Q1

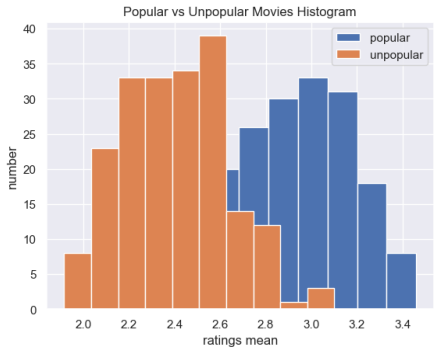
Are movies that are more popular (operationalized as having more ratings) rated higher than movies that are less popular? [Hint: You can do a median-split of popularity to determine high vs. low popularity movies]

Separate Popular and Unpopular Movies

First we calculate the median of valid rating numbers for all 400 movies, which is 197.5. For movies with number of valid ratings > 197.5, we put it into the popular movie category. For movies with number of valid ratings < 197.5, we put it into the unpopular movie category. After this separation, we get 200 popular movies and 200 unpopular movies.

Method1: T Test on Mean

For each movie, we use the mean of its valid rating to represents how audience perceive it and then append this number to either popular or unpopular category rating list (elementwise removal of missing data). Without knowing the population parameters and assuming similar variance, we makes a null hypothesis that the unpopular movies and popular movies have same mean (the means for the two populations are equal). Then we perform independent t test for above 2 groups on their rating mean value. The result t statistic is 17.756 and p value is 2.27e-39, which is less than 0.005 alpha level. This gives a significant result. Hence, we conclude that popular rated higher than movies that are less popular movies. This result can also be intuitively seen by below histogram. As we can see, popular movie's mean rating on higher on average.



However, this method has some limitations. First, it is dubious to reduce each movie to its mean. Also, we gonna lose lots of information with this method. Hence, we apply below method to try to fix above problem.

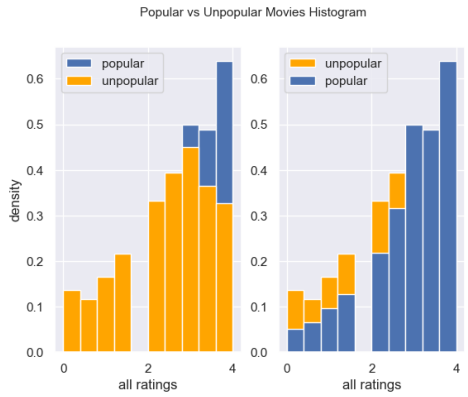
Method 2: Chi Sqaure Test on All Data

In this method, we are trying to lose as much information as possible. We store ratings of all of the popular movies in one list, and all ratings of unpopular movies in another list and try to compare them directly. In this way, we don't have to reduce each movie into a single mean or median, and hence preserve more information of each category.

After this operation, we get a list of popular movie ratings with 90214 ratings, which contains all the rating from popular movies, and a list of unpopular movies ratings with 22000 ratings.

To perform Chi square test, we treat each rating level as one category, where we have 9 categories in total: {4, 3.5, 3, 2.5, 2, 1.5, 1, 0.5, 0}. Our null hypothesis is: the frequency in each rating category are proportionately the same in unpopular movies(observed) as they are in popular movies (expected). Besides, as we noticed, the sum of popular movie frequency is different from the sum of unpopular movie frequency. To make their sum agree with each other, we decide to shrink the number in each category of popular movie rating with same ratio. Specifically, originally there are {1830, 2362, 3459, 4591, 7872, 11386, 18032, 17642, 23040} observed number in popular movies, corresponding to the category we mentioned above. It has a sum of 90214, which is different from 22000 of unpopular movies. So we multiply number in each category with $N_{\text{unpopular rating}}/N_{\text{popular rating}}$ and get a modified frequency {446.27, 576.01, 843.53, 1119.58, 1919.7 , 2776.64, 4397.37, 4302.26} for popular movies.

