**Team 2A**- **Rating comparison of popular genres in Hollywood and Bollywood: Statistical Analysis**

**The research question under study. Why is it important and interesting, but difficult to answer?**

We believe this question is important because it speaks to the cultural and traditional differences between the two societies, as well as shine a light on various situations that are inherent to those areas. Interestingly enough, we generally see that Bollywood is focused internally, opting to tackle topics within India, whereas with Hollywood, we moreso see productions with a focus on global influence. This isn’t to say that one is better than the other, however, it is important to highlight these differences so we can understand different people and different perspectives.

This question is also difficult to answer because of differing opinions. What is a 9/10 movie for someone, could be a 4/10 for someone else. That is the nature of the entertainment industry and it can be quite challenging to navigate that as we find data and make recommendations.

**How do you get the data for your statistical analysis?**

We collected data for our analysis from Kaggle.com, which included movie ratings for Hollywood and Bollywood films in various genres from the years 2012 to 2022, rated on a scale of 1 to 10. We focused on five common genres (Drama, Comedy, Action, Romance, and Crime) to compare ratings between the two industries. Ratings were obtained using the TMDB API, and we included all movies tagged with any of our researched genres.

<https://www.kaggle.com/datasets/darpananand07/hollywood-and-bollywood-movies-dataset>(Anand07, 2023)

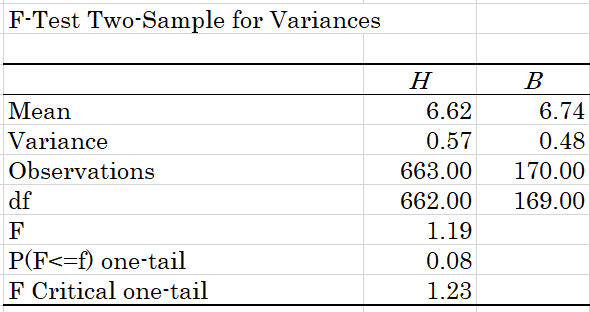
**Describe the statistical analysis and provide the statistical results in support of your conclusion**

In our analysis, we focused on four key variables from the dataset: genre, original\_language, vote\_average, and vote\_count. To ensure the accuracy of our analysis, we filtered the data to include only rows with a vote count of 30 or more. Next, we calculated the vote averages for each genre within the Hollywood and Bollywood industries.

To satisfy the assumptions required for the t-test, we initially conducted an F-test two-sample analysis for variances within each genre using Excel. At a significance level of 0.05, we confirmed that all pairs of comparable samples exhibited equal population variances.

Subsequently, we proceeded with the two-sample t-test, using mean and standard deviation of vote-average from the dataset for both Hollywood and Bollywood for hundreds of movies. We employed p-values to determine whether or not the average rating of one of the industry is significantly higher than the other in each genre at 𝝰=0.05.

**Drama**

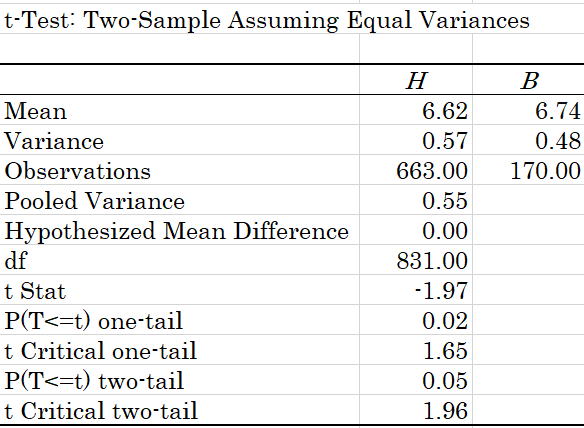


H0: ; Ha:

p-value×2=0.16 >𝝰=0.05

do not reject H0

The variances of two population are equal at 0.05 level of significance

****

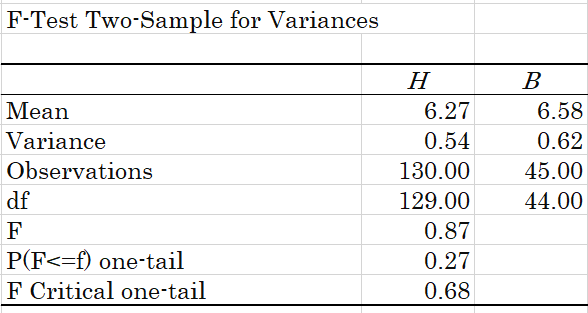
H0: ; Ha:

p-value=0.02 <𝝰=0.05

reject H0

The average rating for Bollywood is significantly higher than that for Hollywood under the Drama genre.

