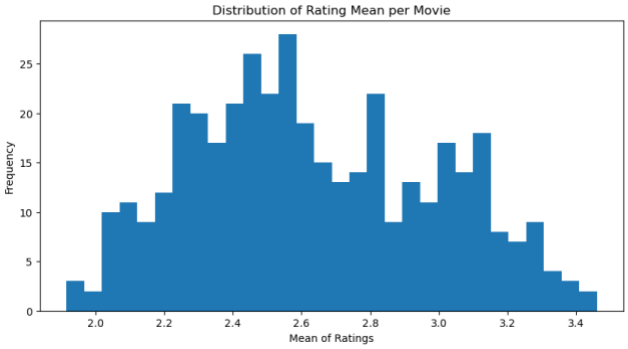
Project 01
Group 16

Group Number:
Asif Tauhid
Xiaokan Tian
Haojie Cai

Q1: Are movies that are more popular (operationalized as having more ratings) rated higher than movies that are less popular?

D: 1. Split movie rating df into 2 groups by the median of rating count. 2. Calculate the means of these 2 groups. 3. Check the distribution of rating means. 4. Using Levene's Test to check if variances are similar. 5. Welch t-test. 6. Compare the mean of 2 groups



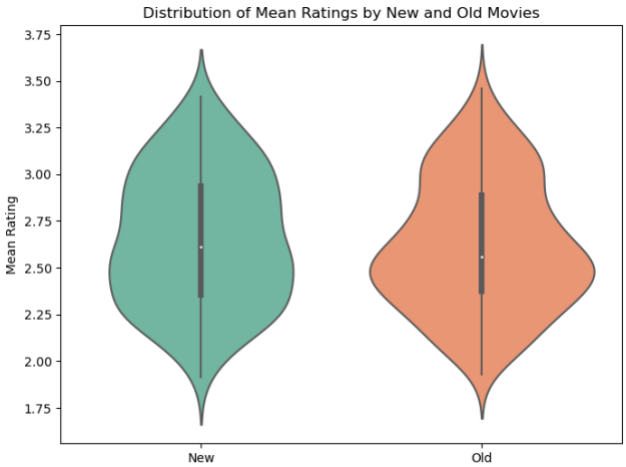
Y: The H0 is "The average ratings for popular and unpopular movies are the same." -> the distribution is normal -> 2 sample means -> population parameters are unknown -> variances are not similar -> use Welch t-test.

F: t-statistic = 17.7560; p-value = 9.5369E-52, which is less than 0.005; we can reject the H0, there is a statistically significant difference in average ratings. By comparing the mean of popular and less popular movie groups, popular rating(2.8683) > unpopular rating(2.4009)

A: The average ratings for popular and unpopular movies are not the same. And popular movies' rating is higher than less popular movies' rating

Q2: Are movies that are newer rated differently than movies that are older?

D: 1. Split movie rating df into 2 groups by the median of year. 2. Calculate the means of these 2 groups. 3. Check the mean rating distribution of old and new movies 3. Using Levene's Test to check if variances are similar. 4. Independent samples t-test



Y: The H0 is "The average ratings for new and old movies are the same." -> the distribution is normal -> 2 sample means -> population parameters are unknown -> variances are similar -> independent samples t-test.

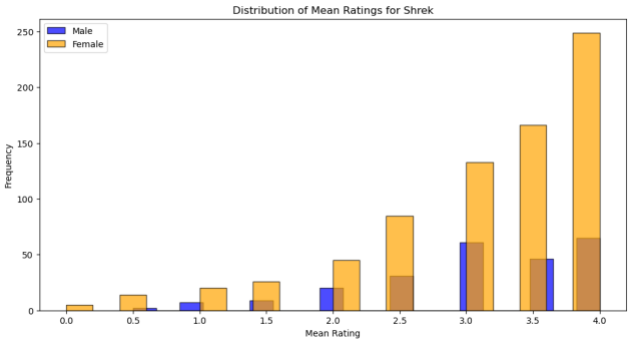
F: t-statistic = 1.1238; p-value = 0.2618, which is higher than 0.005; we cannot reject the H0, there is no statistically significant difference in average ratings.

