Data Analysis Project 1

Francisco Fabian Juarez Padilla

The following project answers 10 specific questions utilizing different hypothesis testing tools and using data from 400 movies from 1097 research participants. The per-test significance level is set to 0.005[1]. The removal of missing data was done in the following ways:

  i. Element-wise removal (EWR): remove all missing data individually from each group. This is used when the participants are assumed to come from different populations.

  ii. Row-wise removal (RWR): if at least one of the observations contains missing data, the entire participant information is removed. This is used when comparing two or more movie ratings (the participant is required to see all movies being compared).

The mean (m), median (me), standard deviation (SD), and standard error (SEM) are used to compare the ratings, in most of the questions the T-test and U-test[2] (ANOVA and Kruskal-Wallis when there are multiple groups) are both used[3]. Unless specified differently the default two-sided test is considered. In most questions the effect size (of the mean) is also calculated to show the relevance of the result[4].

Executive summary answers

1. Yes, more popular movies are rated higher than less popular, the effect is very large.
2. Yes, newer movies are rated higher than older ones, but the effect is very small.
3. We can't say, there is not a significant difference between male and female viewers on their enjoyment of Shrek.
4. 11% of the movies are significantly rated different by female and male reviewers, half of them show moderate or strong effects.
5. We can't say, there is not a significant difference between only children and persons with siblings on their enjoyment of Lion King.
6. 1.8% of the movies are significantly rated different by only children and people with siblings, more than half of them show moderate or strong effects.
7. We can't say, there is not a significant difference between social watchers and persons who prefer to watch movies alone on their enjoyment of The Wolf of Wall Street.
8. 1.3% of the movies are significantly rated different by social watchers and people who prefer to watch movies alone, almost all movies have very small effects.
9. Yes, the ratings distributions of Home Alone and Finding Nemo are significantly different.

---

[1] If the p-value is smaller than α, we will conclude the result is significant.

[2] Mann-Whitney U test for medians of ordinal data, which could be suitable for movie ratings.

[3] As shown by (Wallisch & Whritner, 2017) in a similar analysis, T-tests by themselves can show significant results, but medians also provide important information.

[4] The following criteria is used to interpret the |effect size|: Very Small [0.01,0.2), Small [0.2, 0.5), Medium [0.5,0.8), Large [0.8,1.2), Very Large [1.2,2), Huge [2, infinity)

10. Seven franchises (from 16 found in the data set) show significant inconsistent ratings (quality) between their movies.

Detailed answers

1) Are movies that are more popular (operationalized as having more ratings) rated higher than movies that are less popular?[5]

After an EWR, a median split of the number of ratings (me = 197.5 ratings) was used to classify the 400 movies in two groups. Then the average rating of each movie was calculated and then compared using a one tailed paired samples T-test and a U-test. I concluded that more popular movies (m = 2.868, me = 2.897, SD = 0.291, n = 200) are rated higher than less popular movies (m = 2.401, me = 2.402, SD = 0.231, n = 200), this result is significant with a very large effect size (t = 17.822, p = 2.14 -43, d = 1.780) and $(u = 4596, p = 8.49e\text{-}41)$.

2) Are movies that are newer rated differently than movies that are older?

After an EWR, a median split of the release year (median = 1999) was used to classify the movies in two groups[6]. Then the mean rating of each group was compared using an independent t test. I concluded that newer movies (mean = 2.840, SD = 1.053, n = 57,266, SEM = 0.004) are rated higher than older (mean = 2.814, SD = 1.063, n = 46,344, SEM = 0.005), this result is significant but with a very small effect size, making it of questionable relevance (t = 3.945, p = 0.000, d = 0.025) and $(u = 1308552619.500, p = 0.000)$.

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

After an EWR and classifying the sample between female (743) and male (241)7, the mean rating of each gender was compared using and independent t test. Male participants (mean = 2.083, SD = 0.823, n = 241, SEM = 0.053) show a lower rating of Shrek in comparison to female participants (mean = 3.155, SD = 0.906, n = 743, SEM = 0.033). Nevertheless, this result is not significant and has a very small effect size (t = -1.102, p = 0.271, d = -0.082, df = 982) and (u = 82232.500, p = 0.025).

