Junwen Fang, Lyric Li, Rosy Xu

1 Introduction and Assumptions

This report explores movie ratings to uncover audience preferences. We will apply hypothesis tests to address questions about genre appeal, release year impact, demographic differences, and evolving trends in movie preferences. We assume the sample data is **randomly** drawn from the population and is **representative and independent with each other**. By the end of this analysis, we aim to provide insights valuable to the entertainment industry and film enthusiasts.

2 Basic Analysis

2.1 Are movies that are more popular rated higher than movies that are less popular?

We first categorize movies into two groups: more popular and less popular, based on the median number of ratings per movie. Since rating data is not continuous, we cannot reduce our data to the mean ratings. Therefore, instead of a two-sample t-test, we decide to use **one-sided Mann–Whitney U test** to compare the median ratings of more popular and less popular movies. In order to do this, we combine ratings into one list and drop all NaN value element wise for each group. Our **p-value is 0.0** and **test statistic is 7.42e8**. As a result, we have ample evidence to reject the null hypothesis and conclude that more popular movies tend to have higher average ratings than less popular movies. Though in this question, we are concerned about the inconsistent distributions of two groups because the Mann-Whitney U test can only test on the median if two independent variables have the same distribution. The **histograms** for two groups' distributions are shown in Graph 2.1.1.

2.2 Are movies that are newer rated differently than movies that are older?

We conduct a median split based on the year of release to categorize movies into "older" and "newer" groups. Initially, we determine the median release year, which is 1999, by scraping the years from the movie name. Movies released before the median year are assigned to the "older" group, while those released at and after are included in the "newer" group. Subsequently, for each group, we combine ratings in one list and drop NaN element wise. We also draw the **histograms** for two groups as shown in Graph 2.2.1, and find their distributions are **similar**. By running a **two-sided Mann–Whitney U test** to assess the median of older movies and newer movies, we get **p-value** is 1.28e-6 < 0.005 and **test statistic is 1.50e9**. Therefore, we have sufficient evidence to reject the null hypothesis. Older movies tend to have the similar average mean as newer movies.

2.3 Is enjoyment of 'Shrek (2001)' gendered, i.e. do male and female viewers rate it differently?

First, we group the rating data by the gender of raters, and drop NaN value element wise. Then, we plot the histogram (Graph 2.3.1) to see the distribution of our data and they are skewed to the left. Because it is unreasonable to reduce our data to sample means, we instead answer the question in two ways: if male and female rating distributions are different and if their rating medians are different. To test distribution, we utilize **KS test** and **p-value is 0.06** and **test statistic is 0.10**. To evaluate test median, we utilize **Mann-Whitney U test** and **p-value is 0.05** and **test statistic is 9.68e4**. Both of them are larger than 0.005 so we do not have sufficient evidence to reject the null hypothesis. Thus, movie ratings that are from men tend to have the same distribution and median as women. In this test, we assume the men's ratings and women's ratings are **independent** with each other, though in reality, this might not be the case.

2.4 What proportion of movies are rated differently by male and female viewers?

Similar to the previous question, we divide raters based on gender for each movie and process the data. We then repeat the **KS test** for each movie by using a for loop and count the significant tests. At last, we yield 25 significant tests out of 400. As a result, we conclude that there are **6.25%** movies that tend to rate differently by male and female viewers in terms of distribution. In terms of median, we repeat the **Mann-Whitney U test**, there are 50 significant tests out of 400, which means **12.5%** movies tend

to have different median ratings between males and females. We are concerned about applying these tests recursively on movies because we are not able to check if distributions between female and male are similar for each movie. In order to test the median, we are assuming the rating distributions are similar.

2.5 Do people who are only children enjoy 'The Lion King (1994)' more than people with siblings?

First, we group the rating data by if people are only children, and drop NaN value element wise. Then, we plot the graph to see the distribution of our data and found distributions similar between two groups (Graph 2.5.1). Although the number of people who are only children is very different from the number of people with siblings, we found the distributions for both groups are a bit skewed to the left and not satisfied the conditions for independent t test. So we employed the **one-sided Mann–Whitney U test**. The **p-value is 0.98 > 0.005** and **test statistic is 5.29e4**, we do not have sufficient evidence to reject the null hypothesis. Thus, movie ratings that are from only children tend to have identical distribution and there are no difference in the medians as people with siblings.