A: The average ratings for new and old movies are the same.

Q3: Is the enjoyment of 'Shrek(2001)' gendered, i.e. do male and female viewers rate it differently?

D: 1. Extracted the ratings of 'Shrek (2001)' and split them into male and female groups. 2. Check the rating distribution of male and female 3. U test 4. KS test

Y: The H0 is "The men's and women's average ratings for 'Shrek(2001)' are same." -> the distribution is not normal -> not categorical -> 2 groups -> compare both -> U test, KS test.

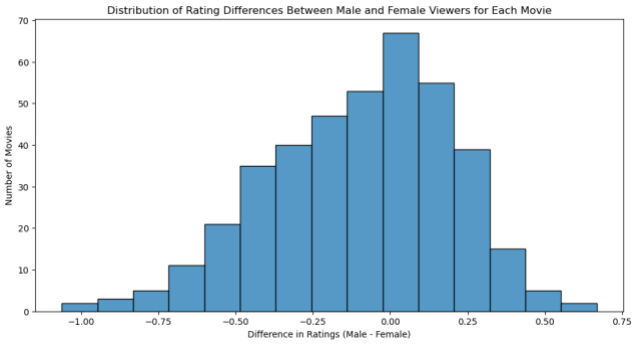


F: U-statistic = 82232.5; p-value-U = 0.051, which is higher than 0.005; we cannot reject the H0, there is no statistically significant difference in ratings medians. KS-statistic = 0.098; p-value-U = 0.056, which is higher than 0.005; we cannot reject the H0, there is no statistically significant difference in ratings distributions.

A: The men's and women's ratings for 'Shrek(2001)' are the same in both rating medians and distributions.

Q4: What proportion of movies are rated differently by male and female viewers?

D: 1. We separated the ratings of each movie by gender. 2. Used the Mann-Whitney U test to compare male and female ratings for each movie. 3. Calculated the proportion of movies with statistically significant gender-based differences (Those who had p-value < alpha).



Y: The H0 is "There is no difference in ratings for movies between male and female viewers." Since the ratings are ordinal, and we assume the data is not normally distributed, we used the Mann-Whitney U test.

F: The proportion of movies rated differently by male and female viewers is 0.1250.

A: Approximately 12.5% of movies show significant differences in ratings, showing a measurable gender-based difference in ratings for a subset of movies.

Q5: Do people who are only children enjoy 'The Lion King (1994)' more than people with siblings?

D: 1. Separated The Lion King ratings by only-child status. 2. Conducted the Mann-Whitney U test to compare the ratings between only children and those with siblings.

Y: The H0 is "The enjoyment of The Lion King is the same for only children and people with siblings." Given the ordinal scale of the data, the Mann-Whitney U test seemed suitable for comparing the two groups.

F: The p-value = 0.0432, which is greater than the significance level $\alpha = 0.005$; thus, we cannot reject the null hypothesis.

A: There is no statistically significant difference in enjoyment of *The Lion King (1994)* between only children and people with siblings.

Q6: What proportion of movies exhibit an "only child effect," i.e., are rated differently by viewers with siblings vs. those without?

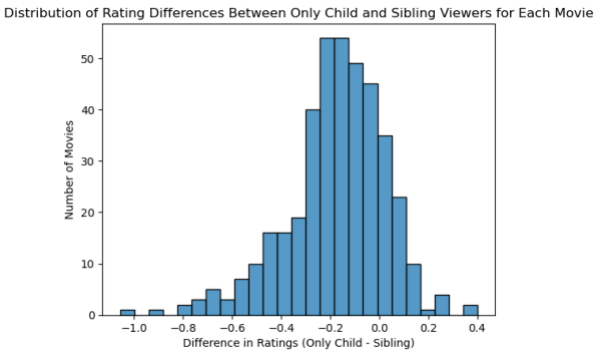
D: 1. We separated ratings for each movie by only-child status. 2. Used the Mann-Whitney U test for each movie to identify significant differences. 3. Calculated the proportion of movies with significant "only child effect" (Those who had p-value < alpha).

