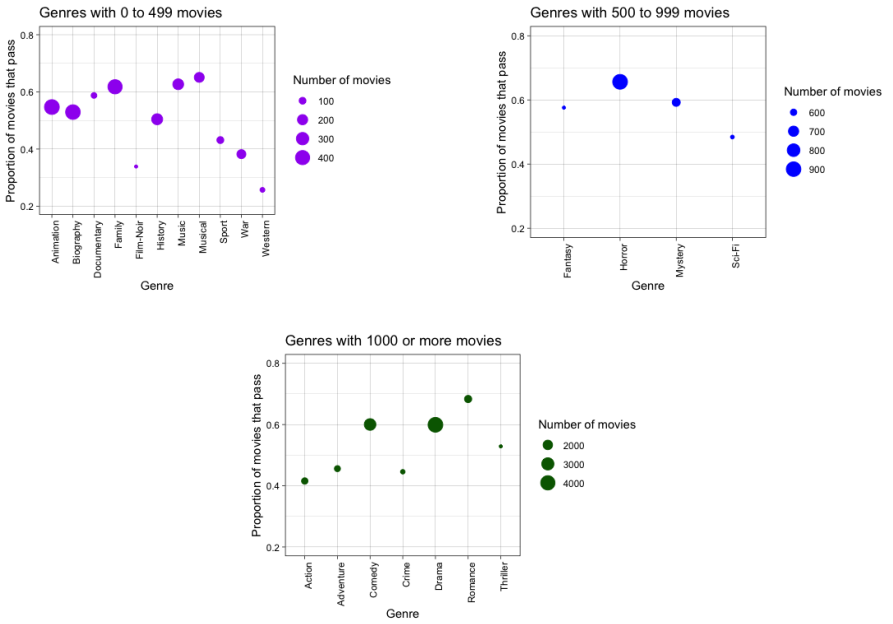
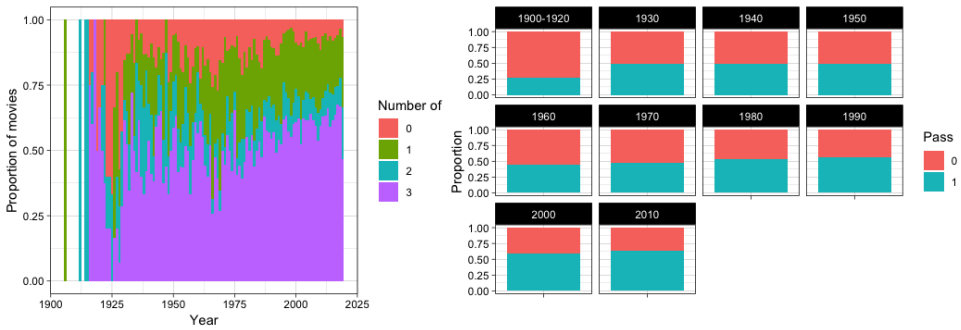


Figure 4: Proportion of movies that passed the Bechdel test by genre

We also want to explore the influence of time on the proportion of movies that pass the Bechdel test. Figure 5(a) shows the proportion of movies that pass broken down by year. In the early years, there are fewer observations, to the proportion that pass (in purple) is more variable. However, as time goes on and the number of observation increases, there appears to be an upward trend in proportion of movies that pass. This trend also appears in Figure 5(b) which shows the proportion of movies that pass broken down by decade. There is a potential slow upward trend visible as we move through time. This motivates an exploration into the effect of time on movies passing the Bechdel test.



(a) By Year (b) By Decade

Figure 5: Proportion of movies that passed the Bechdel Test

3 HIERARCHICAL MODELS

3.1 *Assessing the Proportion of Movies that Pass By Decade*

For our analysis, a movie must meet all three criteria of the Bechdel test as defined above in order to be considered a pass. In order to determine how women have been represented in movies over time, we look at the proportion of movies that pass in each decade. Since our data includes movies from the early 20th century, we group the first two decades in order to account for the sparsity of movies during this time period. All other years are grouped by decade for a total of ten decades that we will analyze.

