

Data Analysis Project 1 — Team Report

Dataset: Wallisch & Whritner (2017) Movie Ratings Replication Set

$\alpha = 0.005$

Q1 – Popular vs. Unpopular Movies

Do:

Compare the mean viewer ratings between the top 50 most popular and the 50 least popular movies.

Why:

To determine whether popularity corresponds to higher audience appreciation.

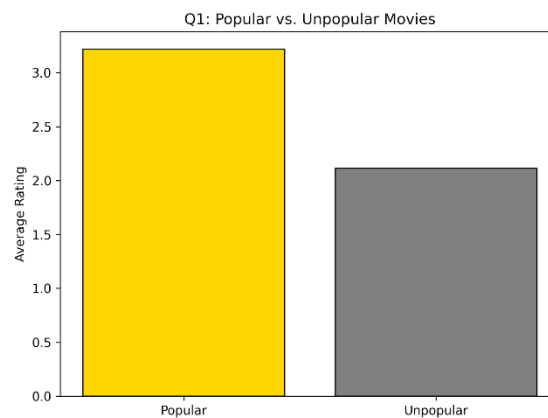
Find:

Mann–Whitney U test $\rightarrow U = 2500.00, p = 7.07 \times 10^{-18}$.

Mean(popular) = 3.22, Mean(unpopular) = 2.11.

Answer:

Significant difference — popular movies are rated considerably higher, indicating that popularity is closely associated with higher viewer satisfaction.



Q2 – New vs. Old Movies

Do:

Compare ratings between movies released before 2000 and after 2000.

Why:

To test whether newer movies receive higher ratings due to modern filmmaking standards.

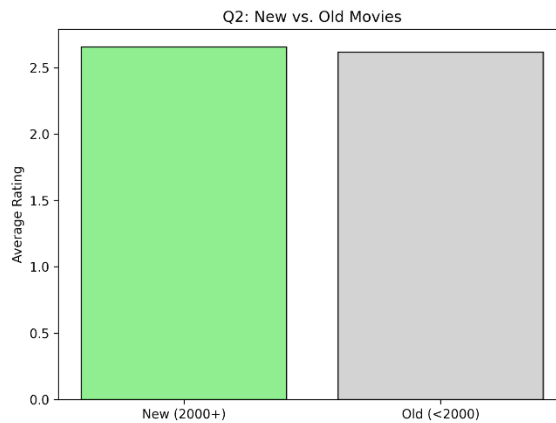
Find:

Welch's t-test $\rightarrow t = 1.120, p = 0.26336$.

Mean(new) = 2.66, Mean(old) = 2.62.

Answer:

No significant difference — newer films are not rated higher than older ones, suggesting nostalgia and classic appeal may balance out newer production quality.



Q3 — Is enjoyment of ‘Shrek (2001)’ gendered, i.e. do male and female viewers rate it differently?

D (Do):

Performed Welch’s independent-sample *t*-tests comparing male (code 2) and female (code 1) ratings for Shrek (2001). Excluded missing data and self-described gender responses.

Y (Why):

The Welch’s *t*-test was chosen because it does not assume equal variances and is robust to unequal sample sizes. Our purpose is to determine whether someone’s gender statistically impacts whether they enjoyed ‘Shrek (2001)’

F (Find):

Mean rating (male) = 3.08, mean rating (female) = 3.16.

t (df \approx variable) = -1.156, $p = 0.24835 > 0.005$.

See **Figure Q3** for a visual comparison.

A (Answer):

Since the *p*-value taken from the *t*-test is greater than 0.005, there is no statistically significant difference between how males and females’ rate ‘Shrek (2001)’.