Y: The H0 is "There is no difference in ratings for movies between only children and people with siblings." The Mann-Whitney U test is used due to the ordinal nature of the ratings.

F: The proportion of movies with a significant "only child effect" is 0.0175.

A: There is a small "only child effect" in the dataset. Only 1.75% of movies were rated differently by only children and people with siblings.

Q7: Do people who like to watch movies socially enjoy '*The Wolf of Wall Street (2013)*' more than those who prefer to watch them alone?



D: Based on the assumption that there's no significant difference in enjoyment of '*The Wolf of Wall Street*' between people who enjoy watching movies socially and those who prefer watching alone, we extracted people's ratings for the movie, separated them into two groups based on their watching preference: watching alone or socially, did Mann-Whitney U test, and calculated the effect size and the statistical power.

Y: Even though the sample size of 663 is relatively large, we can't assume cardinality and normality of the ratings for the movie, and we have only 2 groups to study. In this case, we choose the Mann-Whitney U test. We also calculate effect size and power to help us understand the practical significance and the likelihood of detecting a true effect.

F: 1. Mann-Whitney U test: Statistic: 49303.5; P-value: 0.113.

2. Power analysis: Cohen's d: 0.123; Power: 0.099. (under 10000 simulations)

A: Given that the p-value is 0.113, we are unable to reject the null hypothesis that there's no significant difference in enjoyment of '*The Wolf of Wall Street*' between people who enjoy watching movies socially and those who prefer watching alone. This is further supported by the small effect size. However, we get a low effect size, which means we might not detect a true effect if one exists. Thus, further research with a larger sample size would be better.

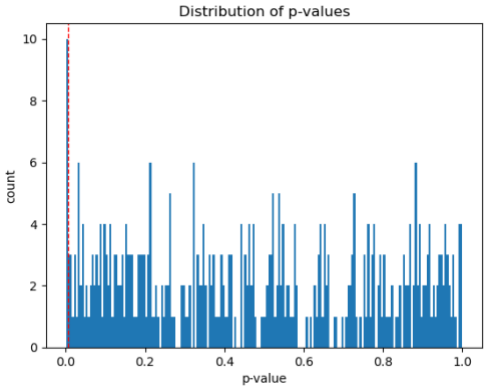
Q8: What proportion of movies exhibit such a "social watching" effect?

D: We did Mann-Whitney U tests for each of the 400 movies under the null hypothesis that there's no significant difference in the ratings of the movie between people who enjoy watching it socially and watching alone.

Y: For each movie, we can't assume the cardinality and normality of each group. For some movies, their sample size is too small to use a t-test, and the samples don't have the same variance. Thus, it's more reliable to use Mann-Whitney U tests.

F: Among 400 movies analyzed, 10 showed statistically significant differences in ratings between two groups at an alpha level of 0.005. The distribution of p-values from the Mann-Whitney U tests is visualized in the histogram on the right:

A: We conclude that approximately 2.5% of movies exhibit a "social watching effect". However, since we have multiple comparisons for all 400 movies, alpha inflation may exist, making the probability of having at least 1 type I error be 86.5%.



Q9: Is the ratings distribution of *'Home Alone (1990)'* different than that of *'Finding Nemo (2003)'*?

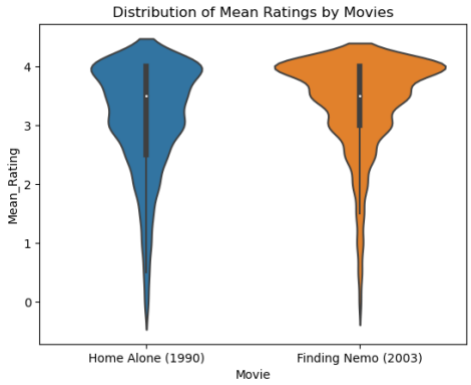
D: Based on the null hypothesis that there's no significant difference between the ratings distribution of the two movies, we did Kolmogorov-Smirnov test, visualize the distribution of the ratings, and find effect size and the statistical power through simulations.

