Data Analysis Project 1
Kristin Mullaney

1. **Are movies that are more popular also rated higher?**

We can conclude that movies that are more popular are also rated higher.

To answer this question, I split the population into two groups based on the number of ratings for each movie. The "Less Popular" group contained data from the 200 movies with that had fewer ratings. The "More Popular" group contained data from the 200 movies with more ratings. I then ran one-tailed Mann Whitney U tests. I ran "one-tailed" tests because the question is asking if more popular movies are rated "higher" as opposed to if they are "different".

In the first test I compared the median rating for the movies in each group. In the second test I compared the mean rating for the movies in each group. The results of the two tests are shown in the table below. In each instance, the p-value is below the alpha level of 0.005. Consequently, in each test we reject the null hypothesis and accept the alternative hypothesis that more popular movies are rated higher than less popular movies.

Results of Man Whitney U tests				
Test Type	**Less Popular**	**More Popular**	**p-Value**	**Conclusion**
Median	2.5000	3.0000	<0.0001	Reject the null hypothesis
Mean	2.4020	2.8973	<0.0001	Reject the null hypothesis

2. **Are movies that are newer rated differently than movies that are older?**

We cannot conclude that movies that are newer are rated differently than movies that are older.

To answer this question, I determined that the median movie "release" year was 1999. I split the population into groups based on the year that the movie was released. The data set did not have a release year for "Rambo: First Blood Part II". The release year was 1985. I amended the data to include this. The "Older" group contained 197 movies that were released in 1998 or earlier. The "Newer" group contained 203 movies that were released in 1999 or later. I was unable to break the data into two even sized groups because the median year was 1999 and there were 29 movies in the data set that were released that year.

Year of Release	Count
1998 and prior	197
2000 and later	203
Total	400

I then ran a two-tailed Mann Whitney U test where I compared the median rating for the movies in each group. The Mann Whitney U test is still applicable because it does not require the sample sizes to be the same. I ran a two-tailed test because the question asks if males and females rate the movies "differently". The resulting p-value was 0.0887. The p-value is well above the alpha level of 0.005. Consequently, we failed to reject the null hypothesis.

To further confirm this, I ran a two-tailed Kolmogorov-Smirnov tests to compare the distributions of each group. The resulting p-value was 0.3067. The p-value is well above the alpha level of 0.005. The Kolmogorov-Smirnov results confirm the results of the Mann Whitney U tests. Consequently, we cannot reject the null hypothesis and cannot conclude that newer movies are rated differently than older movies.

3. Is enjoyment of 'Shrek' gendered?

We cannot conclude that enjoyment of 'Shrek' is gendered.

To answer this question, I first separated the ratings for Shrek into two groups based upon gender. I removed rows from the data set when: 1) the respondent either did not identify their gender or responded as "self-described", or 2) there was no rating given. This reduced the data set from 1,097 respondents down to 984. In particular, there were 743 female respondents and 241 male respondents. As I stated above, the Mann Whitney U test does not require the sample sizes to be the same and is still applicable. I ran a two-tailed Mann Whitney U test to compare the medians. I ran a two-tailed test because the question asks if males and females rate the movie "differently". The null hypothesis is that the median rating for males is the same as the median rating for females. The resulting p-value was 0.0505. This is significantly greater than the alpha level of 0.005. Consequently, I did not reject the null hypothesis.

Additionally, I ran a two-tailed Kolmogorov-Smirnov test to see if the distributions were similar. The Kolmogorov-Smirnov test also does not require the samples to be the same size. For the Kolmogorov-Smirnov test the null hypothesis is that the two samples have the same distribution. The resulting p-value was 0.3067. This is also significantly higher than the alpha level of 0.005. Therefore, I am again unable to reject the null hypothesis.

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

12.5% of the 400 movies are rated differently by male and female viewers.

This question is similar to question three except that it applies to all movies in the data set as opposed to just Shrek. The underlying methodology here is the same as in question three. After cleaning the data, I ran a two-tailed Mann Whitney U test to compare the ratings from females and males for each movie. Here, the null hypothesis is that the two samples have the same median rating. If the resulting p-value was equal to or below the alpha level of 0.005, I rejected the null hypothesis for that movie. If the resulting p-value was greater than the alpha level, I then ran a one-tailed Kolmogorov-Smirnov test to compare the distributions. Here, the null hypothesis is that the two samples have the same distribution. If the resulting p-value from Kolmogorov-Smirnov test was equal to or below the alpha level of 0.005, I rejected the null hypothesis for that movie. In total there were 50 movies out of 400, or a proportion of 12.5%, where we rejected the null hypothesis in either the Mann Whitney U or Kolmogorov-Smirnov test. In those instances, we accept the alternative hypothesis and conclude that male and females rated those movies differently.