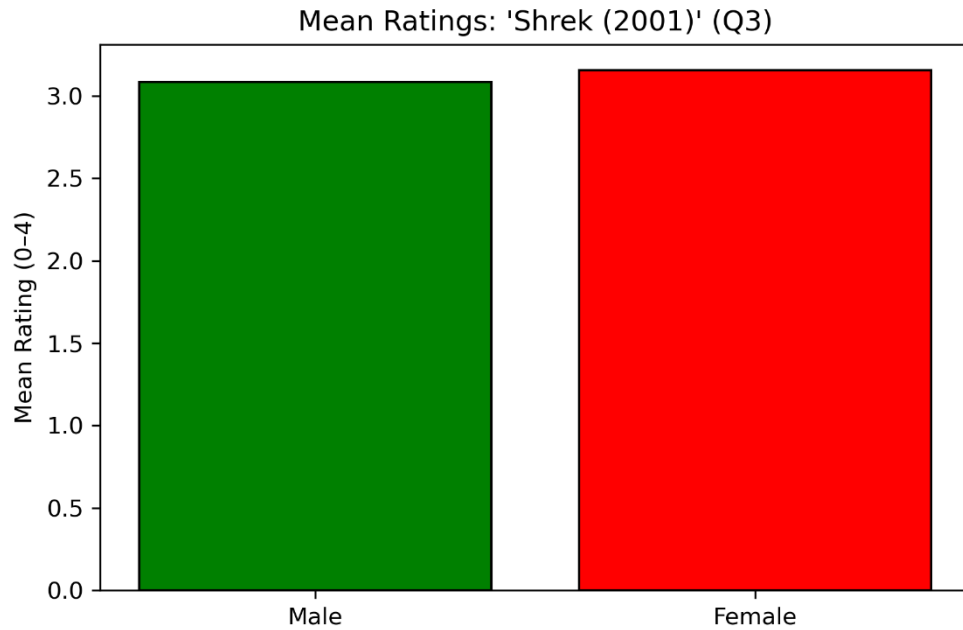


Figure Q3: “Mean Ratings by gender for ‘Shrek (2001)’

Q4 – Gender Differences Across All Movies

Do:

Test for rating differences between male and female viewers across all movies.

Why:

To explore whether gender consistently influences film ratings.

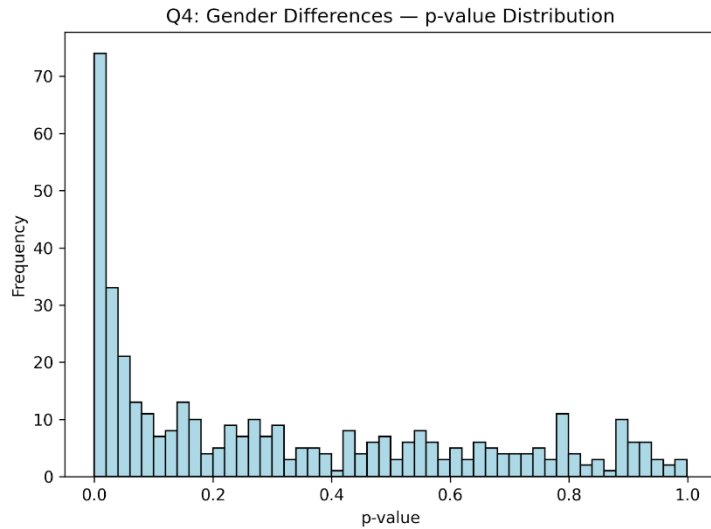
Find:

Bonferroni-corrected $\alpha = 0.00001$ (400 movies).

3 movies (0.75%) differ significantly by gender.

Answer:

Insignificant overall gender effect — male and female viewers rate nearly all movies similarly, with only a few exceptions.



Q5 – Only Child Effect (The Lion King, 1994)

Do:

Compare *The Lion King* (1994) ratings between participants who are only children and those with siblings.

Why:

To examine whether family structure influences emotional engagement with nostalgic or family-oriented films.

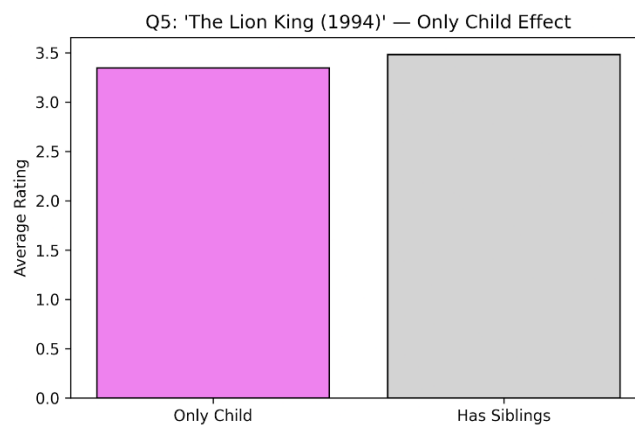
Find:

$t = -1.884$, $p = 0.06103$.