From our raw proportions (Figure 5), there does seem to be a general trend towards more representation as year and decade increase. To estimate the underlying proportion of movies that do pass in each decade, we run a Bayesian analysis using a hierarchical model with hyperparameters. The necessary distributions are as follows:

- **Hyperpriors** - $P(\alpha, \beta) \sim (\alpha + \beta)^{-\frac{5}{2}}$
- **Prior** - $P(\theta_i|\alpha, \beta) \sim Beta(\alpha, \beta)$
- **Likelihood** - $P(y_i|\theta_i, n_i) \sim Bin(n_i, \theta_i)$

Where y_i is the number of movies in decade i that pass the Bechdel test and θ_i is the correspdoning underlying proportion of movies that pass the Bechdel test in decade i. Thus we have a joint posterior density of:

$$P(\theta_i, \alpha, \beta|y_i, n_i) \propto \prod_i^n \theta_i^{y_i}(1-\theta_i)^{n_i - y_i}(\alpha + \beta)^{-\frac{5}{2}}$$

We see from our marginal density of the hyperparameters (Figure 6), that the most likely value of our alpha and beta for our model by decade is roughly 9.5 for both. Using this marginal density and sample of alpha and beta, we then sample for our underlying proportions in each decade (θ_i) from the prior density specified above.

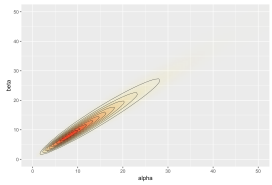


Figure 6: Marginal Distribution of α and β

For this model, we implement the Stan package in R and draw 2000 samples for each θ_i and obtain the 95% confidence intervals seen in Figure 7. We can see that the representation of women in movies is increasing over time, on average. After the 1970's over half of movies produced pass the Bechdel test. Furthermore, the estimates obtained for the underlying proportion of movies that pass is higher than the raw proportion in the early decades. These proportions converge in later decades as more movies are produced.

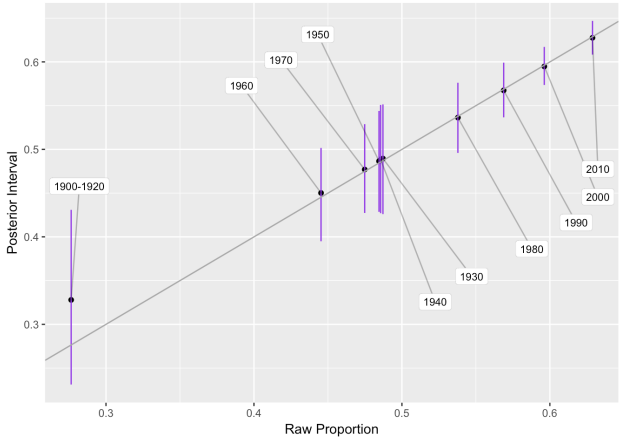


Figure 7: Estimates vs Raw proportions by Decade

3.2 *Assessing the Proportion of Movies that Pass By Genre*

A similar analysis is run to determine how women are represented across different movie genres. There are a total of 22 different genres. It should be noted that many movies are listed with multiple genres, however none of the genres are strongly correlated, with the maximum correlation being 0.3 between Action and Adventure genres.

This analysis is done using a hierarchical model with hyperparameters as above. The necessary distributions are as follows:

- **Hyperpriors** - $P(\alpha, \beta) \sim (\alpha + \beta)^{-\frac{5}{2}}$
- **Prior** - $P(\theta_j|\alpha, \beta) \sim Beta(\alpha, \beta)$
- **Likelihood** - $P(z_j|\theta_j, n_j) \sim Bin(n_j, \theta_j)$

Where z_j is the number of movies in genre j that pass the Bechdel test and θ_j is the corresponding underlying proportion of movies that pass the Bechdel test in genre j. We have a posterior distribution similar to what we saw in the decades analysis.