5. Do people who are only children enjoy 'The Lion King' more than people with siblings

We cannot conclude that 'only children' enjoy the 'Lion King' more than people with siblings.

To answer this question, I first separated the ratings for The Lion King into two groups based upon whether the respondent was an only child or had siblings. I removed rows from the data set when: 1) the respondent did not respond to the question asking if they were an only child, or 2) there was no rating given. This reduced the data set from 1,097 respondents down to 927. In particular, there were 151 respondents who were an "only child" and 776 respondents who had "siblings". As I stated above, the Mann Whitney U test does not require the sample sizes to be the same and is still applicable. I then ran a one-tailed Mann Whitney U test to compare the medians. I ran a one-tailed test because the question being asked is if one group enjoys the movie "more" than another group, as opposed to if they rated the movie "differently". The null hypothesis is that the median rating for 'only children' is the same as the median rating for people with siblings. The resulting p-value was 0.9784. This is significantly greater than α of 0.005. Since I could not reject the null hypothesis based on the Mann Whitney U test, I ran a one-tailed Kolmogorov-Smirnov test to see if the distributions were similar. Here, the null hypothesis is that the two samples have the same distribution. The resulting p-value was 0.077. This is also significantly higher than α of 0.005. Therefore, I am again unable to reject the null hypothesis.

6. What proportion of movies exhibit an "only child" effect?

Eight of the 400 movies, or 2.0% of the movies, were rated differently by "only children" than people with "siblings".

This question is similar to question five except that it applies to all movies in the data set as opposed to just The Lion King. Additionally, this question asks whether only children rate movies "differently". As a result, I used a two-tailed Mann Whitney U test for this question and a two-tailed Kolmogorov-Smirnov test. Otherwise, the underlying methodology here is the same as in question five. After cleaning the data, I ran a two-tailed Mann Whitney U test to compare the ratings from "only children" and "siblings" for each movie. Here, the null hypothesis is that the two samples have the same median rating. If the resulting p-value was equal to or below the alpha level of 0.005, I rejected the null hypothesis for that movie. If the resulting p-value was greater than the alpha level, I then ran a two-tailed Kolmogorov-Smirnov test to compare the distributions. Here, the null hypothesis is that the two samples have the same distribution. If the resulting p-value from Kolmogorov-Smirnov test was equal to or below the alpha level of 0.005, I rejected the null hypothesis for that movie. In total there were 8 movies out of 400, or a proportion of 2.0%, where we rejected the null hypothesis in either the Mann Whitney U or Kolmogorov-Smirnov test. In those instances, we accept the alternative hypothesis and conclude that "only children" rated the movies differently than people with "siblings".

7. **Do people who like to watch movies socially enjoy 'The Wolf of Wall Street' more than those who prefer to watch them alone?**

We cannot conclude that people who watch movies socially enjoy the 'The Wolf of Wall Street' more than people who enjoy watching movies alone.

To answer this question, I first separated the ratings for The Wolf of Wall Street into two groups based upon whether the respondent preferred to watch movies socially or alone. I removed rows from the data set when: 1) the respondent did not respond to the question asking if movies are best enjoyed alone, or 2) there was no rating given. This reduced the data set from 1,097 respondents down to 663. In particular, there were 393 respondents who were an enjoyed movies more "alone" and 270 respondents who enjoyed movies more "socially". As I stated above, the Mann Whitney U test does not require the sample sizes to be the same and is still applicable. I then ran a one-tailed Mann Whitney U test to compare the medians. I ran a one-tailed test because the question being asked is if one group enjoys the movie "more" than another group, as opposed to if they rated the movie "differently". The null hypothesis is that the median rating for the "alone" group is the same as the median rating for "socially" group. The resulting p-value was 0.9437. This is significantly greater than α of 0.005. Since I could not reject the null hypothesis based on the Mann Whitney U test, I ran a one-tailed Kolmogorov-Smirnov test to see if the distributions were similar. Here, the null hypothesis is that the two samples have the same distribution. The resulting p-value was 0.2533. This is also significantly higher than α of 0.005. Therefore, I am again unable to reject the null hypothesis.

8. **What proportion of movies exhibit a "social watching" effect?**

Seven of the 400 movies, or 1.75% of the movies, exhibit a "social watching" effect.