Mean(only) = 3.35, Mean(sibling) = 3.48.

Answer:

No significant difference — both only children and those with siblings rated the film equally highly, suggesting shared cultural appeal regardless of upbringing.



Q6 – Number of Movies Showing an Only-Child Effect

Do:

Determine how many movies show significant rating differences between only children and those with siblings.

Why:

To evaluate whether family background influences movie preferences more broadly.

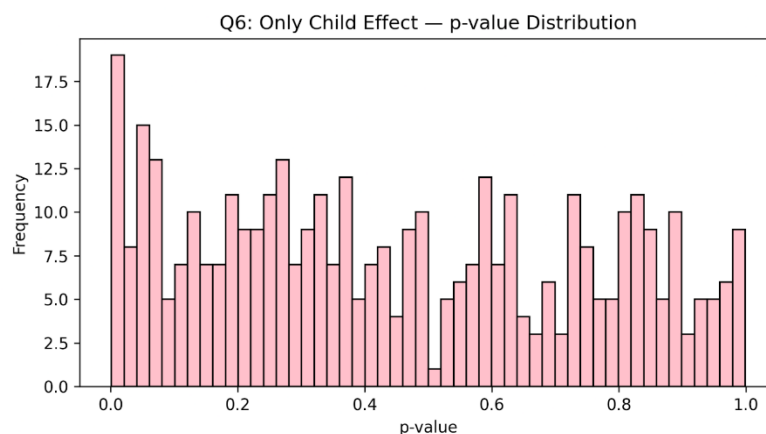
Find:

Bonferroni-corrected $\alpha = 0.00001$.

0 movies (0.00%) showed a significant “only-child” effect.

Answer:

No evidence of a general only-child effect in the dataset.



Q7 – Viewing Preference: Social vs. Alone

Do:

Compare ratings for *The Wolf of Wall Street* (2013) between participants who enjoy movies socially vs. alone.

Why:

To test if social preference impacts appreciation for socially stimulating films.

Find:

$t = -1.551$, $p = 0.12139$.

Mean(social) = 3.03, Mean(alone) = 3.14.

Answer:

No significant difference — viewing preference does not meaningfully affect how participants rate *The Wolf of Wall Street*.

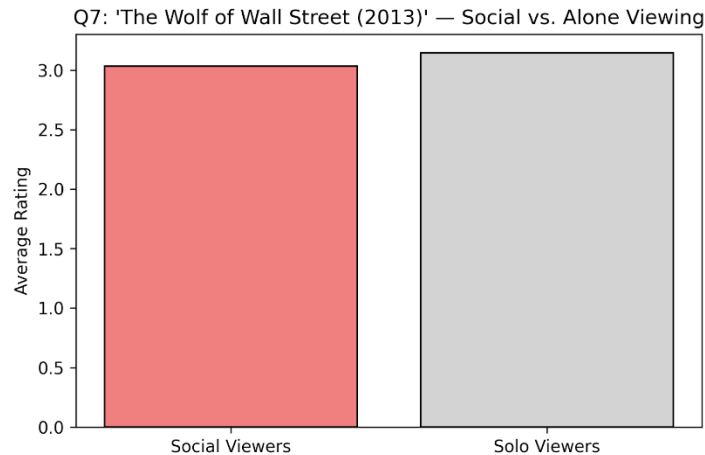


Figure Q7: Mean ratings of *The Wolf of Wall Street (2013)* by viewing preference.

Question 8 — What proportion of movies exhibit a “social watching” effect?

D (Do):

Repeated Welch’s t -tests from Question 7 for each of the 400 movies, comparing only children and participants with siblings.

Y (Why):

This systematic approach estimates how widespread “social watching” effects are across the entire dataset.

F (Find):

Only **6 movies (1.50%)** showed significant differences ($\alpha = 0.005$).

As seen in **Figure 3**, the p -value distribution is heavily skewed above 0.005, indicating minimal systematic effects.