Y: Since we can't assume cardinality and normality of the data, and we are studying the distribution, we use the Kolmogorov-Smirnov test. Visualizing the distribution allows us to intuitively see the difference, and finding effect size and power helps to detect the true difference between the distributions

F: 1. KS test: Statistics: 0.153; P-value: 6.38 * 10^ {-10}

2. Visualization: violin plots indicate a noticeable difference in the rating distributions of the two movies as shown on the right.

3. Power analysis: Cohen's d: 0.303; Power: 0.998. (under 10000 simulations)



A: Given that the p-value is smaller than the significant level, we reject the null hypothesis that there's no significant difference between the rating distribution of the two movies. This is further supported by the different shapes in the violin map, and the high power of 0.998.

Q10: 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?

D: We conducted significance tests for each of the franchises under the null hypothesis that there are no significant differences between the quality of its movies. For franchises with exactly 2 movies, we did Mann-Whitney U tests. For franchises with more than 2 movies, we performed Kruskal-Wallis H tests.

Y: Since we can't assume cardinality and normality of the data, we use the non-parametric Mann-Whitney U test for 2 samples, and the Kruskal-Wallis H test for more than 2 samples.

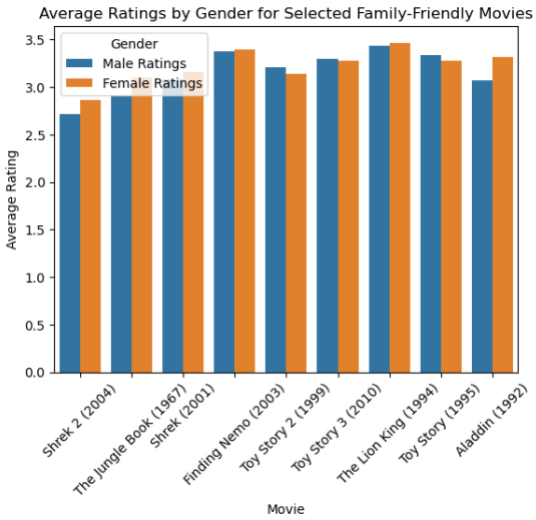
F: 7 out of 8 franchises have a p-value less than the significant level. Those franchises are *'Star Wars'*, *'The Matrix'*, *'Indiana Jones'*, *'Jurassic Park'*, *'Pirates of the Caribbean'*, *'Toy Story'*, and *'Batman'*.

A: Among the 8 franchises, 87.5% of the franchises have inconsistent quality as experienced by viewers. However, since we have multiple comparisons for 8 franchises, there's potential alpha inflation, which means that the probability of having at least 1 type I error is 4% though it's not high.

Extra Credit: Is there a significant gender-based difference in ratings for selected family-friendly movies?

D: 1. Filtered the movie dataset for titles containing keywords indicating family-friendly animated movies (e.g., "Toy Story," "The Lion King") and was able to retrieve 9 movies. 2. Separated ratings by gender for each movie and used the Mann-Whitney U test to assess differences. 3. Calculated the proportion of movies with statistically significant gender-based rating differences.

Y: The H0 is “There is no difference in ratings for family-friendly movies between male and female viewers.” The Mann-Whitney U test was used to compare ratings due to the ordinal nature of the data.



F: The proportion of selected family-friendly movies with significant gender-based rating differences is 0.1111. Out of 9 movies, 1 movie turned out to be a significant movie with a p-value of ratings less than alpha.

A: There is a gender-based difference in ratings for family-friendly animated movies, with 11.11% of the selected titles showing statistically significant differences in enjoyment between male and female viewers. This finding may change if we have a big sample data of family-friendly movies.