This question is similar to question seven except that it applies to all movies in the data set as opposed to just The Wolf of Wall Street. Otherwise, the underlying methodology here is the same as in question seven. After cleaning the data, I ran a one-tailed Mann Whitney U test to compare the ratings from people who like watching movies “alone” to those who like watching movies “socially”. Here, the null hypothesis is that the two samples have the same median rating. If the resulting p-value was equal to or below the alpha level of 0.005, I rejected the null hypothesis for that movie. If the resulting p-value was greater than the alpha level, I then ran a one-tailed Kolmogorov-Smirnov test to compare the distributions. Here, the null hypothesis is that the two samples have the same distribution. If the resulting p-value from Kolmogorov-Smirnov test was equal to or below the alpha level of 0.005, I rejected the null hypothesis for that movie. In total there were 7 movies out of 400, or a proportion of 1.75%, where we rejected the null hypothesis in either the Mann Whitney U or Kolmogorov-Smirnov test. In those instances, we accept the alternative hypothesis and conclude that people who like to watch movies “socially” enjoyed the particular movie more than those people who like to watch movies “alone”.

9. Is the ratings distribution of ‘Home Alone’ different than that of ‘Finding Nemo’?

We can conclude that the ratings distribution for ‘Home Alone’ is different from that of ‘Finding Nemo’.

To answer this question, I ran a two-tailed Kolmogorov-Smirnov test to see if the distributions were similar. I ran a two-tailed test because the question asks if the movies ratings distributions are “different”. Consequently, I did not need to run a Mann Whitey U test for this question. The null hypothesis is that the two samples have the same distribution. The resulting p-value was 0.00000. This is lower than α of 0.005. Therefore, I am able to reject the null hypothesis.

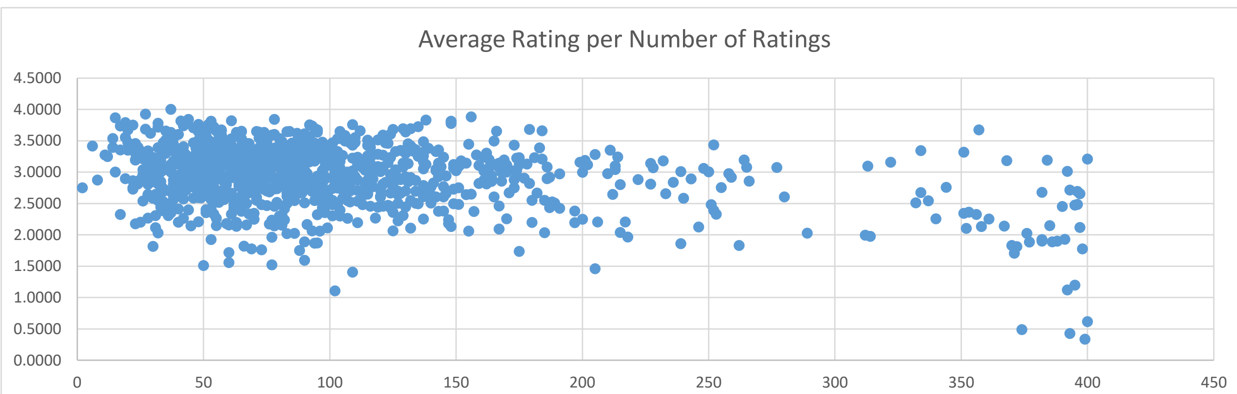
10. How many franchise movies are of inconsistent quality?

There are eight different movie franchises that are being evaluated for this question. All eight have more than two movies in the franchise. Consequently, I used a Kruskal-Wallis test on each franchise to determine if the particular movies were of inconsistent quality. The null hypothesis is that each of the movies in the particular franchise are of consistent quality. Seven of the eight movie franchises had p-values that were below the alpha level of 0.005. For those seven franchises, we are able to reject the null hypothesis and accept the alternative hypothesis that the movies in the franchise are of inconsistent quality. The Harry Potter franchise is the only franchise where we did not reject the null hypothesis. We cannot conclude that the Harry Potter movies are of inconsistent quality.

Franchise	Number of Movies	p-value	Conclusion
Star Wars	5	<0.0001	Reject the null hypothesis
Harry Potter	4	0.343	Do not reject the null hypothesis
The Matrix	3	<0.0001	Reject the null hypothesis
Indiana Jones	4	<0.0001	Reject the null hypothesis
Jurrasic Park	3	<0.0001	Reject the null hypothesis
Prates of the Caribbean	3	<0.0001	Reject the null hypothesis
Toy Story	3	<0.0001	Reject the null hypothesis
Batman	3	<0.0001	Reject the null hypothesis

Extra Credit

I was wondering if there was a relationship between the number of ratings submitted by a participant and how they rate movies. For example, does someone who rates more movies, enjoy movies more and tends to give higher ratings? To explore this question, I determined the total number of ratings per respondent and their average rating. The graph below summaries this data.



