Data Analysis Project (Hypothesis Testing)
Ansh Riyal (ar7964)

Common Assumptions (D):
1. Mann-Whitney test:
The ratings observations for 2 groups being compared are treated as ordinal and mutually independent.
Null Hypothesis(H_0) is that the 2 groups are from same distribution (same median), and the **Alternate Hypothesis(H_1)** is that median of Group 1 is higher than median of Group 2 (for one sided test) and different than median of Group 2 (for two sided)

2. Kolmogorov Smirnov test
The ratings observations for 2 groups being compared are treated as ordinal and mutually independent.
Null Hypothesis(H_0) is that the 2 groups are from same distribution, and the **Alternate Hypothesis(H_1)** is that distribution of Group 1 is different than different of Group 2 (for two sided)

Treatment:
1. We first use the question condition in order to split our dataset into 2 groups (old vs new, only_child vs siblings, etc.)
2. Then, for Ques 1-2, we calculate the column wise median (the median of ratings of every movie in both groups). While doing this, we assume that the median of ratings of a movie are a good representation of that movie.
3. Then, we clean the dataset by dropping the nan and null values in the dataset in order to use significant tests (since scikitlearn treats nan values as infinity)
4. Then, we have 2 distributions: Group 1: (median rating of movie 1 in G1, median rating of movie 2 in G1, ..., median rating of movie n in G1) & Group 2: (median rating of movie 1 in G2, median rating of movie 2 in G2, ..., median rating of movie n in G2).
5. For Ques 3-10, we do not calculate column wise medians and simply use the values we have in our groups.
6. Now we perform the reasonably correct significance test based on our assumptions.

Questions

1. Are movies that are more popular (operationalized as having more ratings) rated higher than movies that are less popular?
Assumptions(D): Assumptions of Mann-Whitney test with Group 1: popular movies and Group 2: unpopular movies. Alternate Hypothesis being defined for a one sided test.
(Y): Using the assumptions mentioned above, we are trying to determine whether the median of the popular movies distribution is higher than the unpopular movies distribution using one-sided Mann Whitney test
Findings(F): Median ratings for H and L were 3.0 and 2.5 with sample sizes 200 and 200 respectively; Mann-Whitney Test results in U-statistic as 6572.5 and the p-value as 9.929 e-35
Answer(A): Since the p-value $< 0.005(\alpha)$, we can conclude with statistical significance that popular movies have a higher median rating than the unpopular movies.
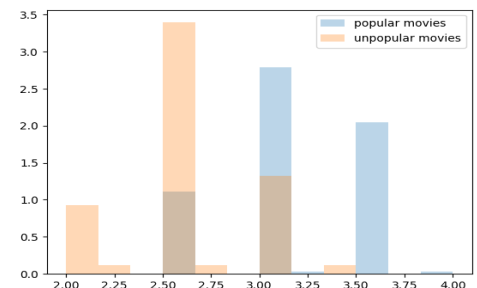

2. Are movies that are newer rated differently than movies that are older?
Assumptions(D): Assumptions of Mann-Whitney with Group 1: median ratings observed for older movies and Group 2: median ratings observed for newer movies. The Alternate Hypothesis being defined for a two sided test.
(Y): Using the assumptions mentioned above, we are trying to determine whether the distribution of the medians of new movies is different than the old movies median distribution using two-sided Mann Whitney test
Findings(F): Median of median ratings for new movies was 3.0 and for old was also 3.0 with sample sizes 203 (new) and 197 (old) respectively; Mann-Whitney Test results in U-statistic as 18127.5 and the p-value as 0.08873
Answer(A): Since the p-value $> 0.005(\alpha)$, we can not conclude with statistical significance that new movies have a different rating distribution than old ones.
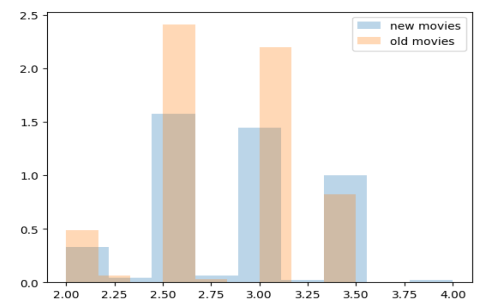
3.   Is enjoyment of 'Shrek (2001)' gendered, i.e., do male and female viewers rate it differently?

