Project Part 3

Joseph Schneeberger

Background

When researching the film industry, I discovered that there are a couple of months throughout the year called "dump months," where film companies release movies they expect to perform poorly. Film companies strategically release their best films in the early summer and "dump" the other ones in January-February [Reference 1]. After this discovery, I wanted to perform a statistical test to find out if this phenomenon truly exists. Luckily, I had a dataset that possesses information on every Disney movie released since the company's inception. Some of the variables include movie name, year of release, genre, and MPAA rating. From the multitude of variables available from the dataset, the two that pertain to our inquiry are: **month released** and **inflation-adjusted earnings**.

Research Question

Do films released in January truly perform worse than films released during the remaining months of the year?

Data Description and Generalization

Disney is one of the largest film companies in the world with a vast collection of films belonging to every popular genre [Reference 2]. Their presence across generations gives us data that behaves alongside technological shifts and changes in consumer additudes and tastes. Therefore, the samples collected on Disney movies possess all the necessary characteristics that can be generalized to the movie industry as a whole. In other words, since Disney has been subject to every major breakthrough in the film industry, possesses films from every popular genre, and has existed across many generations, it makes for a comprehensive source for samples about the film industry.

Chosen Test

The test chosen in this analysis was the left-sided Wilcoxon Rank Sum Test because the samples were from two independent, continuous populations with the same shape. The test was left-tailed because we wanted to know if the earnings for January are smaller than the earnings for the remaining months. The two samples were right skewed based on histograms and boxplots performed for each [code included as comment]. The two samples were independent because a film can only be released in one month and has its own box-office earnings.

Wilcoxon Rank Sum Test

Null Hypothesis: The earnings for films released in January are the same as the earnings for films released in the remaining months. (Ho: M1=M2)

Alternative Hypothesis: The earnings for films released in January are less than the earnings for films released in the remaining months. (Ha: M1 < M2)

```
disney <- read.csv("/Users/joey/Desktop/Spring 2020/disney.csv")
disney <- disney[,c(-1,-2)]

disney_jan <- disney$Real_Earnings[disney$Month=="January"]
disney_other <- disney$Real_Earnings[disney$Month!="January"]

# boxplot(disney_jan,outline=F)
# boxplot(disney_other,outline=F)
# hist(disney_jan,"FD")
# hist(disney_other,"FD")

wilcox.test(disney_jan,disney_other,alternative = "less")
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  disney_jan and disney_other
## W = 7835.5, p-value = 0.007127
## alternative hypothesis: true location shift is less than 0
```

Conclusion

Based on the Wilcoxon Rank Sum Test, we have enough statistical evidence to conclude that the earnings in January are less than those in the remaining months. Because the samples were gathered from a dataset that possesses attributes that allows it to be generalized to the film industry as a whole, we can conclude that January is, in fact, a dump month across the industry.

References

1. https://en.wikipedia.org/wiki/Dump_months#cite_note-Guardian_article-1.
2. https://www.britannica.com/topic/Disney-Company