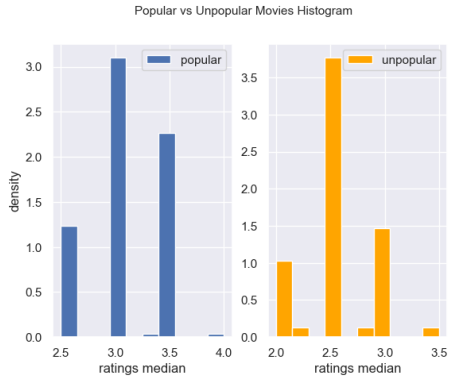
Finally we perform our chi square test and get p value equal to 0, which leads to a significant result. Hence, we overthrow the conclusion that popular and popular movies are rated similarly. This result can also be intuitively seen by below graph, where we plot the the histogram of all the rating of 2 categories. As we can see, their distribution is quite different.



However, this method also has some hidden danger. There is no strong theoratical support on the step we modify the data to make sum of the observed frequencies agree with the sum of the expected frequencies. This method is more of expeimantal.

Method3: U test on median

For each of the movie, we use the median of its valid rating to represents its rating (elementwise removal of missing data). We makes a null hypothesis that the unpopular movies ratings have same population with popular movies. Then we perform independent U test for above 2 groups on their rating mean value. The result t statistic is 33427.5 and p value is 9.93e-35, which is less than 0.005 alpha level. This gives a significant result. Hence, I conclude that popular rated higher than movies that are less popular movies.



However, this method is also dubious when we reduce each movie to its median. Also, we still gonna lose lots of information with this method.

quesion 2

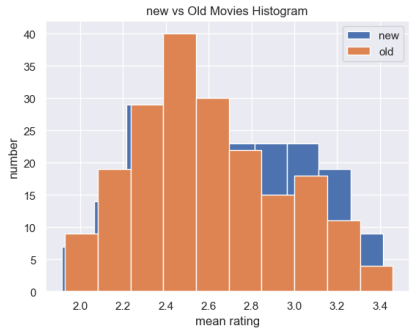
Are movies that are newer rated differently than movies that are older? [Hint: Do a median split of year of release to contrast movies in terms of whether they are old or new]

Separate New and Old Movies

First we calculate the median median movie issue year for all 400 movies, which is 1999. For movies issued before 1999, we put it into the old movie category. For movies that issued after or at 1999, we put it into the new movie category. After this separation, we get 203 new movies and 197 old movies.

Method1: T Test on Mean

For each movie, we use the mean of its valid rating to represents how audience perceive it and then append this number to either new or old movie category rating list (elementwise removal of missing data). Without knowing the population parameters and assuming similar variance, we makes a null hypothesis that the new movies and old movies have same rating mean (the means for the two populations are equal). I perform independent t test for above 2 groups on their rating mean value. The result t statistic is 1.61 and p value is 0.109, which is greater than 0.005 alpha level. This gives a not significant result. Hence, we maintain the null hypothesis and conclude that newer rated similarly with movies that are older. This result can also be intuitively seen by below histogram. As we can see, there is a huge overlap between histogram of mean rating of new movies and mean rating of old movies.



However, this method has some limitations. First, it is dubious to reduce each movie to its mean. Also, we gonna lose lots of information with this method. Hence, we apply below method to try to fix above problem.

Method2: U test on median

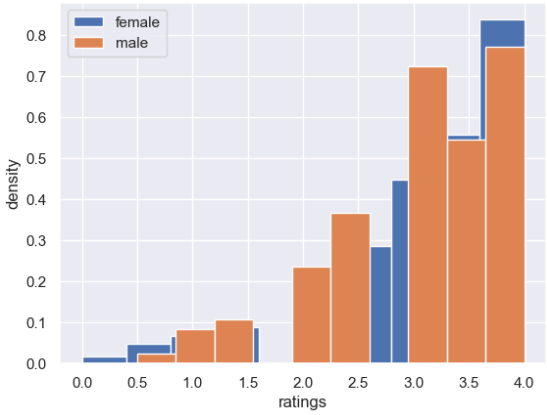
For each of the movie, we use the median of its valid rating to represents its rating (elementwise removal of missing data). We makes a null hypothesis that the old movies ratings have same population with new movies. Then we perform independent U test for above 2 groups on their rating mean value. The result t statistic is 18127.5 and p value is 0.089, which is grater than 0.005 alpha level, which is not significant. Hence, I conclude that movies that are newer rated similarly with movies that are older.