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

---

[5] Nevertheless, by categorizing a continuous predictor (number of ratings) there is a loss of power (Aiken, 1991). By doing this we assume all movies in each category are the same, with almost no separation between the values closer to the median. A general linear model would be more adequate to answer this question and the following that involve categorizing continues predictors.

[6] The movies released in 1999 (39) were excluded from the analysis to avoid biased assignation to either group. One movie was also excluded for not showing the release year in the data set.

[7] If the population we are trying to describe is all the people, there could be a sampling error, as in general, there is a 50-50 ratio of female and male population if sampling randomly. Even before the EWR, the ratio was almost the same.

After an EWR and classifying the sample between female (807) and male (260), 400 (one for each movie) t-tests and u-tests were conducted between male and female ratings. The results of the tests are summarized below.

Table 1. Summary results of 400 independent T tests and U tests between female and male viewer's ratings.

Test	Number of significant results	Proportion of significant results	Significant results with $|d| > 0.2$	Significant results with $|d| > 0.5$	Significant results with $|d| > 0.8$
T-test	46	11.5%	44	20	5
U-test	61	15.25%	53	20	5
Both tests	44	11%	44	20	5

Taking both tests into consideration we can say that 11% (44 movies) are rated differently by male and female viewers. Nevertheless, only 5 movies showed strong effect size, 15 moderate and 19 small, which means that almost half of the significant results are of questionable relevance.

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

After an EWR and classifying the sample between only children (151) and people with siblings (776), the mean rating of each group was compared using a one tailed independent t test. Only children (m = 3.348, SD = 0.814, n = 151, SEM = 0.066) show a lower rating of Lion King in comparison to people with siblings (mean = 3.482, SD = 0.718, n = 776, SEM = 0.026). The test result is not significant and has a very small effect size (t = -2.054, p = 0.98, d = -0.183, df = 925) and (u = 52929.000, p = 0.978).

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

After classifying the sample between only children (177) and people with siblings (894) and an EWR, 400 (one for each movie) t-tests and u-tests were conducted between these two group's movie ratings. The results of the tests are summarized below.

Table 2. Summary results of 400 independent T tests and U tests between only children and people with sibling's movie ratings.

Test	Number of significant results	Proportion of significant results	Significant results with $|d| > 0.2$	Significant results with $|d| > 0.5$	Significant results with $|d| > 0.8$
T-test	10	3.5%	10	4	1
U-test	14	2.5%	14	7	1
Both tests	7	1.8%	7	4	1

Taking both tests into consideration we can say that 1.8% (7 movies) are rated differently by only children and people with siblings. Nevertheless, only 1 movie showed strong effect size, 3 moderate and 3 small, which means that almost half of the significant results are of questionable relevance.

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?

After an EWR and classifying the sample between social watchers (270) and alone watchers (393), the mean rating of each group was compared using a one tailed independent t test. Social watchers (mean = 3.033, SD = 0.919, n = 270, SEM = 0.056) show a lower rating of the Wolf of Wall Street in comparison to people who prefer to watch movies alone (mean = 3.143, SD = 0.869, n = 393, SEM = 0.044). The result is not significant and has a very small effect size (t = -1.568, p = 0.941, d = -0.124, df = 661) and ($u = 49303.500, p = 0.944$).

8) What proportion of movies exhibit such a "social watching" effect?

After classifying the sample between social watchers (462) and people who prefer to watch movies alone (610) and an EWR, 400 (one for each movie) t-tests and u-tests were conducted between these two group's movie ratings. The results of the tests are summarized below.

Table 3. Summary results of 400 independent T tests and U tests between movie ratings from social watchers and people who prefer to watch movies alone.

| Test	Number of significant results	Proportion of significant results	Significant results with $\|d\| > 0.2$	Significant results with $\|d\| > 0.5$	Significant results with $\|d\| > 0.8$
T-test	6	1.5%	6	1	0
U-test	13	3.3%	11	1	0
Both tests	5	1.3%	5	1	0

Taking both tests into consideration we can say that 1.3% (5 movies)[8] are rated differently by social watchers and people who like to watch movies alone. Nevertheless, not a single movie showed strong effect size, 1 showed moderate and 4 small, which means that almost all results are of questionable relevance.

