Gender Presentation and Successfulness of Movie

Statistics 302 Final Report

The 4 Avengers
Haowei Feng
William Han
Yeonseo Park
Yuxin Liu

May 10, 2019

Contents

Abstract

This project aim is to analyze the gender bias in the hollywood movies. A prior study showed that there is no significant evidence that movies with strong female roles perform worse in regards to its gross profit and investment return. This study further examines whether the opposite of Hollywood's belief is true, and to show that movies with strong female roles actually do better worldwide. The dataset is from the previous study with additional data collected from IMDb. The population of interest is the popular movies from 2011, which is representative of the recent trends of Hollywood movies. To operationalize the concept of wellness of movies, this study uses the variables of investment return, measured by international gross per dollar budget, and the success of movie, measured by whether the international gross is greater than three times the budget. The explanatory variables proxying the concept of gender presentation in movies are the Bechdel test and proportion of female characters on the theatrical poster. The research finds that movies that passed the Bechdel test and movies with higher proportion of female characters on the theoretical poster have a greater investment return. However, there is no sufficient evidence that movies that passed the Bechdel test have a higher probability of success. The result indicates that the belief that movie industry is holding is not only inaccurate, but rather opposite from the reality. Movies highlighting female characters in fact do better in the gross profit and investment return.

Introduction

In the past decades, the issue of gender inequality has risen in prominence. More recently, the "Me Too" movement has taken the hollywood industry by storm, and has shed a light on the issues that still remain prevalent today. In the past, women would either not get prominent roles or recieve roles that enforced long-standing stereotypes on females in American society. This is still a prevalent issue today. According to the "2019 Hollywood Diversity Report" published through UCLA, women make up only 39% of all lead roles. Although the situation may seem better now, the truth is, by looking closer at the specific way female characters are presented, women are still stereotyped and underrepresented in Hollywood.

A previous study reflects that increasing women representation may be in the best interest of Hollywood. In a study done through USC's Annenberg Inclusion Initiative (2018), 40 of the top 100 highest grossing films were led by women. This represents a 20% increase from 2007, where only 20 films had a female lead. This report will try to establish and confirm the relationship between gender representation and box office revenue from USC's study, and examine the gender representation in Hollywood in a deeper way than just focusing on the presence of the leading female roles.

This study examines sexism and gender inequality within the US movie industry. For the scope of the report, sexism will be separated into how the moviegoing audience inadvertently presents sexism through their purchasing preferences. This is defined by how they respond to the movies with different representations of gender. For the project, the level of response will be measured by the film's success in the box office. The purpose of the research report is to examine the effect of gender representation on the movie's box office success. The hypothesis is that movies where female characters do not play a significant role perform worse.

The explanatory variable, the gender presentation, will be measured through several criterias. The gender bias in the movie will be determined through the Bechdel test and how many women play a significant role. The Bechdel test examines three things: there are at least two named women in the movie, these two women talk to each other, and the women talk about something other than a man. The number of female characters will be estimated by the number of women featured in the movie's main poster and the percentage of female characters among all characters presented in the movie's main poster. These proxy variables are good indicators of the explanatory variable for several reasons. First, the Bechdel test, developed by Alison Bechdel's "Dykes To Watch Out For" in 1985, is one of the most widely-acknowledged tool used to measure the gender bias in movie. Second, the number of female characters featured on the main poster of the movie serves as a direct indicator for how many leading characters are portrayed in the movie. The proportion of female characters among all characters in the poster serves as a supplement indicator of how important the female roles are as opposed to the male roles.

The response variable, box office success, will be determined by the movie's gross revenue. Gross revenue will be split between domestic US and international box office revenues. There are several criterias to evaluate the movie's gross performance. The first two are the domestic and the international gross revenue, adjusted by inflation. The last two criteria are whether the domestic and international gross are each greater than three times the budget. Gross revenue is the direct measurement on how well US audiences responded to the movie. According to Appleton and Yankelevits (2010), a Hollywood movie would need roughly 3 times the budget in order to earn positive profit. Therefore, whether the box office revenue exceed 3 times the budget is an indicator for whether the movie is successful. International gross revenue is a essential part of the income for the Hollywood movies, and the scope may be extended to an international audience. In a report done through the Motion Picture Association of America (MPAA) [2016], international box office revenue makes up more than 70% of a film's revenue. The study will then also be able to examine sexism in the international audience.

The Bechdel test serves as a baseline floor for the minimum amount of female representation in movies. Some may argue that the test requires too little of movie directors and producers. As Alyssa Rosenberg(2018), a film critic for The Washington Post, notes, "the Bechdel test doesn't even get close to reorienting the entertainment industry towards women's stories, or to subject matter and emotional tones that have been traditionally considered female rather than excitingly masculine." However, the Bechdel test has done enough to help frame the conversation around certain form of evidence.

FiveThirtyEight did a famous report on this issue back in 2014. To advance on their findings, the study will include the number and proportion of females featured on the movie's poster. Yet as stated before, there are many arguments against using the Bechdel test as the baseline for female representation. Although this is not a perfect indication of a woman's role in the film, it will highlight how willing the movie studios are to promote the women of the movie. Besides, additional statistics about the number of female characters and the difference in proportion of gendered characters are included in this survey, which provides it a more solid understanding of the concept of interest, the gender presentation in Hollywood movies, than the previous research.

Methods

The population of interest is popular Hollywood movies from 2011. Hollywood movies made in 2011 are chosen as the population of interest since Hollywood, the world's largest movie producer, is essential in shaping the public view on genders worldwide. By examining the potential bias on how different genders are portrayed in the Hollywood movies, people could get a better sense on individual audiences' preference on gendered movie characters, as well as the public recognition on genders. There is a general view in Hollywood that "women will go to a 'guy's movie' more easily than guys will go to a 'woman's movie' " (Hickey). With the understanding of audience preference and public recognition, directors and production studios might argue that they do not feature women because these movies do not do as well in theaters. Only movies from 2011 are chosen from the original data set because they are more representative of the recent trend of sexism in movie industry. Furthermore, the data on movie box office revenues are adjusted by the inflation, and the closer the movies are published, the more precise the data would be. The year 2011 is specifically chosen since there is enough movies published that year for a statistical test to be done. The dataset provides information on the popular, mainstream movies. Because the more popular movies have a higher chance of focusing on the general public as targeted audience group, while the little-known movies tend to be only targeting at a particular, small group of audience, these popular movies are representative of the target population. Such movies with little popularity might not reflect how gender presentation is related to the performance of the movies, or movie producers' perception on the well-being of the movie, which are the concepts that this research report is interested in testing.