A (Answer):

Because only 1.50% of movies exhibit significant differences, we conclude that **“social watching” preferences rarely affects movie ratings**. This suggests that social movie watching habits rarely influence someone’s enjoyment of a film

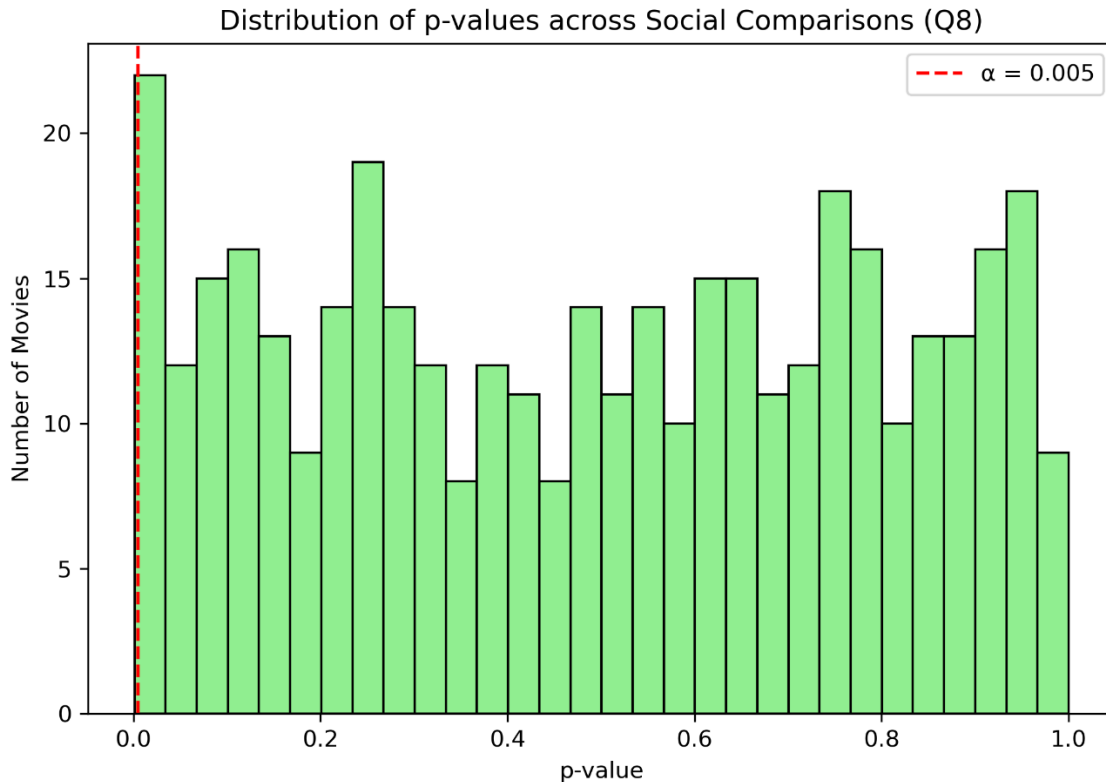


Figure Q8: Distribution of p-values across Social Comparisons (Q8)

Question 9 — Is the ratings distribution of ‘Home Alone (1990)’ different than that of ‘Finding Nemo (2003)’?

D (Do):

Run KS tests on ‘Home Alone (1990)’ and ‘Finding Nemo (2003)’ to determine if there is a significant difference in distribution for the two sets of ratings.

Y (Why):

These tests determine both the probability that the two samples have two different distributions – if there is strong enough evidence to reject the null hypothesis that the two samples have the same distribution- and at what rating in the data set the largest difference between the CDF’s of the two

F (Find):

The test found that $p=6.38e-10 < 0.005$, indicating the max value of the difference of the two cumulative distributions is significant. This is shown in *Figure Q9* below.

A (Answer):

Since $p < 0.005$, there is significant evidence that the distribution of ratings of ‘Home Alone (1990)’ and ‘Finding Nemo (2003)’ are different, meaning that we reject the null hypothesis.

