

# Hypothesis Testing of Movie Ratings Data

Zitong Luo

## Q1 Popularity and Ratings

To distinguish between high and low popularity movies, we first did a median split based on the number of valid ratings. Given the relatively high proportion of missing data, we applied **element-wise removal** of null values. Since the psychological distance between neighboring ratings may not be consistent, and with the alternative hypothesis that more popular movies are rated higher than less popular ones, we employed a **one-sided Mann-Whitney U test**.

We found that the p-value is extremely small (**0.000**), which is less than the significance level of 0.005, indicating that the result is statistically significant. In other words, we have reason to conclude that more popular movies are rated higher than less popular ones.

## Q2 Years of Release and Ratings

Similar to the first one, we first performed a median split based on the year of release and applied **element-wise removal** of null values. With the alternative hypothesis that newer movies are rated differently than older ones, we employed both a **two-sided Mann-Whitney U test** and a **KS test** to emphasize differences in median and distribution.

We found that the p-values for both tests are less than 0.005 (**0.000** and **0.002**, respectively), indicating that the results are statistically significant. In other words, we have reason to conclude that newer movies are rated differently than older movies.

## Q3 Gender Effect: *Shrek (2001)*

Similar to the last two tests, with the alternative hypothesis that male and female viewers rate the movie *Shrek (2001)* differently, we employed both a **two-sided Mann-Whitney U test** and a **KS test** to emphasize differences in median and distribution.

We found that the p-values for both tests are larger than 0.005 (**0.051** and **0.056**, respectively), indicating that the results are not statistically significant. So, we conclude that the enjoyment of *Shrek (2001)* is not gendered.

## Q4 Gender Effect in General

We applied the same tests as in the previous question to all 400 movies and found that the proportion of movies rated differently by male and female viewers is **12.5%** and **6.25%** for the Mann-Whitney U test and KS test, respectively.

In this case, a separate test was conducted for each movie. However, a potential enhancement could be the implementation of a correction, such as False Discovery Rate (FDR), to mitigate the likelihood of false discoveries and reduce the proportion of erroneous findings.

#### **Q5 Only Child Effect: *The Lion King* (1994)**

With the alternative hypothesis that people who are only children enjoy *The Lion King* (1994) more than people with siblings, we employed a **one-sided Mann-Whitney U test**.

We found that the p-value is quite large (**0.978**), which is larger than the significance level of 0.005, indicating that the result is not statistically significant. In other words, we have reason to conclude that people who are only children don't enjoy it more than people with siblings.

#### **Q6 Only Child Effect in General**

With the alternative hypothesis that movies are rated differently by viewers with siblings and those without, we applied both a **two-sided Mann-Whitney U test** and a **KS test** to all 400 movies and found that the proportion of movies rated differently by viewers with siblings and those without is **1.75%** and **0.75%** for the Mann-Whitney U test and KS test, respectively.

Again, a potential enhancement could be the implementation of a correction, such as FDR, to mitigate the likelihood of false discoveries.

#### **Q7 Social Watching Effect: *The Wolf of Wall Street* (2013)**

With the alternative hypothesis that people who like to watch movies socially enjoy *The Wolf of Wall Street* (2013) more than those who prefer to watch them alone, we employed a **one-sided Mann-Whitney U test**.

We found that the p-value is quite large (**0.944**), which is larger than the significance level of 0.005, indicating that the result is not statistically significant. In other words, we have reason to conclude that people who like to watch movies socially don't enjoy this movie more than those who prefer to watch them alone.

#### **Q8 Social Watching Effect in General**

Here, we assume the "social watching" effect refers to "people who like to watch movies socially enjoy them more than those who prefer to watch them alone", so we applied the **one-sided Mann-Whitney U test** to all 400 movies and found that **1.5%** of movies exhibit such a "social watching" effect.

If we applied a two-sided Mann-Whitney U test instead, the result would be larger, yielding a proportion of 2.5%. Again, a potential enhancement could be the implementation of a correction.

**Q9 Ratings Distribution: *Home Alone (1990)* vs. *Finding Nemo (2003)***

To mitigate a potential source of bias, we applied **row-wise removal** of null values. With the emphasis on “ratings distribution”, we only employed a **KS test** in this case.

We found that the p-value is extremely small (**0.000**), which is less than the significance level of 0.005, indicating that the result is statistically significant. So, we have reason to conclude that the ratings distribution of *Home Alone (1990)* is different than that of *Finding Nemo (2003)*.

**Q10 Quality Consistency of Movies from Different Franchises**

To appropriately assess the consistency of movie quality, we employed **row-wise removal** of null values to mitigate a potential source of bias. Considering that all movie series include at least three movies, we used a **Kruskal-Wallis test**.

As shown in the table below, we can conclude that six series (*Star Wars*, *The Matrix*, *Indiana Jones*, *Jurassic Park*, *Toy Story* and *Batman*) are of inconsistent quality.

Table 1 The Kruskal-Wallis Test Results

	<i>Star Wars</i>	<i>Harry Potter</i>	<i>The Matrix</i>	<i>Indiana Jones</i>	<i>Jurassic Park</i>	<i>Pirates of the Caribbean</i>	<i>Toy Story</i>	<i>Batman</i>
p-value	0.000	0.118	0.000	0.000	0.000	0.036	0.000	0.000
<0.005	True	False	True	True	True	False	True	True

**An Extra Finding**

We found that, on average, men tend to have a greater interest in watching movies compared to women, quantified by the number of movies watched. Assuming heteroskedasticity, we used a **one-sided Welch t-test** and found that the p-value is extremely small (**0.000**), which is below the significance level of 0.005, indicating that the result is statistically significant.

However, we also found a large discrepancy in sample sizes (807 females and 260 males), which indicates a potential sampling bias and may be a restraint on the reliability of the comparison between the two groups.