This research is an observational study, which includes a dataset for a population of US movies from 1990 to 2013. Observational study is chose due to the lack of feasibility of experimental design. The data was collected through accessing data from BechdelTest.com and The-Numbers.com. The site BechdelTest.com provides information of the Bechdel test collected by committed moviegoers who analyze films and ascertain if they pass the Bechdel test. The site has detailed, coded information for about 5,000 films. The-Numbers.com is a leading site for box office and budget data, collected through public information released from the production studios. It inventories financial information for roughly 4,500 films. When considering the financial information, all numbers are adjusted for inflation, using 2013 dollars. While hardly a complete record of contemporary films, this gave us a sample that has both rigorous evaluations of female character involvement as well as the most accurate financial data available online. The other relevant information is collected through IMDB.com. Information regarding the gender of characters featured on the movie's theatrical poster is collected from IMDB through observing and recording the gender of characters on the cover poster of movies.

The study uses convenience sampling, which is a type of non-probability sampling. The movies being sampled are selected based on two data sets: BechdelTest.com and The-Numbers.com. The site BechdelTest.com is operated by committed moviegoers who analyze films and ascertain if they pass the Bechdel test. The-Numbers.com is a leading site for box office and budget data. By extracting data from those two datasets, the sample is selected from a dataset of 1,615 films released between 1990 and 2013. Sampling bias may be present, since the sampling is not completely randomized and the list of movies being conducted the Bechdel test on the BachdelTest.com are chosen by individual reviewers. However, such bias is limited in several ways.

Although the sampling strategy of the study is not probability sampling, the data collected is still representative of the population of interest, Hollywood movies from 2011. On one hand, the BachdelTest.com is a website that reflects the unequal representation of gender in Hollywood movies. Reviewers from the website tend to include movies that portray gender in varied ways. Such technique ensures that neither the movies that have significant female roles, i.e. have higher possibility of passing the tests, or those that do not signify the female characters, i.e. have lower possibility of passing the tests, would be over-represented in their review list. Therefore the sample frame generated from the reviewed movies would be representative of both types of movies in the targeted population, popular Hollywood movies from 2011. On the other hand, since only the gender presentation in the popular movies produced in 2011 are of interest, the sample from the BachdelTest.com is a good indicator of the population since more popular and well-known movies tend to be viewed by committed moviegoers. Moreover, the sample size is large enough to be representative of the targeted population (about 200 movies out of all 600 movies published in 2011) .

As discussed in the previous section, the sample of 2011 is chosen, given that the sample size from our data package in year 2011 is sufficiently large. In order for a z-test to be done on the data, for the categorical variables, a minimum of 50 responses in each level is needed, and for numerical data, 200 cases is needed. That is because a z-test requires a roughly normal distribution of the observations.

Variable Name	**Var Category**	**Var Type**	**Var Level**
onlymen	Explanatory	Categorical(Nominal)	pass/fail
no talk	Explanatory	Categorical(Nominal)	pass/fail
posternum	Explanatory	Numerical	NA
posterprop	Explanatory	Numerical	NA
domgross	Response	Numerical	NA
intgross	Response	Numerical	NA
domsuc	Response	Categorical(Nominal)	Yes/No
intsuc	Response	Categorical(Nominal)	Yes/No

- onlymen: whether the movie passes test 3 of Bechdel test, which is whether the conversation between women is only about men
- notalk: whether the movie passes test 2, which is whether women talk to each other
- posternum: the number of female characters on the poster
- posterprop: the proportion of female characters on the theoratical poster among all characters
- domgross: Domestic gross (US) of the movies in 2013 inflation adjusted dollars
- intgross: Total International gross of the movies in 2013 inflation adjusted dollars
- domsuc: whether the movie has a domestic gross that is greater than 3 times its budget
- intsuc: whether the movie has a international gross that is greater than 3 times its budget
- return: the return of investment, calculated by intgross/budget, which is dollars return per dollar invested

R is the software that is going to be used to analyze the data for the study. R is a free software environment for statistical computing and the R language is broadly used among statisticians to develop data analysis. RStudio is powerful, open-source integrated development environment (IDE) for R, which runs on either desktop (Windows, macOS, and Linux) or in a browser connected to RStudio server. Additionally, R Markdown provides straightforward methods to conduct reproducible analyses, and allows users to share a single file that contains R code, comments, and metadata that is needed to do the analysis from start to the end. R Markdown helps the users combine chunks of R code and produce formatted HTML, PDF, or Word file. The research includes a dataset for a population of US movies from 1990 to 2013 and the data will be collected from BechdelTest.com and The-Numbers.com. BechdelTest.com has information of 5000 movies which committed moviegoers analyzed and ascertained if the films pass the Bechdel test. The-Numbers.com provides financial information(box office and budget data) for 4500 films. Additional information will be collected from movie posters posted on IMDB.com. This will provide the information on the number of women featured prominently on the movie's promoting material.

The software R is going to be used to collect the data effectively from the websites and the samples are going to be selected based on the two data sets. Packages used to conduct the statistical tests are knitr, ggplot2, dplyr, stargazer, and xtable. A special package of FiveThirtyEight is introduced to R to access the data frame from the research done by FiveThirtyEight. The data that are collected will be analyzed through R Markdown and it will be used as evidence to check if it converges to support the hypothesis.