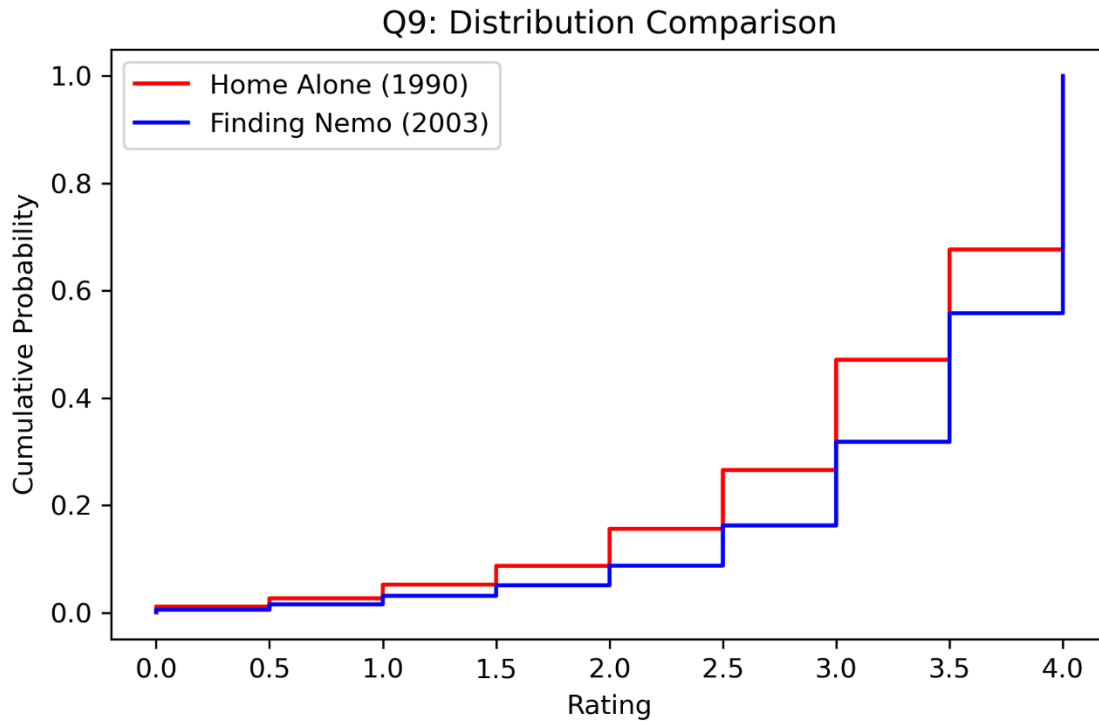


Figure Q9: Distribution Comparison

.Q10 – Franchise Consistency Analysis

Do:

Evaluate whether viewer ratings are consistent within each major movie franchise.

Why:

To identify which franchises have stable quality versus fluctuating audience satisfaction.

Find:

ANOVA and Kruskal–Wallis tests:

7 of 8 franchises show $p < 0.05$.

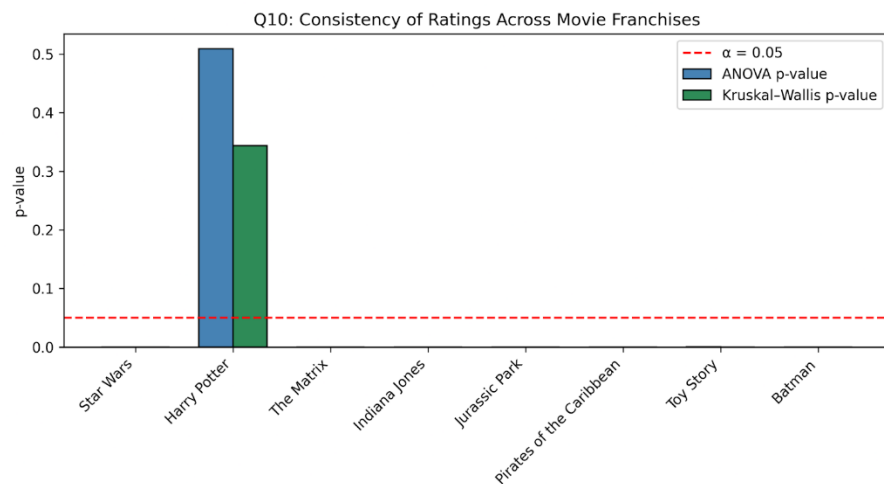
Only *Harry Potter* maintains consistent quality across films.