2.6 What proportion of movies exhibit an "only child effect", i.e. are rated differently by viewers with siblings vs. those without?

2.6 is an variated question of 2.5 with almost similar background but more movies, so we use the similar method to find the answer. For each movie, we separate rating data by whether the audience has siblings or not, then apply the Mann–Whitney U test. The only difference is that we run a **two-sided Mann–Whitney U test** to see whether viewers with siblings rate **differently** as viewers without and found out that **1.75% of movies** exhibit the "only child effect". Similar to our concern in 2.4, we are not sure if rating distributions between female and male are similar.

2.7 Do people who like to watch movies socially enjoy 'The Wolf of Wall Street (2013)' more than those who prefer to watch them alone?

We first divide the rating data of movies 'The Wolf of Wall Street (2013)' by people who watch movies socially and alone. Then, we checked the distribution of people who watch movies socially and people who watch movies alone. The two distributions look similar, and both are highly left-skewed (Graph 2.7.1). After the observation, we used a test to prove it. Since the highly skewed rating data doesn't fulfill the assumption of independent t-test, we apply the **one-sided Mann–Whitney U test** to compare the median of the two groups of ratings, then we get the **p-value 0.94 which is higher than 0.05**. The **test statistic is 5.68e4**. So we conclude that we do not have enough evidence to reject the null hypothesis. Thus, people who like to watch movies socially "The Wolf of Wall Street (2013)" do not tend to enjoy it more than those who prefer to watch them alone.

2.8 What proportion of movies exhibit such a "social watching" effect?

Since this question is an extension of 2.7, we apply the same method on every movie. As we did on 2.7, applying **one-sided Mann–Whitney U test** on sample medians, using the one tail p-value to conclude. We then get the result that **1.5% of movies** exhibit such a "social watching" effect. We assume that rating distributions among people who watch movies socially and people who prefer to watch movies alone are similar for each movie, so that the significant tests representing the medians between two groups are different.

2.9 Is the ratings distribution of 'Home Alone (1990)' different than that of 'Finding Nemo (2003)'?

First, we draw the distribution of both movies. From the histogram (Graph 2.9.1), we notice the distribution of Finding Nemo is more skewed than that of Home Alone. Then we apply the **KS test**, which is used for analyzing the differences between two distributions to find the answer. With KS test, we get the **p-value** which is **3.26e-10** < 0.005, so we have sufficient evidence to prove movie rating distributions of 'Home Alone' tend to have different distributions as 'Finding Nemo'.

2.10 There are ratings on movies from several franchises (['Star Wars', 'Harry Potter', 'The Matrix', 'Indiana Jones', 'Jurassic Park', 'Pirates of the Caribbean', 'Toy Story', 'Batman']) in this dataset. How many of these are of inconsistent quality, as experienced by viewers?

To answer this question, we first define inconsistent quality as movies having different medians. For each franchise, there are more than two movies in and distributions of some movies do not fulfill normality as shown in Graph 2.10.1. So instead of using ANOVA, we apply the **Kruskal-Wallis H-test** on each of the franchises. As shown in table below, the p-value for "Harry Potter" franchise yields a **p-value of 0.34** which is larger than 0.005. The Kruskal-Wallis H-test with rest franchises all yield **p-value lower than 0.005** and **test statistics larger than 20**. Thus, we conclude that **all movies except Harry Potter** have inconsistent quality.