**Crime**

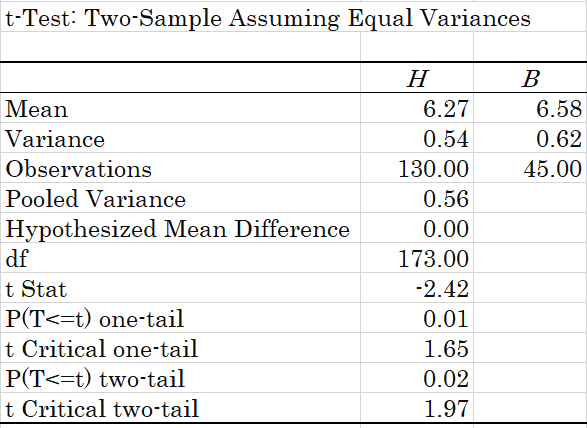
****

H0: ; Ha:

p-value×2=0.52 >𝝰=0.05

do not reject H0

The variances of two population are equal at 0.05 level of significance

****

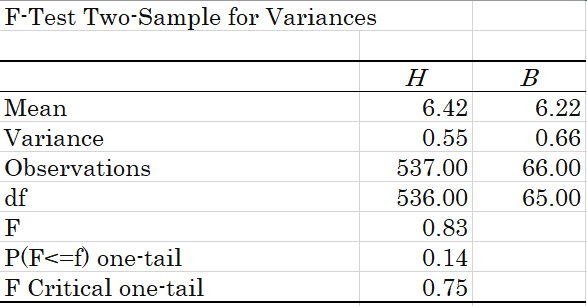
H0: ; Ha:

p-value=0.01 <𝝰=0.05

reject H0

The average rating for Bollywood is significantly higher than that for Hollywood under the Crime genre.

**Action**

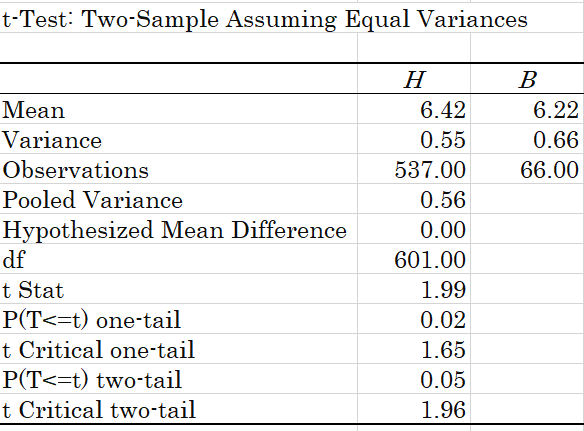
****

H0: ; Ha:

p-value×2=0.28 >𝝰=0.05

do not reject H0

The variances of two population are equal at 0.05 level of significance

****

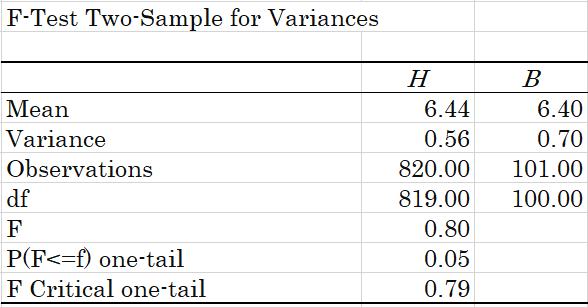
H0: ; Ha:

p-value=0.02 <𝝰=0.05

reject H0

The average rating for Hollywood is significantly higher than that for Bollywood under the Action genre.