However, this method is also dubious when we reduce each movie to its median. Also, we still gonna lose lots of information with this method.

quesion 3

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

For each of the valid rating of Shrek (2001), if its corresponding gender identity is 1, I put it into the famale rating category. If its corresponding gender identity is 2, I put it into the male rating category. In the end we have 743 valid female ratings and 241 valid male ratings. Here mean is not a reasonable statistics for a movie, hence we decide to perform Mann-Whitney U test. Our null hypothesis that male's rating on Shrek (2001) have same population with female's rating.The u value is 82232.5 and p value is 0.05, which is greater than 0.005 alpha level. This result is not significant. Hence, we maintain our null hypothesis and conclude that male and female viewers rate Shrek (2001) similarly. This result can also be intuitively seen by plotting the histogram of mean rating of 2 categories. As we can see from below graph, there is no big different between 2 categories on the density of distribution of ratings.



quesion 4

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

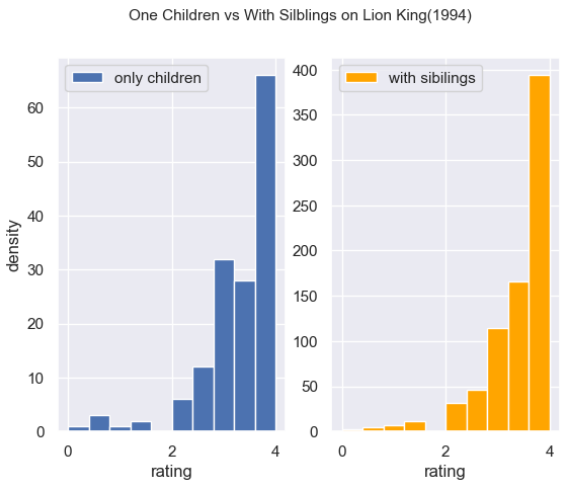
For each movie, we record the an list of valid male ratings and a list of female ratings, as what we did in question 3. With median as a reasonable choice for summary statistics, then I performed Mann-Whitney U test on the male rating list and female rating list for each movie. Our null hypothesis that male's rating on this movie have same population with female's rating. If the

corresponding p value is less than chosen alpha level 0.005, I count this movie as a different rating one. Otherwise, I count this movie as a similar rating one. After traversing through all the movie, I find that there are 50 movies in the different rating category and 350 movies in the similar rating category. Hence we have 50/(50+350) = 12.5% movies that are rated differently by male and female viewers.

quesion 5

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

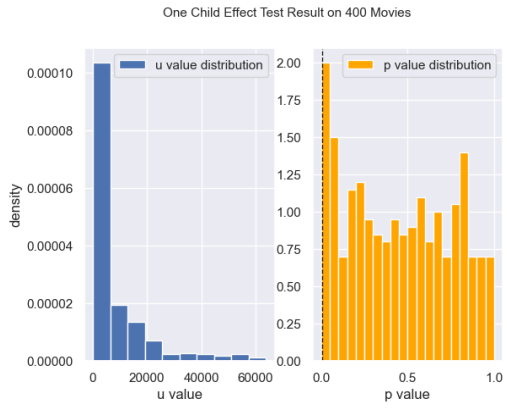
For each of the valid rating of The Lion King (1994), if its corresponding answer on Are you an only child is 1, I put it into the only children rating category. If its corresponding answer on Are you an only child is 0, I put it into the with siblings rating category. In the end we have 151 valid recordings in only children category and 776 valid recordings on with siblings rating category. Here mean is not a reasonable statistics for a movie, hence we decide to perform Mann-Whitney U test. Our null hypothesis that people who are the only child's rating on The Lion King (1994) have same population with people with siblings's rating. The resulted u value is 52929.0 and p value is 0.043, which is greater than 0.005 alpha level. This result is not significant. Hence, we maintain our null hypothesis and conclude that that people who are the only children and people with siblings enjoy The Lion King (1994) at similar level . This result can also be intuitively seen by plotting the histogram of mean rating of 2 categories. As we can see from below graph, there is no big different between 2 categories on the density of distribution of ratings.



quesion 6

What proportion of movies exhibit an “only child effect”, i.e. are rated different by viewers with siblings vs. those without?