	Franchise	Test_statistics	P-Value	Reject Null Hypothesis?
0	Harry Potter	3.331231	3.433195e-01	False
1	Pirates of the Caribbean	20.643998	3.290129e-05	True
2	Toy Story	24.385995	5.065805e-06	True
3	Indiana Jones	45.794163	6.272776e-10	True
4	Jurassic Park	46.590881	7.636930e-11	True
5	The Matrix	48.378867	3.123652e-11	True
6	Batman	190.534969	4.225297e-42	True
7	Star Wars	230.584175	8.016477e-48	True

There are still some concerns about this Kruskal-Wallis H-test, because it presumes that within each franchise, the movie ratings are independent of each other. In reality, however, they might not be independent. For example, the ratings for Harry Potter and Chamber of Secrets (2002) can be influenced by the rater's feeling on Harry Potter and the Sorcerer's Stone.

3 Further Analysis (Extra Credit)

To further explore the dataset, we are interested in that for people who "feel the things are happening to them" when watching a movie, if they tend to give higher movie ratings. Several potential intuitions are following:

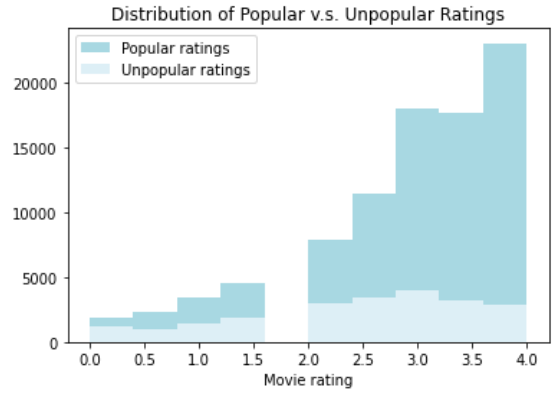
1. It is possible that the movie utilizes technology which made them vivid, so these people might be more likely to give higher ratings.
2. If one is engaged in the movie, he/she must be immersed in the movie and fully absorbed in the story. As a result, their emotion is closely related to the characters and give a higher rating.

As a result, we propose our null hypothesis as "For raters who are more likely to feel things on screen are happening to them, their median movie ratings will be the same as raters who are less likely to feel things on screen are happening to them.", and alternative hypothesis as "For raters who are more likely to feel things on screen are happening to them, their mean movie ratings will be higher than raters who are less likely to feel things on screen are happening to them."

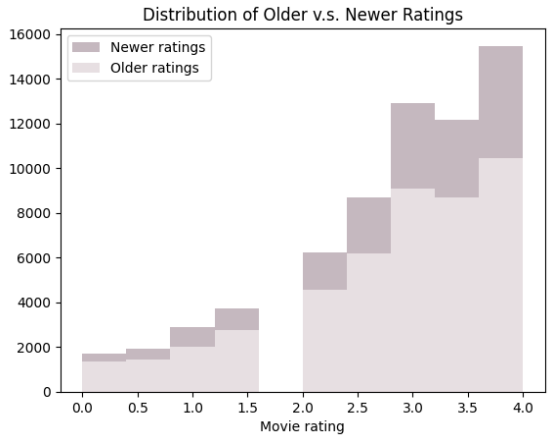
To test our hypothesis, we first divide people into two groups, based on the column "When watching a movie I feel like the things on the screen are happening to me". If ratings in this column is higher than 3, then the rater belongs to the "high engagement" group. If it is lower than 3, the rater belongs to the "low engagement" group. We then stack ratings for each rater in each group, and visualized the histogram (Graph 3.1) of rating for each group, which both showed similar distributions. Then, we ran a **one sided Mannwhitneyu test**, and the **p-value is 1.08e-119** and **test statistics is 7.86e8**. Because the p-value is less than 0.005, we

have sufficient evidence to **reject** the null hypothesis. We conclude that the median movie ratings from raters who are more likely to 'feel the things happening on them' tend to have a higher rating on a movie compared to those who are less likely to feel so.