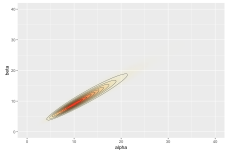


Figure 8: Marginal Distribution of α and β

For this analysis, we find from the marginal density that our α and β hyperparameters are now both at approximately 10 (Figure 8). Using the STAN package in R we again draw 2000 samples for each θ_j and obtain 95% confidence intervals as seen in Figure 9.

We find that representation of women in films varies across different genres. Western films have the fewest moves that pass the Bechdel test with an estimated proportion of about 0.33 while Romance movies have the most with an estimated proportion of about 0.68. Genres with fewer total movies typically have an over estimate of the underlying proportion as compared to the raw proportion of movies that pass.

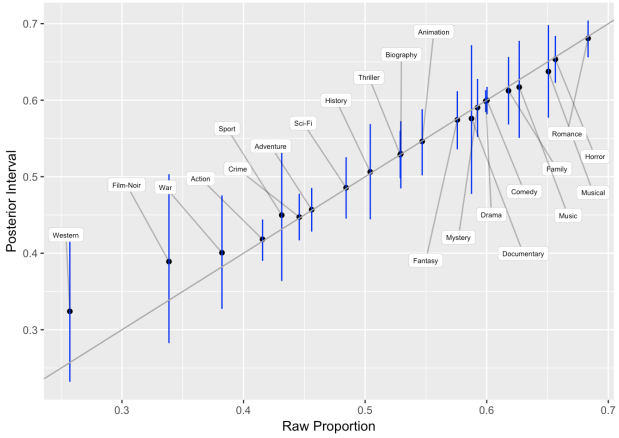


Figure 9: Estimates vs Raw proportions by Genre

4 RANDOM SAMPLING FOR HIERARCHICAL REGRESSION

In order to understand the relationship between year and pass rate by genre we needed each movie to have a unique genre. IMDb assigns each movie up to three genres. We found 56% of our movies had 3 genres, 30% had 2, and 14% had 1. In order to avoid additional bias we randomly choose

a genre among the three to assign as the sole genre. Below shows the randomized counts (Figure 11) of movies in each genre against the original counts (Figure 10), and we can see they follow almost the same order.

Rank	Genre	Count
22	Film-Noir	62
21	Western	70
20	Documentary	80
19	Sport	102
18	War	157
17	Musical	189
16	Music	225
15	History	234
14	Family	403
13	Biography	427
12	Animation	437
11	Fantasy	573
10	Sci-Fi	574
9	Mystery	633
8	Horror	911
7	Thriller	1023
6	Crime	1070
5	Adventure	1226
4	Action	1280
3	Romance	1428
2	Comedy	2749
1	Drama	4259
0	Total	18112

Figure 10: Full counts

Rank	Genre	Count
22	Film-Noir	19
21	Western	39
20	Sport	40
19	Documentary	61
18	War	72
17	Musical	78
16	History	84
15	Music	92
14	Family	137
13	Biography	157
12	Animation	179
11	Sci-Fi	186
10	Fantasy	196
9	Mystery	218
8	Thriller	394
7	Adventure	396
6	Crime	409
5	Horror	421
4	Action	462
3	Romance	581
2	Comedy	1243
1	Drama	2033
0	Total	7497

Figure 11: Random sample counts

Full and Randomly Sampled Counts for Regression

5 HIERARCHICAL REGRESSION MODEL

To determine if there is an association between time and passing the Bechdel test by genre we fit a hierarchical regression model. Our exploratory analysis and hierarchical models indicates a potential time trend, and we test to see whether there is an overall significant time trend for each genre. As our data is a convenience sample, any results are biased towards data collected, which seems to be newer movies, and is overall not representative of the movie population as a whole for any given year.

To determine whether year is positively associated with passing the Bechdel test we fit 22 slopes, while allowing different intercepts, for each genre. A positive slope indicates for a given genre, a newer movie has higher odds of passing the Bechdel test than an older movie.

In our model we center and scale year, then transform our coefficients back and exponentiate our results to obtain odds ratios instead of log odds ratios.