For each movie, I record the an list of valid with sibling ratings and a list of without sibling ratings, as what we did in question 5. With median as a reasonable choice for summary statistics, then I performed Mann-Whitney U test. Our null hypothesis that people who are the only child’s rating on current movie have same population with people with siblings’s rating. If the corresponding p value is less than chosen alpha level 0.005, I count this movie as a different rating one. Otherwise, I count this movie as a similar rating one. After traversing through all the movie, I find that there are 7 movies in the different rating category and 350 movies in the similar rating category. Hence we only have 7/400 = 1.75% movies that exhibit an “only child effect”. We can also see the result intuitively when we plot the histogram of u value and p value for each movie. As we can see, most of the p values are > 0.005.

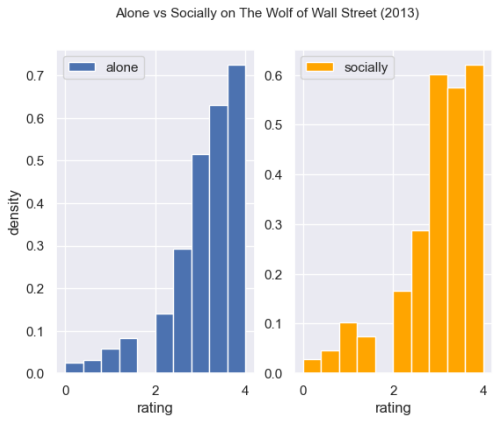


quesion 7

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

For each of the valid rating of The Lion King (1994), if its corresponding answer on movies are best enjoyed alone is 1, I put it into the prefer alone rating category. If its corresponding answer on movies are best enjoyed alone is 0, I put it into the prefer socially rating category. In the end we have 393 valid recordings in prefer alone category and 270 valid recordings on prefer socially category. Here mean is not a reasonable statistics for a movie, hence we decide to

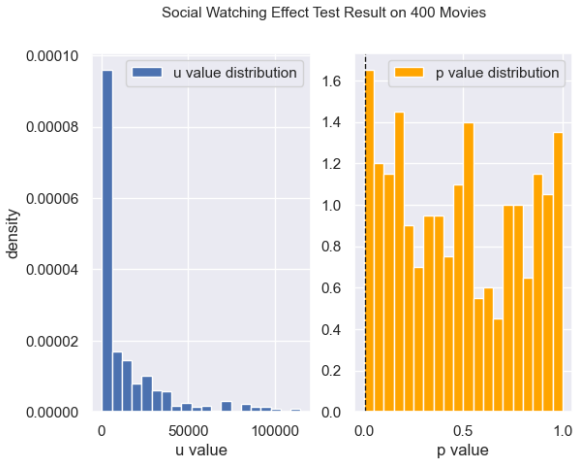
perform Mann-Whitney U test. Our null hypothesis that rating from people who like to watch movies socially on The Wolf of Wall Street (2013) have same population with rating from people who prefer to watch them alone. The resulted u value is 56806.5 and p value is 0.11, which is greater than 0.005 alpha level. This result is not significant. Hence, we maintain our null hypothesis and conclude that people who prefer watching alone and people who prefer watch watching movie socially enjoy The Wolf of Wall Street at similar level. This result can also be intuitively seen by plotting the histogram of mean rating of 2 categories. As we can see from below graph, there is no big different between 2 categories on the density of distribution of ratings.



quesion 8

What proportion of movies exhibit such a “social watching” effect?

For each movies, I record the an list of preder alone ratings and a list of prefer socially ratings, as what we did in question 7. With median as a reasonable choice for summary statistics, then I performed Mann-Whitney U test on preder alone rating list and prefer socially list for each movie. Our null hypothesis that rating from people who like to watch movies socially on current movie have same population with rating from people who prefer to watch them alone. If the corresponding p value is less than chosen alpha level 0.005, I count this movie as a different rating one. Otherwise, I count this movie as a similar rating one. After traversing through all the movie, I find that there are 10 movies in the different rating category and 390 movies in the similar rating category. Hence we only have 10/400 = 2.5% movies that exhibit an “social watching”. We can also see the result intuitively when we plot the histogram of u value and p value for each movie. As we can see, most of the p values are > 0.005.