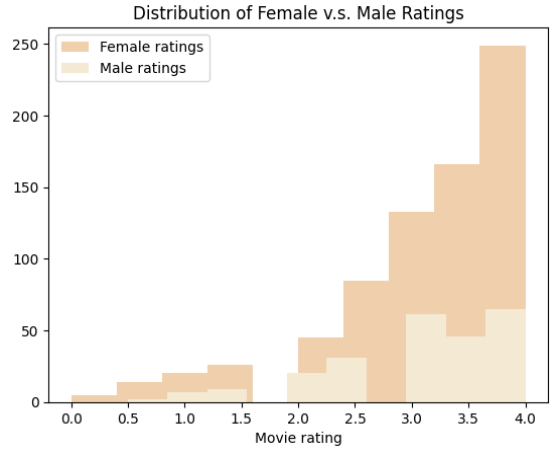
Appendix



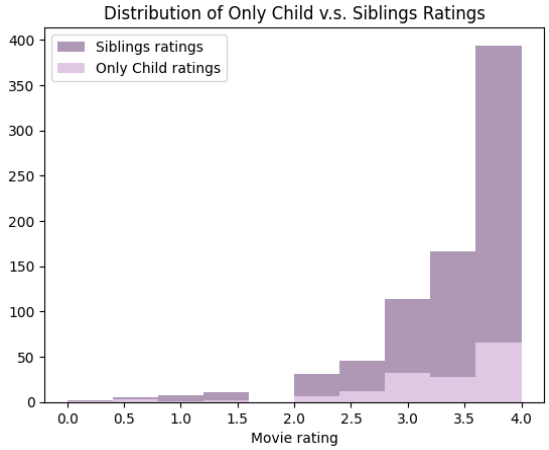
Graph 2.1.1 Distribution of Popular v.s. Unpopular Ratings



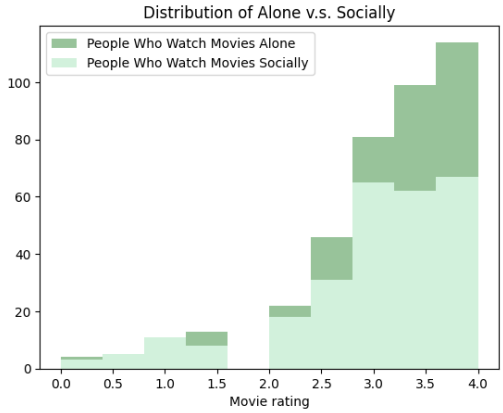
Graph 2.2.1 Distribution of Newer and Older Ratings



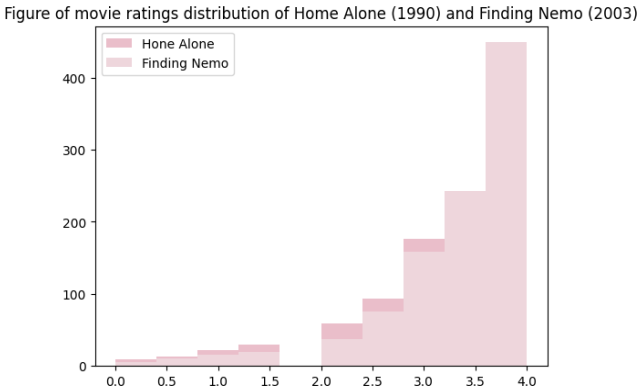
Graph 2.3.1 Distribution of Female v.s. Male Ratings