Assumptions(D): Assumptions of Mann-Whitney with Group 1: ratings observed for male audience and Group 2: ratings observed for female audience. The Alternate Hypothesis being defined for a two sided test.

(Y): Using the assumptions mentioned above, we are trying to determine whether the distribution of the male audience ratings is different than the female audience ratings using two-sided Mann Whitney test

Findings(F): Median ratings for male audience was 3.0 and for female audience was 3.5 with sample sizes 241 (male) and 743 (female) respectively; Mann-Whitney Test results in U-statistic as 96830.5 and the p-value as 0.0505

Answer(A): Since the p-value > 0.005(α), we can not conclude with statistical significance that ratings of Shrek (2001) are gendered.
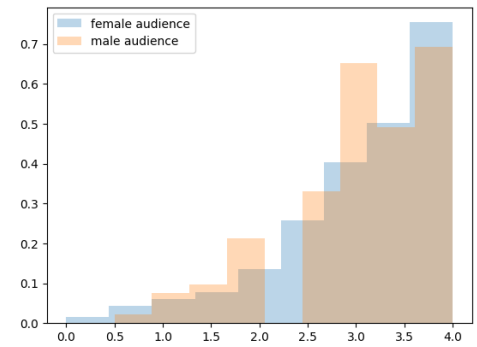
4.   What proportion of movies are rated differently by male and female viewers?

Assumptions(D): Same as Q3 as we are performing the same test mechanism on every movie in the dataset.

(Y): Using the assumptions mentioned above, we are trying to determine in what proportion of movies, is the distribution of male audience ratings is different than the female audience ratings distribution using two-sided Mann Whitney test

Findings(F): Out of 400 movies, 50 movies have p values < 0.005(α).

Answer(A): Using the same statistical significance paradigm as Q3, we conclude that approx. 12.5% of movies show a gendered effect.

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

Assumptions(D): Assumptions of Mann-Whitney with Group 1: ratings observed for only children and Group 2: ratings observed for people with siblings. The Alternate Hypothesis being defined for a one sided test.

(Y): Using the assumptions mentioned above, we are trying to determine whether the rating distribution for only children are higher than the people with siblings median distribution using one-sided Mann Whitney test

Findings(F): Median of ratings for only children was 3.5 and for people with siblings was 4.0 with sample sizes 151 (only) and 776 (siblings) respectively; Mann-Whitney Test results in U-statistic as 64247.0 and the p-value as 0.9784

Answer(A): Since the p-value > 0.005(α), we can not conclude with statistical significance that only children have a higher rating distribution than people with siblings. In fact, if we look at the histogram, we can see that if anything, people with siblings tend to enjoy 'The Lion King (1994) more than only children.
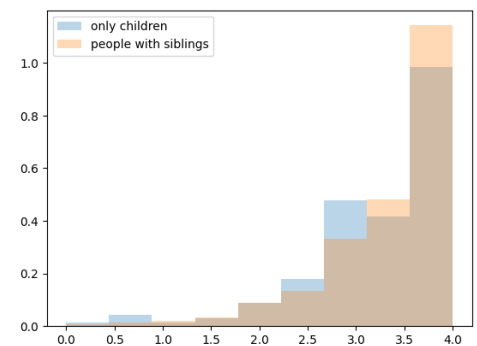
6.   What proportion of movies exhibit an "only child effect", i.e., are rated different by viewers with siblings vs. those without?

Assumptions(D): Same as Q5 as we are performing the same test mechanism on every movie in the dataset.