Statistical Tests

The first test examines the difference in means of domestic return of investment for the movies that passes or fails Test 3. The return of investment is defined as the dollars return for every dollar invested, which is calculated by domestic gross over the budget, both adjusted for the 2013 inflation. The dollar return is chosen as the response variable because it takes into account the domestic gross as well as the budget of movies, which gives production companies and investors in Hollywood a better sense of how much they earn from investing a movie. Test 3, which is whether the women in the movie have a conversation not about men with each others, is chosen as the explanatory variable because it is a good indicator for how well female characters are portrayed in the movie. If the movie passes Test 3, it also passes all the three tests, meaning that the movie does have stronger female characters. The mean return of movies that passed the test is denoted as μ_p and the mean return of movies that failed Test 3 is denoted μ_f.

A randomization test is performed to test the difference in means of domestic return of investments. Randomization tests are a way to monitor the process of sampling with the assumption that the null hypothesis about the population is correct without actually have to sample from the population. By conducting the randomization test, the distribution of test statistics sampled from a population with null hypothesis being true is estimated, and by looking at how extreme the true sample statistic is in regard of the estimated sample statistic, the possibility for which the null hypothesis is actually correct could be calculated. Method 1 of the randomization test is used because there is no assumption that could be made about the features of the two groups, including SE and shape. The sample means for both groups are shifted so that the sample means, μ_p and μ_f are equal, as assumed by the null hypothesis. Each of the two groups are randomly sampled with replacement from the #independent samples# given the sample sizes and the randomization means are compared.

The null hypothesis is that there is no difference in mean domestic return of the groups of movies that passes and fails Test 3. The alternative hypothesis is that the group that passes Test 3 has a greater mean domestic return than the group that does not pass Test 3. The alternative hypothesis is chosen because the original study tests the lower tail, i.e. the Hollywood stereotype that movies that portray female characters have a lower return for investment, and they found that there was no significant evidence to support such stereotype. This research intend to take one step further and test whether the truth is the inverse of the stereotype, i.e. the movies that portray female characters in fact have a higher return for investments than those that does not portray female characters.

Literature shows that the success of movie, which is whether the investment on the movie is a wise choice, is measured by whether the international gross is greater than 3 times the budget, as discussed in the introduction. Therefore in test 2, the success of movies is chosen as the response variable. By testing whether there is a difference in proportion of success between the group of movies that passes and the group that fails the third Bechdel test, this study is able to advance prior findings and see if portraying strong female characters yields more successful movies. The proportion of successful movies, i.e. the movies that have a international gross greater than 3 times the budget, of the group that passes Test 3 is denoted p_p, and the proportion of successful movies of the group that does not pass Test 3 is denoted p_f.

To test the difference in proportion of the two groups, a t-test for the difference of proportions is conducted. The null hypothesis is that there is no difference between the possibility of success of the two groups that pass or fail test 3. The alternative hypothesis is that the group that passes Test 3 has a higher possibility of success than the group that fails Test 3.

Thirdly, a regression test is conducted to test for the correlation between the proportion of female characters on the poster and the international investment return. While the Bechdel test is widely acknowledged, there still exist concerns for whether it is a direct measurement of how well the movie portrays female characters. This research therefore add up to the previous research by creating a new proxy variable for the presentation of female characters that is more perceptual and intuitive–the existence of female characters on the theoretical poster. Since the theoretical poster is the primary source of information that movie audiences can have prior to watching the movie, it therefore should express the most important characters that highlight the key elements of the movie. The new proxy variable, denoted "posterprop," was obtained by measuring the proportion of female characters portrayed among all characters on the theoretical poster of a movie. This research is able to add validity to simply proxying the concept of gender presentation in Hollywood movies, since the Bechdel Test provides a theoretical measurement of prominence of female characters while the proportion of female characters on posters provide a intuitive sense of how many important female characters are in the movie. In order to test for whether the proportion of female characters on poster is correlated with the wellness of movie, a regression test is conducted, where the slope between the proportion of female characters on movie poster and the international return of investment, calculated by international gross over the budget, both adjusted for the inflation of 2013.

Results

Test 1

Hypotheses

The null hypothesis is that there is no difference in mean domestic return of the groups of movies that passes and fails Test 3. The alternative hypothesis is that the group that passes Test 3 has a greater mean domestic return than the group that does not pass Test 3.

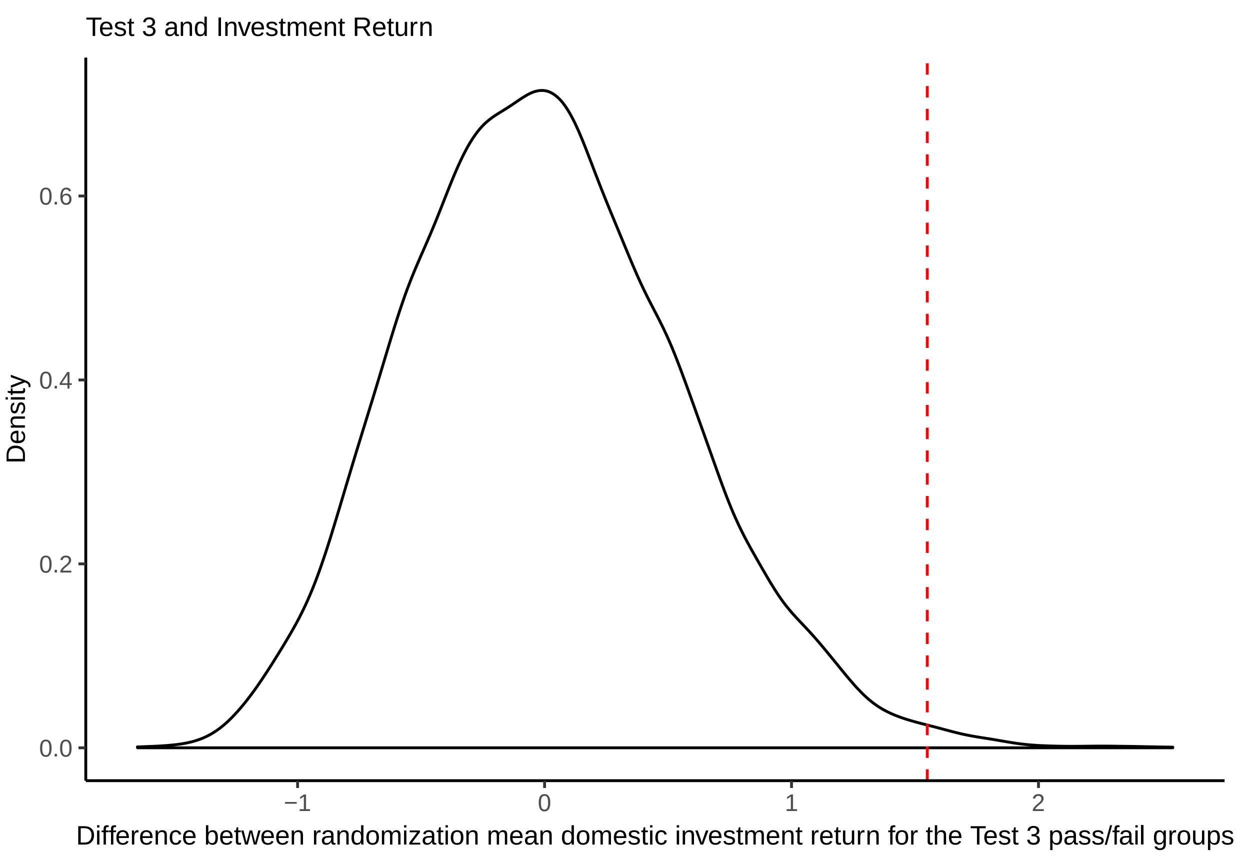
$$H_0 : \mu_p = \mu_f$$
$$H_a : \mu_p > \mu_f$$