There were 1,097 participants in the study. One participant did not submit any ratings and was excluded.

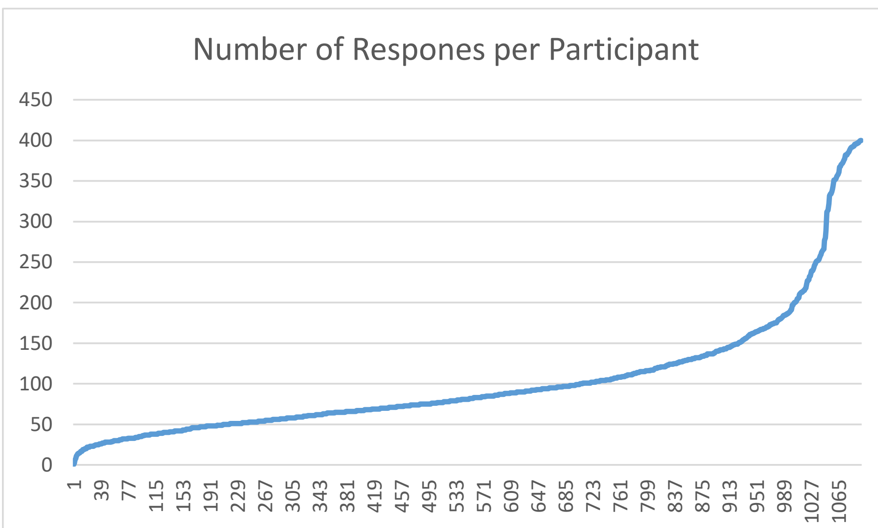
I made a few superficial observations from looking at the graph. First, there appeared to be a lower average rating as the number of movies rated increased. Second, their appeared to be some participants that were relative outliers in how low their average rating was. In particular, I noted that their appeared to be an anomaly in the data with participants who had rated the most movies. The most movies rated by any participant was 400 movies. The most extreme outliers were in instances where the participant had rated over 350 movies. This made me wonder what type of influence this might have on our hypothesis testing. Due to the fact that we are using ratings data, Mann Whitney U and Kolmogorov-Smirnov will be important tests that we will frequently rely on. The p-value in both tests is influenced by the sample size. As the sample size goes up, the p-value will get lower (all other things equal). So I wondered whether these participants with a high volume of ratings might have an outsized effect on the results of our testing. For example, there were only 39 respondents that rated over 350 movies (3.55% of respondents). However, this group submitted 14,827 of the overall 112,214 ratings (13.21% of ratings).

To test this question, I first did a median split of the participants based on the number of ratings. I then did a Mann Whitney U test to see if there was a difference in how the groups rated the movies. The results of the test are shown in the table below.

	Number of Participants	Number of Ratings	Average Rating
Fewer Ratings	548	29,516	3.0353
More Ratings	548	82,698	2.8762
Total	1,096	112,214	2.9558
		Mann Whitney U p-value	0.0012

The low p-value indicates a strong probability that there is a difference in how the groups rate movies. The group where participants rate more movies, rates the movies lower.

The graph below shows the number of ratings per participant in rank order based on number of responses.



The graph has a relatively consistent slope and then increases rapidly. I arbitrarily reduced the population to participants that had less than 150 responses. (This visually appeared to be a point where the slope clearly increased.) I then split the population into two groups based on the number of ratings per participant. The "fewer ratings" group consisted of participants that rated less than 75 movies. The "more ratings" group consisted of participants that rated 75 or more movies but less than 150. I then performed a Mann Whitney U test to see if there was a difference in how the groups rated the movies. The table below details the results.

	Number of Participants	Number of Ratings	Average Rating
less than75 ratings	481	24,314	3.0381
75 to 149	443	45,933	2.9784
Total	1,096	70,247	2.9558
		Mann Whitney U p-value	0.0616

With a p-value of 0.0616 we would accept the null hypothesis that the two samples have the same mean.

These tests support the hypothesis that the average rating per participant is similar for lower numbers of ratings per participant but the average ratings declines when the number of ratings for a participant becomes very high. Furthermore, the standard deviation of the ratings is significantly higher when the number of ratings exceeds 350 ratings. The standard deviation is 0.44 when the number of ratings is less than 350. The standard deviation is 0.78 when the number of ratings is above 350. This raises interesting questions about why this would occur and about how this may influence our analysis.