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

---

[8] The result also shows that "cult movies" (Silence of the Lambs and Donnie Darko) are rated higher by alone watchers and "commercial movies" (North, Shrek 2, and Spiderman) are rated higher by social watchers.

After a RWR the distributions of the ratings of 810 participants who saw both Home Alone and Finding Nemo were compared using a Kolmogorov-Smirnov test. The distributions of both movies are significantly different (k = 0.168, p = 2.204 e-10).

Figure 1. Results of the Kolmogorov-Smirnov test, histogram, and probability distributions of Home Alone (blue) and Finding Nemo (orange).
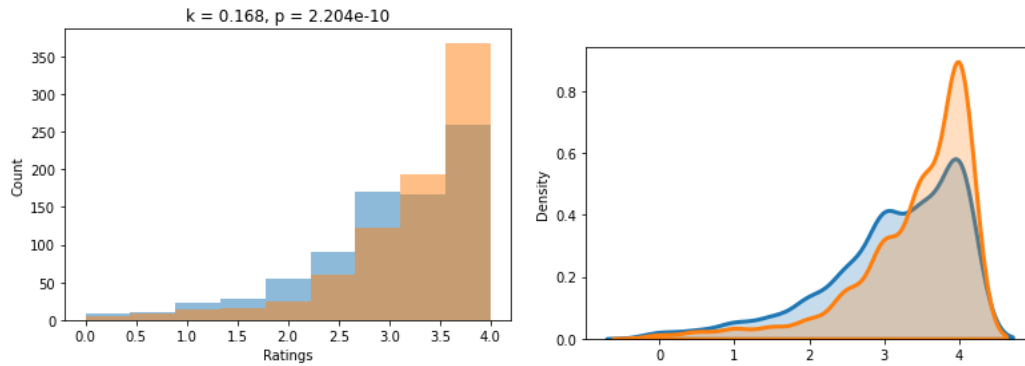
10) There are ratings on movies from several franchises (['Star Wars', 'Harry Potter', 'The Matrix', 'Indiana Jones', 'Jurassic Park', 'Pirates of the Caribbean', 'Toy Story', 'Batman']) in this dataset. How many of these are of inconsistent quality, as experienced by viewers?

16 franchises were identified in the data set. As the number of movies in one franchise can be more than two and to avoid inflating alpha for multiple comparisons, I used a Kruskal-Wallis and ANOVA tests. After a RWR to ensure all the participants have seen the entire franchise, the multiple tests showed the following results:

Table 4. Summary results of 16 ANOVA and Kruskal-Wallis tests between the ratings of the movies in each franchise.

Franchise	ANOVA		Kruskal-Wallis		n	Movies	Significant[9]
	f	p	h	p			
Star Wars	39.030	2.40E-38	193.510	6.94E-40	333	6	Yes
Harry Potter	1.446	0.228	5.874	0.118	710	4	No
Matrix	18.593	1.30E-08	40.323	1.75E-09	260	3	Yes
Indiana Jones	19.051	5.20E-12	54.194	1.02E-11	244	4	Yes
Jurassic Park	22.164	3.54E-10	49.427	1.85E-11	398	3	Yes
Pirates of the Caribbean	3.447	0.032	6.660	0.036	561	3	No

---

[9] Both tests arrived at the same significance conclusion.

Toy Story	7.588	5.19E-04	23.497	7.90E-06	757		3	Yes
Batman	43.626	1.64E-18	84.658	4.14E-19	219		3	Yes
Shrek	64.514	1.68E-15	62.919	2.15E-15	933		2	Yes
The Lord of The Rings	0.155	0.856	0.558	0.76	390		3	No
The Godfather	3.321	0.069	7.124	0.01	284		2	No
Terminator	6.881	0.009	7.655	0.01	263		2	No
Kill Bill	0.713	0.399	0.729	0.39	300		2	No
Bad Boys	1.331	0.250	1.106	0.29	134		2	No
Rocky	4.715	0.031	5.691	0.02	173		2	No
X-men	4.552	0.033	6.372	0.01	429		2	No

The results show that Star Wars, Matrix, Indiana Jones, Jurassic Park, Toy Story, Batman, Shrek show significant differences in their ratings, therefore we can conclude these seven movie franchises have inconsistent ratings (quality) as reported by viewers.