Variables tested

Var Name	Var Category	Var Type	Var Level
Test3 successfulness	Explanatory	Categorical(Nominal)	pass/fail
International return in adjusted dollars	Response	Numerical	NA

Randomization Test Method 1

Check for Assumptions



The randomization distribution is approximately symmetric and bell-shaped, so there is no assumption that could be made about the features of the two groups, including SE and shape, and thus method 1 of randomization test is chasen.

Interpretation of the graph

This graph visually shows the result of the hypothesis test. The randomization samples represent what a random sample would look like if the population means were equal; the graph is centered around 0. Since it is a one-sided test, a reference line(red color) is drawn at the difference in means of the original sample, which is around 1.5 in this graph. A very small portion of the randomization data is greater than this line.

Compute p-value

The p-value of the sample data in a statistical test is the probability, when the null hypothesis is true, of obtaining a sample as extreme as (or more extreme than) the observed sample. The p-value is calculated by looking at the area under the curve to the right of the line in the above graph.

The p value calculated through the randomization test is 0.006. A smaller p-value provides stronger evidence against the null hypothesis. This very small p-value, which is below the significance level, provides strong evidence against the null hypothesis that there is no difference in mean domestic gross between the groups(mu_p) that pass and fail Test 3. Therefore, the null hypothesis is rejected and is in favor of the alternative hypothesis. There is sufficient evidence that the mean domestic investment return of movies that passed the Bechdel test is greater than the mean domestic return of movies that failed the test.

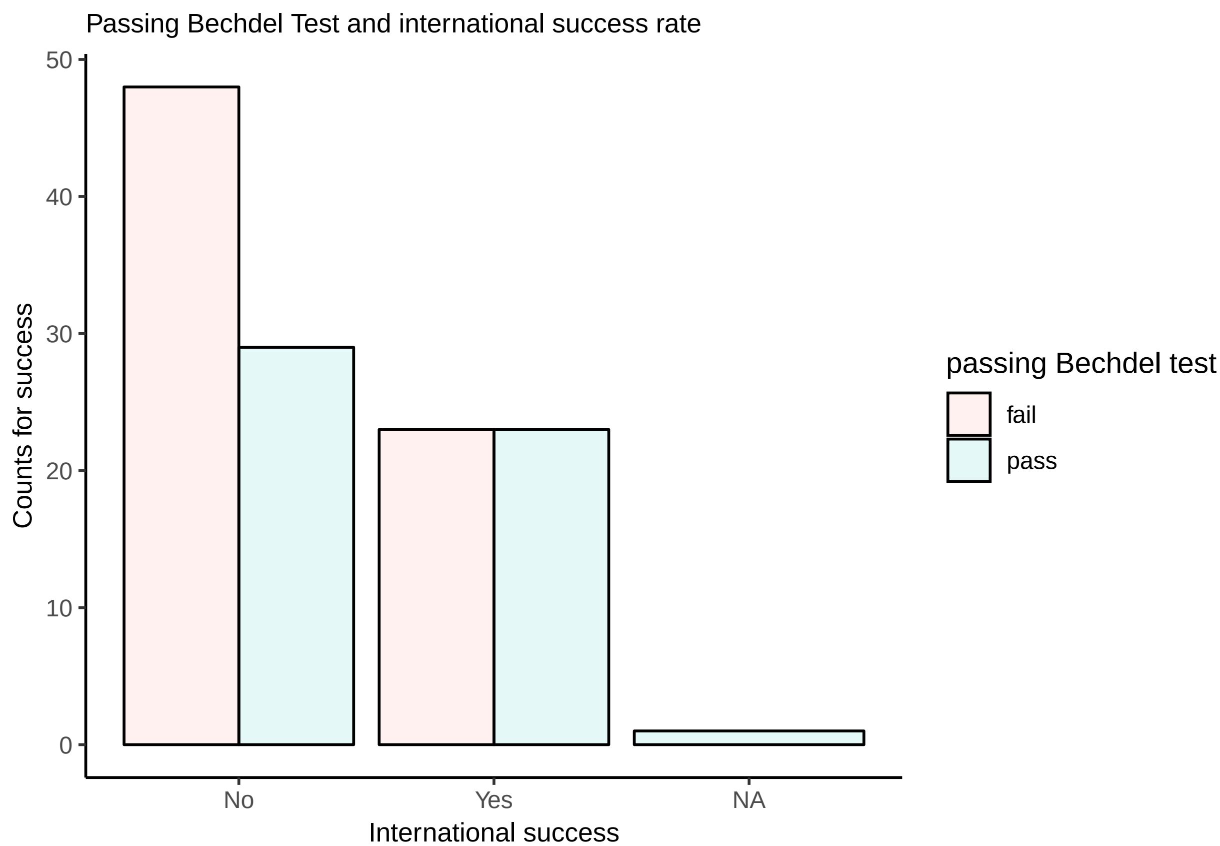
Test 2

Hypothesis

The null hypothesis is that there is no difference between the possibility of success of the two groups that pass or fail test 3. The alternative hypothesis is that the group that passes Test 3 has a higher possibility of success than the group that fails Test 3.