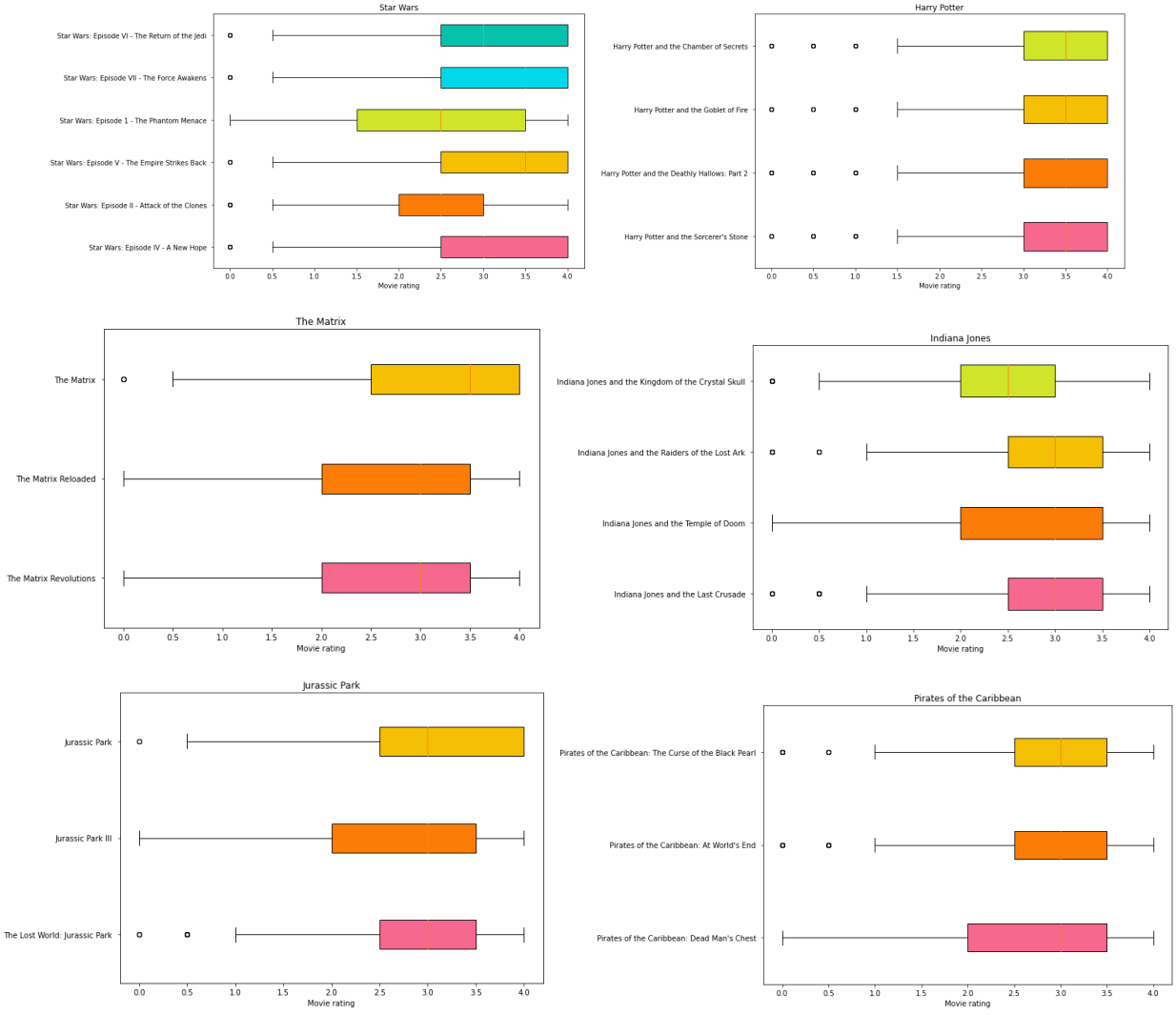
Graph 2.5.1 Distribution of Only Child v.s. Siblings Ratings



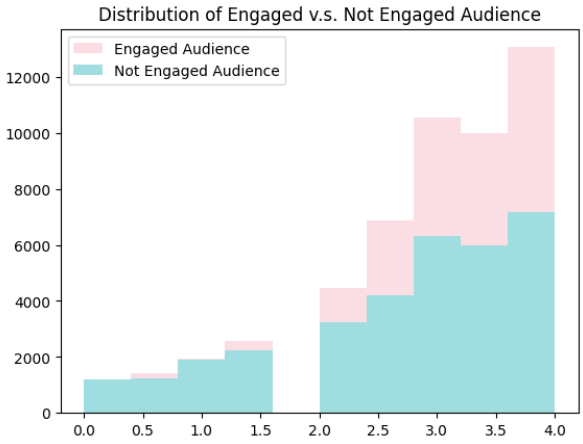
Graph 2.7.1 Distribution of The Wolf of Wall Street



Graph 2.9.1 Distribution of Home Alone and Finding Nemo



Graph 2.10.1 Boxplot of Movies Within Each Franchise



Graph 3.1.1 Distribution of Engaged v.s. Not Engaged Audience