My inferential report has two parts: before movie's announcement and after movie's announcement. If I am an investor, I would love to know whether I can make money from the movie. At least, I would not make some bad decisions leading a huge money loss. It is also better to know about it at the day 1 of my movie plan. This will belong to the part of "Before movie's announcement". However, there is also a lot of signs that we can tell the revenue direction after the movie announced. Even it is the best timing to realize we are gonna lose money (if it is a negative revenue), there is some possible ways to rescue the situation in the early phase. This will belong to "After movie's announcement".

After movie's announcement:

Is IMDB score a good indicator of movie revenue?

Ho: The correlation between IMDB score and revenue is 0
HA: The correlation between IMDB score and revenue is not 0

From the scatter plots, I can easily tell that high IMDB scores can indicate the movies with high budget and high USA gross. However, movies with high IMDB scores can lead the revenue positively and negatively. If I just eyeball it, there is more chance to make money from the higher IMDB score. In a further investigation, there is 23% correlation between revenue and imdb score. I use permutation to simulate the null hypothesis and compare to the correlation of observed samples. The null hypothesis is reject and 23% correlation is significant in the perspective of statistical significance.

Before movie's announcement:

Is the budget correlated to the revenue? Invest more, earn more back?

Ho: The correlation between budget and revenue is 0
HA: The correlation between budget and revenue is not 0

Interestingly, the budget and revenue seem to have a negative correlation instead of positive one if the movies don't make profits. In terms of whole sample population, there is only 2.94% correlation. The p-value of that is 0.0255 which is between alpha = 0.01 and alpha = 0.05. I don't think it is strong enough to conclude the correlation is true in a practical way. I further separate sample population into two parts: the positive revenue and the negative revenue. Then I compare them to the budget, respectively. The result shows that there is a ~76% correlation between revenue and budget in movies with a negative revenue. On the other hand, there is a ~35% correlation between revenue and budget in movies with a positive revenue. Both of them are statistically significant. I would conclude that it is not necessary to make more money when the investors invest more money. The budget has to be in an adequate range for the movie's success (HIGH REVENUE).

Is there a seasonal trend of revenue?

In terms of revenue, June might be the best timing to release the new movie. From the boxplot, June's revenue is obviously higher than other months. However, are they significant in a statistical significance context? I compare the mean revenue of June to all other months individually. In a more stringent condition (alpha = 0.01), the results indicate that May, June, July, November, and December are good for announcing the movies compared to other months. Although the results are pretty reasonable, I now understand it in a more statistical sense.

Is any particular genre making more money?

Animation and family categories could be good profit indicators of movies. The mean revenue difference between animation movies and non-animation movies is 17 million. Also, the mean revenue difference between family movies and non-family movies is 14 million. I compare them to simulated null hypothesis from bootstrapping and calculate their p values. The results show that those mean revenue differences are both significant.