$$H_0 : p_p = p_f$$
$$H_a : p_p > p_f$$

Summary Figure and Check for Assumptions




```
    fail pass
No    48   29
Yes   23   23
```

```
    fail pass
No  TRUE TRUE
Yes TRUE TRUE
```

```
fail pass
  71   53
```

```
         fail      pass
No  0.6760563 0.5576923
Yes 0.3239437 0.4423077
```

The sample sizes are sufficiently large in each group such that $n_1p_1 \geq 10$, $n_1(1-p_1) \geq 10$, $n_2p_2 \geq 10$, and $n_2(1-p_2) \geq 10$ where $n_1 = 53$, $n_2 = 71$, $p_1 = 0.442$, and $p_2 = 0.323$. Therefore, a normal approximation is appropriate.

Calculate Test Statistic

$$\begin{aligned} z &= \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}} \\ &= \frac{0.442 - 0.323}{\sqrt{0.37(1-0.37)(\frac{1}{53} + \frac{1}{71})}} \\ &= 1.348 \end{aligned}$$

Compute p-value

$$\begin{aligned} Z &\sim N(0,1) \\ p - value &= P(|Z| \geq |z|) \\ &= P(|Z| \geq 1.348) \\ &= 0.105 \end{aligned}$$

Interpretation

The test could not reject the claim that there is no difference between the proportion of internationally successful and unsuccessful movies based on if it passed or failed the third Bechdel test. There is no significant evidence to show that the movies that passed the Bechdel test has a higher rate of success than those that does not pass the test. (upper-tail difference between proportions z-test, $n_1 = 53$, $n_2 = 71$, $p_1 = 0.442$, and $p_2 = 0.323$, $p = 0.1048$, $\alpha = 0.05$).

Test 3

Regression

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i,\ i = 1, \ldots, n$$

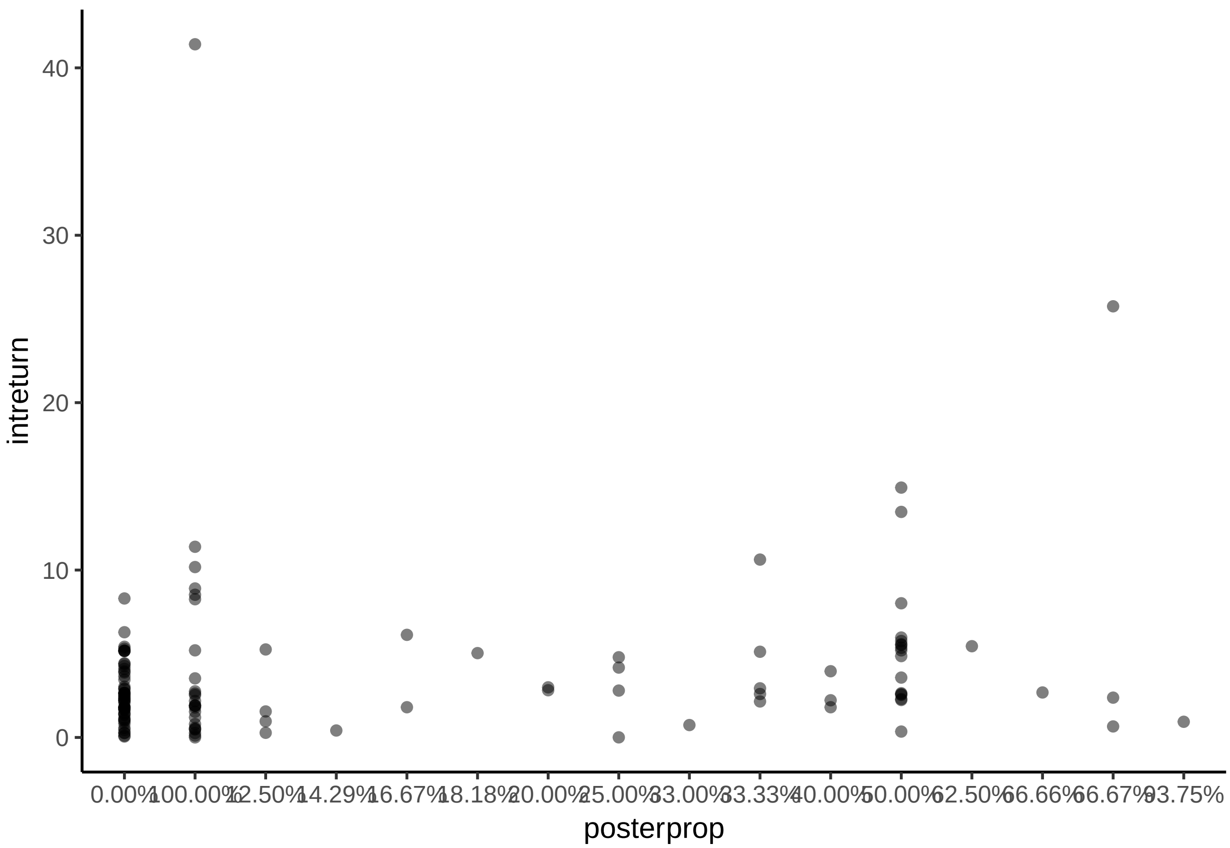
where $\varepsilon_i \sim N(0, \sigma^2)$.