Each of the 22 genres, $j \in 1{:}22$ will have its own intercept α_j and slope β_j. Each β_j represents the odds of passing the Bechdel test as a function of year (x) from 1906 to 2019. Our model feeds data in at the movie level, and then clusters at the genre level. Our only predictor of whether a movie passed or not is year, and the odds of passing for a given genre are decided by year.

Hence, our $y_{ij} \in 0, 1$ is an indicator of whether an individual movie i of genre j passed. Our model is parameterized as follows:

- $\beta_j \sim N(0, \tau^2)$
- $\alpha_j \sim N(0,^2)$
- $y_{ij} \sim Bern(logit^{-1}(\alpha_j + \beta_j x_{ij}))$

Each y_{ij} is modeled Bernoulli with a logit model on the regression equation $\alpha_j + \beta_j x_{ij}$. α_j and β_j both have hierarchical priors that are Normal.

Genre	Beta
Film-Noir	0.0353
Music	0.0454
Documentary	0.0459
Fantasy	0.0468
Action	0.0470
Crime	0.0478
Musical	0.0511
Family	0.0518
Animation	0.0531
Adventure	0.0545
Biography	0.0548
Sci-Fi	0.0569
Comedy	0.0583
History	0.0598
Drama	0.0604
Thriller	0.0607
Sport	0.0609
War	0.0627
Horror	0.0636
Romance	0.0658
Western	0.0667
Mystery	0.0711

Figure 12: βs by Genre

Our $\beta_j s$ (Figure 12) indicate time is positively associated with the proportion of movies that pass the Bechdel test for every genre, some moreso than others. We can interpret our coefficient as follows; $\beta_{Mystery} = 0.0710$ indicates that for each year increase, the odds of passing the Bechdel test increase by 0.710, on average for the Mystery Genre. If one Mystery movie is filmed in 1980 and a different Mystery movie is in 2010, the odds of passing the Bechdel test are about 2.13 times higher for the newer movie, purely as a function of time.

We find the genres with the largest coefficients are Mystery, Western, and Romance. These are a mix of our most represented and least represented genres. On the other hand, our smallest coefficients, for Film-Noir, Music, and Documentary, come from genres with fewer observations.

The difference between our largest coefficient, for Mystery, and our smallest coefficient, for Film-Noir, for a 30 year period of time, is 2.13 increase in odds for Mystery and 1.07 increase in odds for Film-Noir.

6 CONCLUSION

Through our analysis with hierarchical modeling and hierarchical regression, we were able to identify that genre and time influence whether a movie passes the Bechdel test. With hierarchical modeling we learn that the raw and estimated proportions for each decade are similar aside from the 1900-1920 combined group which is the smallest decade. We also learn that there is a possible relationship between time and proportion that passed. We find a similar relationship when looking across genres. There appear to be underlying differences between the genres and the proportion of movies that pass the Bechdel test. In our hierarchical regression we learn that when we fit slopes and intercepts for each genre where the slopes are associated with

year, all of the slopes are positive. This corresponds to an increase in the odds that movies will pass the Bechdel test as time goes on. Overall, it does appear that women are better represented in modern day movies that they have been in the past.

Our data set included some difficulties, the most obvious being that it was not generated through a random sample of the population of movies. The data was created through crowd sourcing, so it includes bias to movies that the "crowd" is interested in and has watched. There is also an potential issue with there being fewer observations in the early years. This could partially be due to the fact that there were fewer movies made earlier in the twentieth century, but also because current watchers and reviewers don't watch those movies as often as they watch the newer releases.

There is still future analysis that could be done beyond what we've done for this project. We could gather more predictors and add them to our hierarchical regression model. We could also investigate fitting splines as there is some potential cyclical behavior in the proportion of movies that pass year to year.

7 REFERENCES

Bechdel Test Movie List, bechdeltest.com/.
title.basics.tsv.gz, IMDb.com, www.imdb.com/interfaces/.
title.ratings.tsv.gz, IMDb.com,www.imdb.com/interfaces/.