Franchise	Movies	ANOVA p	KW p	Consistent?
Star Wars	6	0.00000	0.00000	no
Harry Potter	4	0.50899	0.34332	yes
The Matrix	3	0.00000	0.00000	no

Indiana Jones	4	0.00000	0.00000	no
Jurassic Park	3	0.00000	0.00000	no
Pirates of the Caribbean	3	0.00007	0.00003	no
Toy Story	3	0.00048	0.00001	no
Batman	3	0.00000	0.00000	no

Answer:

Most franchises (7 out of 8) show inconsistent viewer ratings, with *Harry Potter* being the only stable series in perceived quality.



. Bonus — How much does personal life stress impact how someone would rate ‘The Silence of the Lambs (1991)’

D (Do): Sort the scores of individuals who rated ‘The Silence of the Lambs (1991)’ into sets based on stress ratings from 1 to 5 and then run an ANOVA test on the means of each set

Y (Why): The ANOVA test takes analyzes if there is a statistically significant difference between the general ratings of sets with more than 2 elements, so its use applies to looking at the means of ratings of ‘The Silence of the Lambs’ in different stress groups.

F (Find):

p=0.92823, with the figure below showing similar average scoring between all of the test groups. **Figure**

Bonus shows that the mean rating of the movie for people with specific life stress levels is not significantly different.

A (Answer): The data does not show that stress has a significant impact on how individuals rated 'The Silence of the Lambs (1991)'

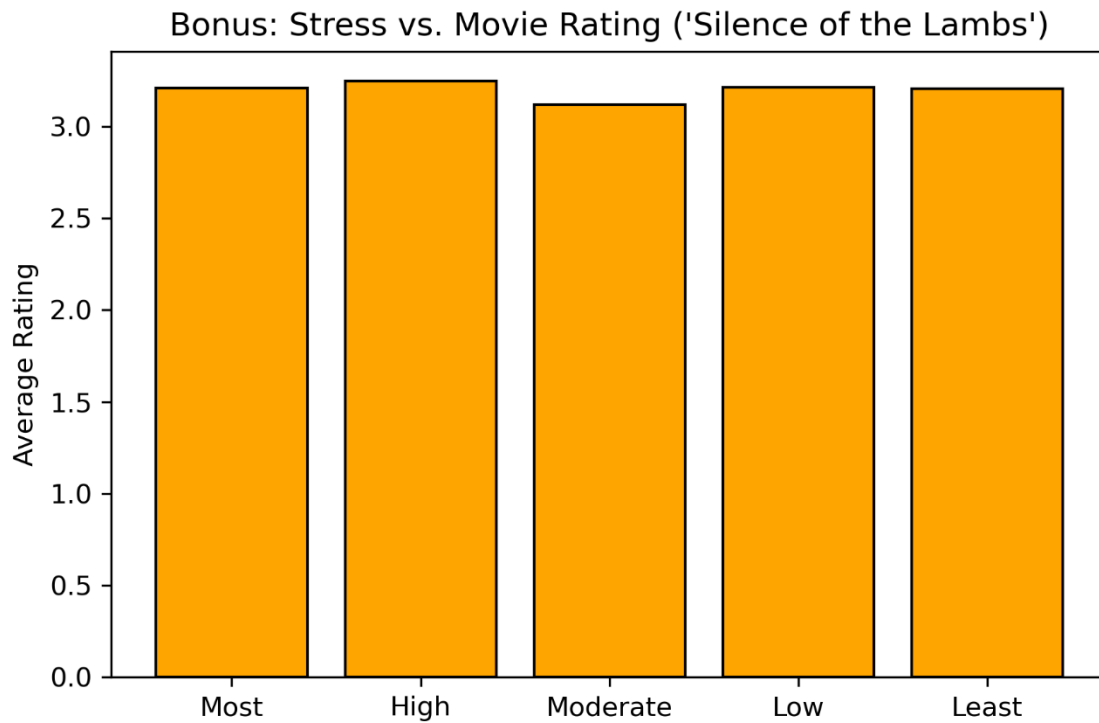


Figure Bonus: Stress vs. Movie Rating ('Silence of the Lambs')