Hypotheses

$$H_0 : \beta_1 = 0$$
$$H_a : \beta_1 \neq 0$$

where β_1 is the slope of the least squares line to predict the investment return based on the proportion of female characters.

Summary figure



There is no obvious curvature in the distribution, and the variability is relatively consistent along with the least square fitting line. However, there are several outliers, which may influence the validity of the estimation for slope.

Model Fitting

Calculate Test Statistic

$$\begin{aligned} t &= \frac{b_1 - 0}{SE} \\ &= \frac{b_1}{SE} \\ &= \frac{2.550}{1.101} \\ &= 2.318 \end{aligned}$$

Compute p-value

$$\begin{aligned} T &\sim t_{n-2} \\ p-value &= P(|T| \geq |t|) \\ &= P(|T| \geq 2.318) \\ &= 0.0222 \end{aligned}$$

Interpretation

There is significant evidence such that the the investment return is affected by the proportion of female characters portrayed on the movie poster. (t-test for slope, $t = 2.318$, $df = 121$, $p = 0.0222$).

Discussion

The randomization and t test performed in the study examined the effect of passing the third and final Bechdel test on the movie's wellness of being in the domestic and international markets. The randomization test focused on the domestic investment return, and the t test focused on the international success.

The regression test performed examined the relationship between the proportion of females on the theatrical poster and its international investment return. The goal was to show that there was a positive slope between the two variables, meaning that a higher proportion yielded a higher return for the companies.

Through the randomization and t-test, the study investigated the impact of passing test 3 of the Bechdel test on the movie's return on budget. Return on budget is calculated through dividing the movie's international gross revenue by its budget. This metric provides production companies a way to measure how much return they received from the cost of hiring all its staff and marketing the movie. Furthermore, the t test performed in the study focused on movies that exceeded a return of 3. These movies are deemed as successful movies for the production companies.

In the initial study done by FiveThirtyEight, the researchers ran a similar test. However, they were solely focused on the lower tail. In their study, they were successful in proving there was no evidence that movies that pass the Bechdel test performed worse than movies that failed. As discussed above, there is a general belief by Hollywood that moviegoers do not want to see movies that have strong female roles. The goal for the test performed in this study was to expand upon the conclusions made by FiveThirtyEight and prove that the movies passing the Bechdel test were actually more successful for the production companies, completely different from the widely held belief in Hollywood.

For the randomization test, the hypothesis was the movies that pass the third Bechdel test will have a greater mean of domestic investment success than the movies that fail the third Bechdel test. The p-value for the randomization test is .006 and therefore the study is able to reject the null that there is no difference in means.

For the t test, the hypothesis was the movies that pass the third Bechdel test will have a greater proportion of international investment success than the movies that fail the third Bechdel test. The t test performed for the difference in proportions yielded a p-value of .1048. The test was unable to reject the null hypothesis that there was no difference in the proportion of successful movies between movies that passed the third Bechdel test.

For the regression test, the estimated slope is 2.55. This positive slope signifies that an increase in the proportion of females on the poster leads to a positive impact on investment return. However, the test was only focused on proving that there was a relationship. The p-value calculated is .022. This means that the study is able to reject the null that the slope for the two variables is zero. Although no conclusions can be made on the exact value of the slope, a relationship does exist between the two variables.

Although the results of these two tests show an interesting difference between the international and domestic markets. Hollywood movies that pass the third Bechdel test are more successful in the US market than the international market. This difference may show an international bias over movies that have strong female roles. In other words, movies that make it to the international audience often do not pass the third Bechdel test. Therefore, these movies that were not successful for the production companies in the domestic release became a successful investment once they were shown abroad.

Error Analysis

The data analyzed only included movies from 2011. A major assumption used was that these movies were representative of the entire population. 2011 was initially chosen because it satisfied the requirements needed to run the three test. However, 2011 could have been an outlier year in terms of female representation. For example, Hollywood may have begun to feature more women with greater roles in their films. The study did not examine the year to year trends to make that distinction.

The poster data collected for this study involved recording the number of females on the IMDb movie poster. The study assumed that there was a relationship between the number of females featured on the poster and its female representation. However, many times the women featured on the posters play very stereotyped and minor roles. Therefore, this may not have been the best casual variable to select. Furthermore, errors could have been made in the data collection process. The study counted the number of females and the total number of individuals featured on the poster. A recording error could have been made in the process, negatively affecting the conclusions made in the study.

Further Studies

Moving forward, there are multiple areas that the study did not cover. Again, the study only used movies made in 2011. The same exact test could also be ran on movies from 2018 to see if there has been any improvement. 2018 saw the rise of the MeToo movement, and a further study could examine if it had any impact on female representation and performance.

The study based performance of the movies primarily by its box office revenue. Another way to examine how a movie is perceived is by looking at how critics and audiences rated the movie. The next study could incorporate Rotten Tomato and IMDb scores, along with the number of awards the movie received. It would be interesting to see if increased female representation is related to the reviews it receives. This also can help account for the box office performance of some movies. Even though a movie may pass the Bechdel test, its directing, writing, or acting can be bad and therefore affect how much money it makes.

Finally, a further study can examine the specific relationship between the proportion of females and its investment return. The study only focused on establishing that the relationship exists. Although the sample had a positive estimated slope, it would be interesting to examine this relationship across a wider dataset.

References

Appleton, D, Yankelevits, D. (2010, January 12). "Hollywood Dealmaking: Negotiating Talent Agreements for Film, TV, and Digital Media." *Allworth*.