(Y): Using the assumptions mentioned above, we are trying to determine in what proportion of movies, is the distribution of only children ratings is different than the people with siblings ratings distribution using one-sided Mann Whitney test

Findings(F): Out of 400 movies, 7 movies have p values < 0.005(α).
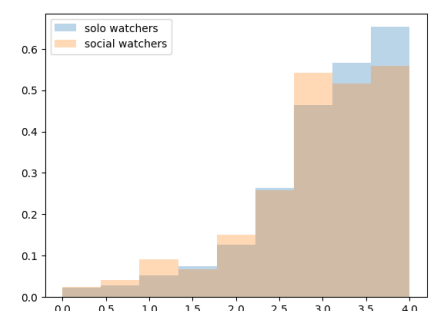
Answer(A): Using the same statistical significance paradigm as Q5, we conclude that approx. 1.75% of movies show a only child effect.

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?

Assumptions(D): Assumptions of Mann-Whitney with Group 1: ratings observed for social watchers and Group 2: ratings observed for solo watchers. The Alternate Hypothesis being defined for a one sided test.

(Y): Using the assumptions mentioned above, we are trying to determine whether the distribution of the ratings of social watchers is higher than the solo watchers rating distribution using one-sided Mann Whitney test

Findings(F): Median of ratings for social watchers was 3.0 and for solo watchers was 3.5 with sample sizes 270(social) and 393 (solo) respectively; Mann-Whitney Test results in U-statistic as 56806.5 and the p-value as 0.9436

Answer(A): Since the p-value > 0.005(α), we can not conclude with statistical significance that new movies have a different rating distribution than old ones.

8.   What proportion of movies exhibit such a "social watching" effect?
Assumptions(D): Same as Q7 as we are performing the same test mechanism on every movie in the dataset.
(Y): Using the assumptions mentioned above, we are trying to determine in what proportion of movies, is the distribution of social watchers ratings is different than the solo watchers ratings distribution using one-sided Mann Whitney test
Findings(F): Out of 400 movies, 6 movies have p values < 0.005(α).
Answer(A): Using the same statistical significance paradigm as Q7, we conclude that approx. 1.5% of movies show a social watching effect.

9.   Is the ratings distribution of 'Home Alone (1990)' different than that of 'Finding Nemo (2003)'?
Assumptions(D): Assumptions of Kolmogorov Smirnov with Group 1: ratings observed for Home Alone and Group 2: ratings observed for Finding Nemo. The Alternate Hypothesis being defined for a two sided test.
(Y): Using the assumptions mentioned above, we are trying to determine whether the distribution of 'Home Alone (1990)' is different from the rating distribution of 'Finding Nemo' using two-sided KS test
Findings(F): Median ratings for Home Alone was 3.5 and for Finding Nemo was also 3.5 with sample sizes 857 (Home Alone) and 1014 (Finding Nemo) respectively; KS Test results in D-statistic as 0.1526 and the p-value as 6.38e-10

Answer(A): Since the p-value < 0.005(α), we can conclude with statistical significance that Home Alone has a different rating distribution than Finding Nemo.
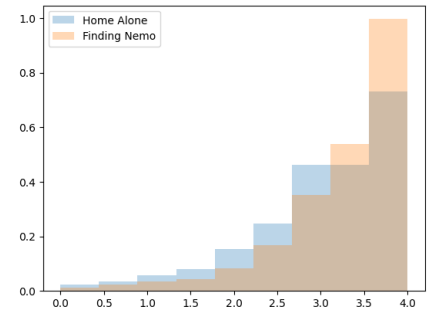
10.   There are ratings on movies from several franchises (['Star Wars', 'Harry Potter', 'The Matrix', 'Indiana Jones', 'Jurassic Park', 'Pirates of the Caribbean', 'Toy Story', 'Batman']) in this dataset. How many of these are of inconsistent quality, as experienced by viewers?
Assumptions(D): The ratings observations for 2 groups being compared are treated as ordinal and mutually independent. Null Hypothesis (H$_0$): All the movies belonging to a franchise have similar ratings (i.e. sampled from the same distribution) Alternate Hypothesis (H$_1$): Movies belonging to a franchise have different underlying distributions.
(Y): Using the assumptions mentioned above, we are trying to determine whether the distribution of the movies belonging to the same franchise have different underlying distribution using Kruskal Wallis test. Since this is a test of dominance, the result is significant even if there is one group with a different distribution than the others.
Findings(F):