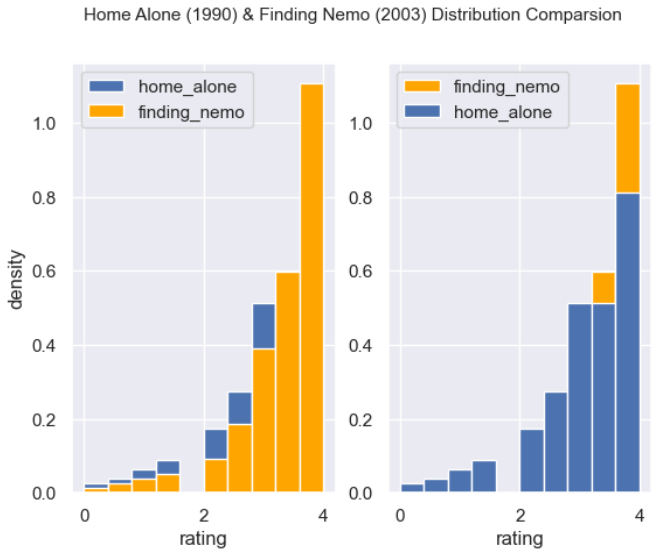


quesion 9

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

Here I perform row wise pruning for each of the movie to delete the nan. Hence to make we are comparing apple to apple. After this operation we are lefting with 810 piece of rating for each of the movie. To perform Chi Square test, we treat each rating level as one category, where we have 9 categories in total: {4, 3.5, 3, 2.5, 2, 1.5, 1, 0.5, 0}. Our null hypothesis is: the frequency in each rating category are proportionately the same in Home Alone (1990) (observed frequency as they are in Finding Nemo (2003) (expected frequency).

Finally we perform our chi square test on the two rating list and get p value equal to 5.60e-24, which leads to a significant result. Hence, We overthrow the null hypothesis and we can conclud that distribution of 'Home Alone (1990)' different than that of 'Finding Nemo (2003)'. We can also see this difference woith below histogram graph. As we can see, their rating distribution are quite different.



quesion 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? [Hint: You can use the keywords in quotation marks featured in this question to identify the movies that are part of each franchise]

For each of above franchises, we perform row-wise pruning to make sure that we are comparing apple to apple. Since we are trying to investigate testing whether multiple samples are originated from the same distribution, we apply kuskal wallis test on each of above franchises. Our null hypothesis is that the mean ranks of the groups are the same.

The results are showed as below:

Frenchsise Name	Star Wars	Harry Potter	The Matrix	Indiana Jones	Jurassic Park	Pirates of the Caribbean	Toy Story	Batman
stats	193.51	5.87	40.32	54.19	49.43	6.66	23.50	84.65
p value	6.94e-40	0.1179	1.75e-09	1.02e-11	1.85e-11	0.035	7.90e-6	4.14e-19
significant	yes	no	yes	yes	yes	no	yes	yes

As we can see, only Harry Poter and Pirates of the Caribbean has no significant result. For them, we maintain our null hypothesis and conclude that all movies rating from that franchise are originated from the same distribution. And we conclude that they have consistent quality. For all other movies, we reject the null hypothesis and conclude that they are of inconsistent quality.

Extra Credit

Tell us something interesting and true (supported by a significance test of some kind) about the movies in this dataset that is not already covered by the questions above [for 5% of the grade score].

Here we want to investigate what proportion of movie has an ‘energy level effect’, that people think they are full of energy rate movies higher than people think they are lack of energy.

Given a movie, for each valid rating, if its corresponding answer on Is full of energy is 4 or 5, I put it into the full of energy category. If its corresponding answer on Is full of energy is 1 or 2, I put it into the lack of energy category.

For each movie, I get a record of rating of above two categories. With median as a reasonable choice for summary statistics, I performed Mann-Whitney U test on the full of energtic rating list and lack of energtic rating list. Our null hypothesis that rating from people who are energetic on current movie have same population with rating from people who are lack of energy. If the corresponding p value is less than chosen alpha level 0.005, I count this movie as a different rating one. Otherwise, I count this movie as a similar rating one. After traversing through all the movie, I find that there are 11 movies in the different rating category and 389 movies in the similar rating category. Hence we only have 11/400 = 2.75% movies that exhibit an ‘energy level effect’.