Hickey, W. (2014, April 1). "The Dollar and Cents Case Against Hollywood's Exclusion of Women." *FiveThirtyEight*. Retrieved from https://fivethirtyeight.com/features/the-dollar-and-cents-case-against-hollywoods-exclusion-of-women/

Motion Picture Association of America (2016). "Theatrical Market Statistics." Retreived from https://www.mpaa.org/wp-content/uploads/2018/03/MPAA-Theatrical-Market-Statistics-2016_Final-1.pdf

Rosenberg, A. (2018, December 21). "In 2019, it's time to move beyond the Bechdel test." *The Washington Post*. Retrieved from https://www.washingtonpost.com/opinions/2018/12/21/its-time-move-beyond-bechdel-test/?utm_term=.d75706128dff

UCLA, College of Social Sciences. (2019, February 21). "Hollywood Diversity Report 2019." Retrieved from https://socialsciences.ucla.edu/wp-content/uploads/2019/02/UCLA-Hollywood-Diversity-Report-2019-2-21-2019.pdf

USC, Annenberg Inclusion Initiative (2018). "Inequality Across 1,200 Popular Films: Examining Gender and Race/Ethnicity of Leads/Co Leads from 2007 to 2018." Retrieved from http://assets.uscannenberg.org/docs/inequality-in-1200-films-research-brief_2019-02-12.pdf

Apendix

Data Cleaning R Code

```
bech = bechdel
num = which(bech$year == 2011)
df = bech[num, ]
table = read.csv("data.csv", header = TRUE)
df$posternum = table$posternum
df$posterprop = table$posterprop

onlymen <- NULL
for(i in 1:nrow(df)) {
  if(as.vector(df$clean_test)[i] == "ok"){
    onlymen[i] <- "pass"}
  else{
    onlymen[i] <- "fail"
  }
}
df$onlymen = onlymen

df$domreturn = df$domgross_2013/df$budget_2013
df$intreturn = df$intgross_2013/df$budget_2013

success <- NULL
for(i in 1:nrow(df)) {
  if(is.na(df$intreturn[i])){
    success[i] = NA
  }
  else if(df$intreturn[i] >= 3){
    success[i] <- "Yes"}
  else{
    success[i] <- "No"
  }
}
df$intsuccess = success
```

Randomization Test R Code

```
df.p <- df %>%
  filter(onlymen=="pass")
df.f <- df %>%
  filter(onlymen=="fail")

return.p <- df.p %>%
  select(domreturn) %>%
  pull()
return.f <- df.f %>%
  select(domreturn) %>%
  pull()

x.bar.p <- mean(return.p, na.rm = T)
```

```
x.bar.f <- mean(return.f, na.rm = T)

return <- df %>%
  select(domreturn) %>%
  pull()

x.bar <- mean(return, na.rm = T)

shift.p <- x.bar.p - x.bar
shift.f <- x.bar.f - x.bar

return.p.0 <- return.p - shift.p
return.f.0 <- return.f - shift.f

set.seed(302)

B <- 10000
n.p <- length(return.p)
n.f <- length(return.f)

mat.rand.p <- matrix(sample(return.p.0,B*n.p,replace=TRUE),
                     byrow = TRUE,
                     nrow = B,
                     ncol = n.p)
mat.rand.f <- matrix(sample(return.f.0,B*n.f,replace=TRUE),
                     byrow = TRUE,
                     nrow = B,
                     ncol = n.f)
rand.mean.p <- apply(mat.rand.p,1,mean, na.rm = T)
rand.mean.f <- apply(mat.rand.f,1,mean, na.rm = T)
rand.diff <- rand.mean.p - rand.mean.f

tol <- 1.0e-12
(p.value <- mean(rand.diff >= x.bar.p-x.bar.f-tol))
```

```
[1] 0.006
```

T-Test R Code

```
x.1 <- sum(df$intsuccess[df$onlymen=="pass"]=="Yes", na.rm = T)
x.2 <- sum(df$intsuccess[df$onlymen=="fail"]=="Yes", na.rm = T)
n.1 <- sum(df$onlymen=="pass", na.rm = T)
n.2 <- sum(df$onlymen=="fail", na.rm = T)


prop.test(x = c(x.1, x.2), n = c(n.1, n.2), alternative = "greater", conf.level = 0.95, correct = FALSE
```

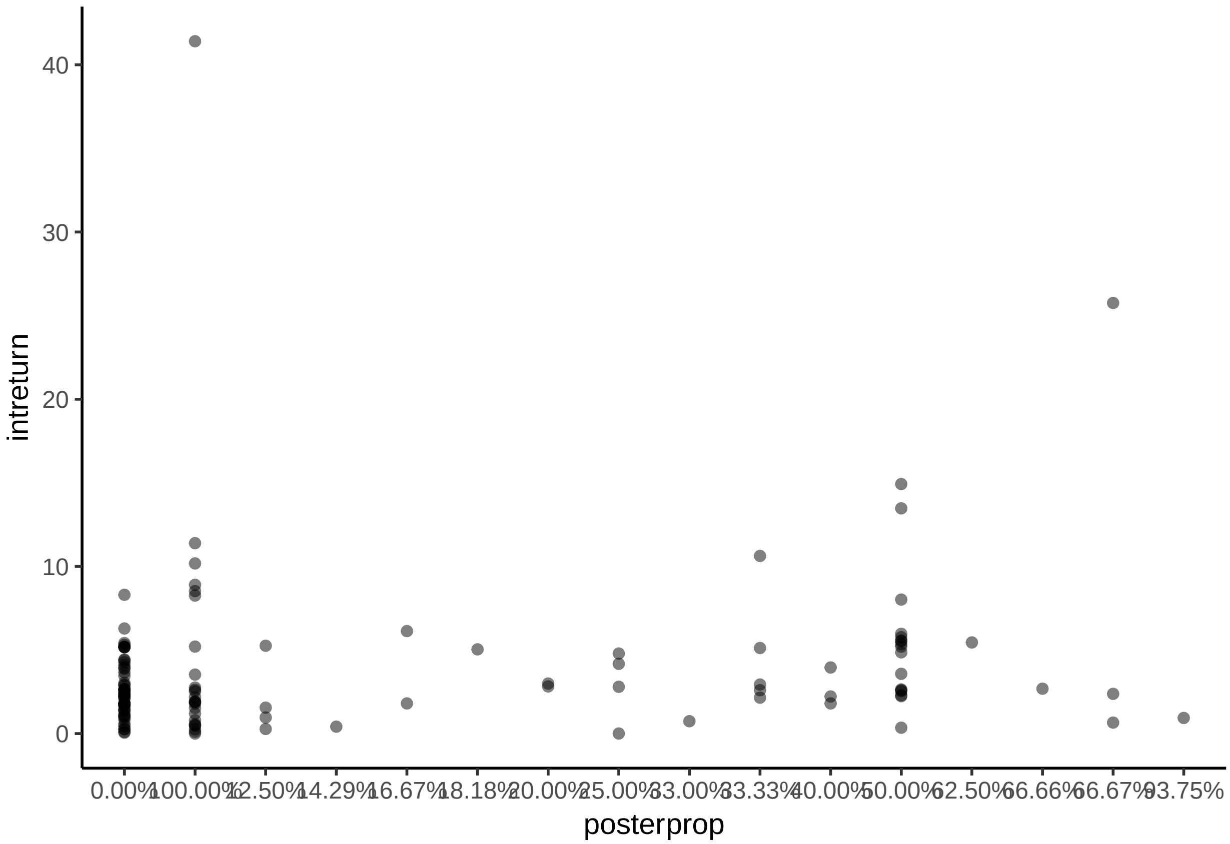
```
    2-sample test for equality of proportions without continuity
    correction

data:  c(x.1, x.2) out of c(n.1, n.2)
X-squared = 1.5741, df = 1, p-value = 0.1048
```

```
alternative hypothesis: greater
95 percent confidence interval:
 -0.03449716  1.00000000
sample estimates:
   prop 1    prop 2
0.4339623 0.3239437
```