Franchise	No. of movies	Test statistic	P value	Conclusion
Star Wars	6	230.58	8.01 e-48	p < 0.005(α) : Significantly different rating distributions
Harry Potter	4	3.33	0.343	p > 0.005(α) : Can not conclude different distributions
The Matrix	3	48.37	3.12 e-11	p < 0.005(α) : Significantly different rating distributions
Indiana Jones	4	45.79	6.27 e-10	p < 0.005(α) : Significantly different rating distributions
Jurassic Park	2	45.17	1.80 e-11	p < 0.005(α) : Significantly different rating distributions
Pirates of the Caribbean	3	20.64	3.29 e-5	p < 0.005(α) : Significantly different rating distributions
Toy Story	3	24.38	5.06 e-6	p < 0.005(α) : Significantly different rating distributions
Batman	3	190.53	4.22 e-42	p < 0.005(α) : Significantly different rating distributions

Answer(A): Since the p-value > 0.005(α) for only 1 out of 8 franchises (Harry Potter), we can conclude with statistical significance that 7 out of 8 franchises are of inconsistent quality (as experienced by viewers).

Extra Credit Questions

(Intuition supported by evidence)

Q1 Do people who watch more movies rate movies differently than the people who watch less movies?

i.e. is there a 'movie experience effect' on the ratings?

Assumptions(D): Assumptions of Kolmogorov Smirnov with Group 1: ratings observed for Frequent movie watchers and Group 2: ratings observed for Rare movie watchers. The Alternate Hypothesis being defined for a two sided test.

(Y): Using the assumptions mentioned above, we are trying to determine whether the rating distribution of rare watchers is different from the rating distribution of frequent watchers using two-sided KS test

Findings(F): Median ratings for frequent watchers was 3.0 and for rare watchers was also 3.0 with sample sizes 385 per group; KS Test results in D-statistic as 0.319 and the p-value as 9.025e-18

Answer(A): Since the p-value $< 0.005(\alpha)$, we can conclude with statistical significance that frequent movie watchers have a different rating distribution than those who watch movies rarely.
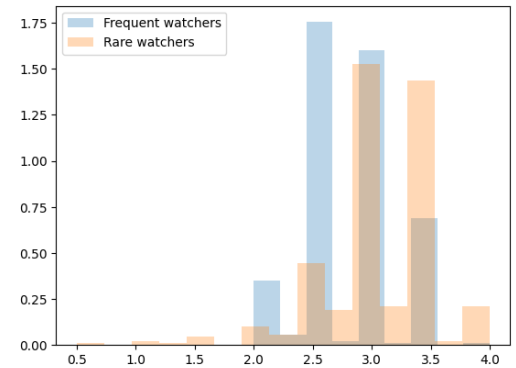
(Intuition not supported by evidence)

Q2 What proportion of movies show a criticism effect?

criticism effect: Do people who tend to find fault with others rate the movie differently from those who do not

Assumptions(D): Assumptions of Mann Whitney with Group 1: ratings observed for critical movie watchers and Group 2: ratings observed for Non critical movie watchers. The Alternate Hypothesis being defined for a two sided test.

(Y): Using the assumptions mentioned above, we are trying to determine in what proportion of movies, is the distribution of critical watchers' ratings is different than the non critical watcher ratings distribution using a two-sided Mann Whitney test

Findings(F): Out of 400 movies, 3 movies have p values $< 0.005(\alpha)$.

Answer(A): Based on our dataset, we conclude that approx. 0.75% of movies show a criticism effect.

Intuitively, one might think that if someone is more critical towards others, then they will probably rate movies more critically i.e. in a different manner. However, given our data and our test, we could only see that for 3 out of 400 movies (given our significance level of 0.005). This might suggest that our Hypothesis is false i.e. critical people rate movies in the same manner as non critical people. However, we would be wrong in drawing that conclusion as the situation here is 'Absence of Evidence' and as we know, 'absence of evidence does not mean evidence of absence'