MOVIES SEASONAL PERFORMANCE STATISTICS

By: Nam Pham, Linh Nguyen, Rosario Guel, Margaret Keenan

UNIVERSITY of WASHINGTON

Introduction

This project investigates if the season of the year a movie is released has any bearing towards its total gross revenue. We were interested in knowing which season generates the most box office revenue and how the seasons compare to each other.

Method

Our data was collected using *ParseHub* (an open source web scraping tool) on Box Office Mojo [1], a web site that stores various data on yearly movies' box office. We pull the top 100 grossing movies of every year from 1980 to 2017 into our data base. (so around 3800 movies total)

We then sort and process our movie data, grouping them by year and then by their opening date. We further divide each year into 4 quarters/ seasons (Winter, Spring, Summer, and Fall) and calculate how each season performs financially.

We then perform a statistical analysis on that dataset in order to see if the season of the year that a movie is released has any influence toward its financial performance.

Claim

The null hypothesis is that the average box office gross per year is the same for all seasons. The alternative hypothesis is that the average box office gross per year is not the same for all seasons.
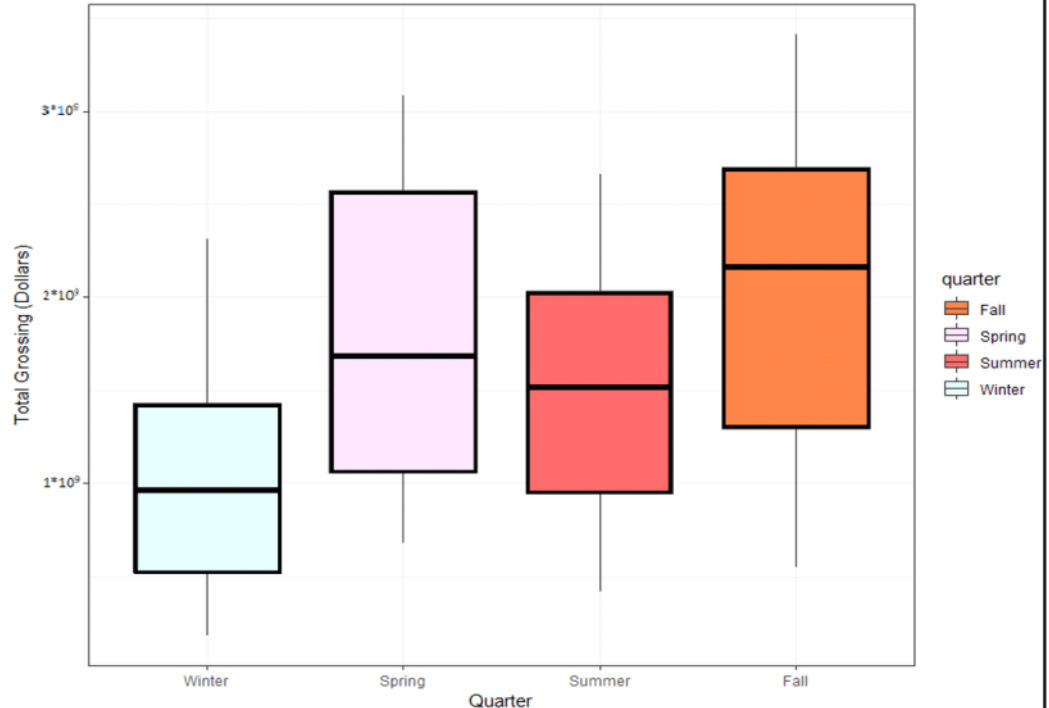
> $H_0 : \mu_{Win} = \mu_{Spr} = \mu_{Sum} = \mu_{Fall}$
> $H_A : \mu_{Win} \neq \mu_{Spr} \neq \mu_{Sum} \neq \mu_{Fall}$
> $\alpha = 0.05$

Result

Distribution of each quarter's total grossing

quarter: Fall, Spring, Summer, Winter

> Winter makes the least revenue, with every other seasons' Q_1 above its mean.
> Spring and Fall's revenue is quite similar, with Fall having a higher mean.
> Surprisingly, Summer's revenue is behind both Spring and Fall. Fall's mean is even above its Q_3. We did not expect this at first, but it is explainable.

Data Summary	Winter	Spring	Summer	Fall
Mean	1,021,233,239	1,817,182,358	1,470,227,204	2,002,648,365
Standard Deviation	574,825,398	798,690,441	616,285,290	868,419,737
Variance	3.30E+17	6.38E+17	3.80E+17	7.54E+17
95% Confidence Interval Lower	867,852,294	1,604,067,394	1,305,783,496	1,770,927,498
95% Confidence Interval Higher	1,174,614,184	2,030,297,323	1,634,670,912	2,234,369,231

Box Office Mojo

ANOVA	$\alpha = 0.05$					
Source of Variation	SS	df	MS	F	P-value	F-crit
Between Groups	2.12473E+19	3	7.08E+18	13.47562	8.02E-08	2.665729
Within Groups	7.77848E+19	148	5.26E+17			
Total	9.9032E+19	151				

Scheffe Table	Winter	Spring	Fall	Summer
Winter				
Spring	22.90			
Fall	34.82	**1.24**		
Summer	7.29	4.35	10.25	

Reference

[1] https://www.boxofficemojo.com/yearly/chart/?yr=xxxx&p=.htm where "xxxx" equals to the year that we want to collect data on

Discussion

At an $\alpha = 0.05$ level of significance, through an ANOVA test, we determine that our p-value is less than α, and that our calculated F-value is larger than our F-statistic. Therefore, we reject the null hypothesis that the average box office gross per year is the same for all seasons. Through our Scheffé method, we then determine that only the Spring-Fall pair are not statistically significance because 1.24 is smaller than 2.66. All the other pair are significantly different from each other.

The reason for this difference is likely because of how each season affects entertainment. It is too cold in winter so people would likely not go out to watch movies. In the summer people have many other choices of entertainment such as going to the pool and summer sports, so they do not go to movies that often. Spring and fall are the closest in temperature/ mood and there are not as much competition for consumer time so people go watch movies more often, explaining why their gross box office revenue is not

mood and there are not as much competition for consumer time so people go watch movies more often, explaining why their gross box office revenue is not significantly different from each other.

R e f e r e n c e

[1] https://www.boxofficemojo.com/yearly/chart/?yr=xxxx&p=.htm where "xxxx" equals to the year that we want to collect data on

Conclusion

At an $\alpha = 0.05$ level of significance, through an ANOVA test, we determine that our p-value is less than α, and that our calculated F-value is larger than our F-statistic. Therefore, we reject the null hypothesis that the average box office gross per year is the same for all seasons. Through our Scheffé method, we then determine that only the Spring-Fall pair are not statistically significance because 1.24 is smaller than 2.66. All the other pair are significantly different from each other.