Regression Test R Code

```r
ggplot(df, aes(x=posterprop,y=intreturn)) +
  geom_point(alpha=0.5) +
  geom_smooth(method="lm",se=FALSE) +
  theme_classic()
```




```r
lm.gross <- lm(intreturn ~ posterprop, df)
coef(lm.gross)
```

```
    (Intercept) posterprop100.00%  posterprop12.50%  posterprop14.29%
     2.67575219        2.14179096       -0.66244480       -2.25987360
 posterprop16.67%  posterprop18.18%  posterprop20.00%  posterprop25.00%
     1.29185111        2.36297810        0.23241944        0.26738731
 posterprop33.00%  posterprop33.33%  posterprop40.00%  posterprop50.00%
    -1.93537567        2.00933700       -0.01491380        2.67101990
 posterprop62.50%  posterprop66.66%  posterprop66.67%  posterprop93.75%
     2.77547609        0.01143472        6.92162790       -1.73853785
```

```
summary(lm.gross)
```

```
Call:
lm(formula = intreturn ~ posterprop, data = df)

Residuals:
   Min     1Q Median     3Q    Max 
-8.939 -2.180 -0.440  0.876 36.590 

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)        2.67575    0.67708   3.952 0.000139 ***
posterprop100.00%  2.14179    1.18826   1.802 0.074289 .  
posterprop12.50%  -0.66244    2.53338  -0.261 0.794220    
posterprop14.29%  -2.25987    4.92918  -0.458 0.647546    
posterprop16.67%   1.29185    3.51819   0.367 0.714201    
posterprop18.18%   2.36298    4.92918   0.479 0.632643    
posterprop20.00%   0.23242    3.51819   0.066 0.947452    
posterprop25.00%   0.26739    2.53338   0.106 0.916141    
posterprop33.00%  -1.93538    4.92918  -0.393 0.695369    
posterprop33.33%   2.00934    2.28607   0.879 0.381399    
posterprop40.00%  -0.01491    2.89906  -0.005 0.995905    
posterprop50.00%   2.67102    1.36407   1.958 0.052819 .  
posterprop62.50%   2.77548    4.92918   0.563 0.574565    
posterprop66.66%   0.01143    4.92918   0.002 0.998153    
posterprop66.67%   6.92163    2.89906   2.388 0.018715 *  
posterprop93.75%  -1.73854    4.92918  -0.353 0.725005    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.882 on 107 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.1028,	Adjusted R-squared:  -0.02297 
F-statistic: 0.8174 on 15 and 107 DF,  p-value: 0.6564
```

```
anova(lm.gross)
```

```
Analysis of Variance Table

Response: intreturn
            Df  Sum Sq Mean Sq F value Pr(>F)
posterprop  15  292.28  19.485  0.8174 0.6564
Residuals  107 2550.71  23.838               
```

```
xtable(summary(lm.gross)) %>%
  kable()
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2.6757522	0.6770752	3.9519277	0.0001394
posterprop100.00%	2.1417910	1.1882621	1.8024566	0.0742888
posterprop12.50%	-0.6624448	2.5333833	-0.2614862	0.7942203
posterprop14.29%	-2.2598736	4.9291817	-0.4584683	0.6475459
posterprop16.67%	1.2918511	3.5181858	0.3671924	0.7142007
posterprop18.18%	2.3629781	4.9291817	0.4793855	0.6326428

	Estimate	Std. Error	t value	Pr(>|t|)
posterprop20.00%	0.2324194	3.5181858	0.0660623	0.9474515
posterprop25.00%	0.2673873	2.5333833	0.1055455	0.9161405
posterprop33.00%	-1.9353757	4.9291817	-0.3926363	0.6953692
posterprop33.33%	2.0093370	2.2860689	0.8789486	0.3813988
posterprop40.00%	-0.0149138	2.8990627	-0.0051444	0.9959050
posterprop50.00%	2.6710199	1.3640710	1.9581238	0.0528189
posterprop62.50%	2.7754761	4.9291817	0.5630704	0.5745650
posterprop66.66%	0.0114347	4.9291817	0.0023198	0.9981534
posterprop66.67%	6.9216279	2.8990627	2.3875399	0.0187154
posterprop93.75%	-1.7385379	4.9291817	-0.3527031	0.7250048

```r
xtable(anova(lm.gross)) %>%
  kable()
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
posterprop	15	292.2773	19.48516	0.8173852	0.656369
Residuals	107	2550.7090	23.83840	NA	NA