**Comedy**

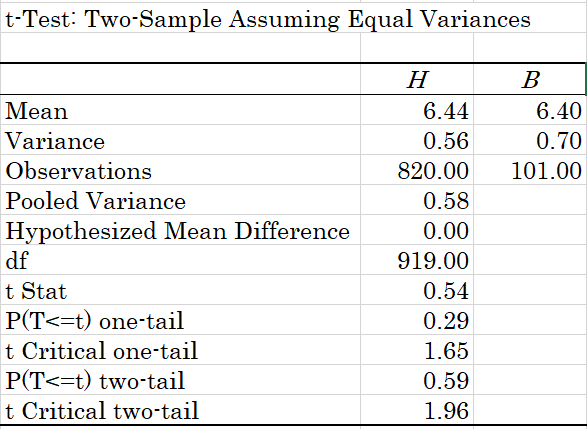
****

H0: ; Ha:

p-value×2=0.1 >𝝰=0.05

do not reject H0

The variances of two population are equal at 0.05 level of significance



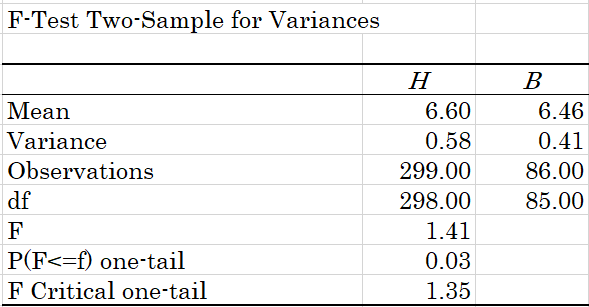
H0: ; Ha:

p-value=0.29 >𝝰=0.05

do not reject H0

We can not conclude that the average rating for Hollywood is significantly higher than that for the Bollywood under the Comedy genre.

**Romance**

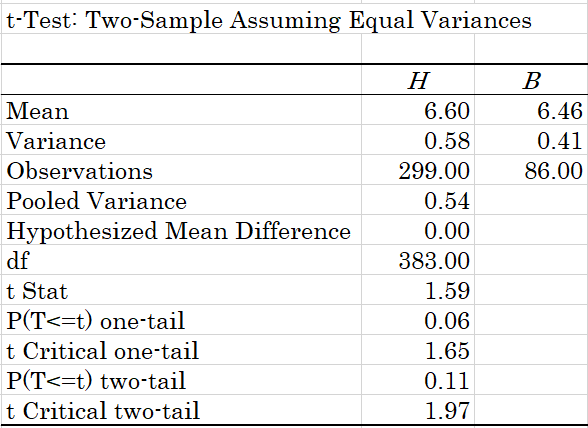


H0: ; Ha:

p-value×2=0.06 >𝝰=0.05

do not reject H0

The variances of two population are equal at 0.05 level of significance



H0: ; Ha:

p-value=0.06 >𝝰=0.05

do not reject H0

We can not conclude that the average rating for Hollywood is significantly higher than that for the Bollywood under the Romance genre.

**What is the conclusion to the research question based on your analysis in layman terms? Any limitations due to data integrity and availability?**

Here’s the conclusions to Hollywood vs Bollywood based on our analysis by the top 5 genres:

Drama- Bollywood

Crime- Bollywood

Action- Hollywood

Comedy-Tie

Romance-Tie

These conclusions provide us with a better understanding of the popularity of Hollywood and Bollywood within 5 popular genres. This will help our classmates to pick a good Bollywood movie to watch. They can pick one from any of the genres that had a result of Bollywood ratings being significantly greater from our conclusions.

From the producers perspective, it is useful to know what genres of movies are popular with the audience. With this knowledge, they can make more movies in that specific genre. So the take away for Bollywood producers is to make more movies in the Drama and Crime genre. And the take away for Hollywood producers is to make more movies in the Action genre. And both Hollywood and Bollywood producers should make more movies in the Romance and Comedy genre.

Data availability was not a limitation to our research. In terms of data integrity, we picked only the ratings which had more than 30 people who rated it. So it would have been helpful if all the movies were rated by more than 30 people. Also since ratings are subjective, it would have been better to have more ratings and sources.

**References:**

Anand07, D. (2023, July 4). *Hollywood and Bollywood Movies Dataset*. Kaggle. https://www.kaggle.com/datasets/darpananand07/hollywood-and-bollywood-movies-dataset