Homework 1
Problem 2
Ayad Masud
733009045

I certify that I have personally done the coding, generated the figures and written the report without aid from anybody else, and that I have not plagiarized, self-plagiarized, or used AI-generated text. I certify that I have acknowledged any sources I used to complete this assignment. ARM.

1 Part 1: Starting and Refining the Question

During question 1 of this homework assignment, we were tasked with performing EDA on a dataset containing movies and information about these movies. While preforming EDA a trend that appeared multiple times was the number of movies and the gross revenue of those movies increasing as time passed. Results are shown in **Figure 1**. From this simple bar plot we can see a trend that the total gross revenue increases over time. Apart from the year 2020. This raises the question, what might have caused this steady increase in gross revenue? Is it factors such as marketing, advertisement, audience preferences or could it be something like the introduction and adoption of CGI in the film industry? I will be exploring the question: Did the invention and adoption of CGI impact the gross revenue of films as the years went by? The dependent variable will be the gross revenue of movies and the independent variable will be whether a movie was made before or after CGI was introduced.

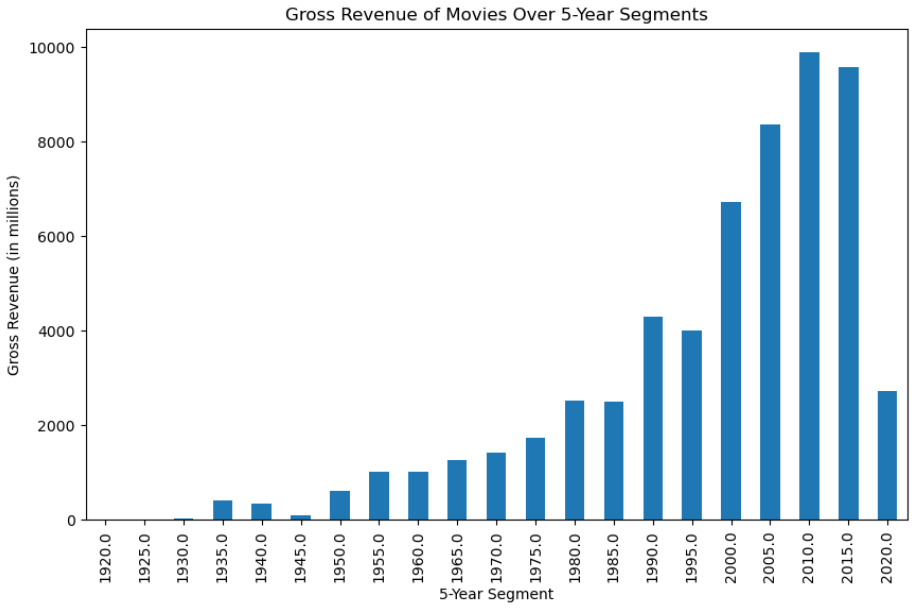

Figure 1. Bar plot showing in 5-year intervals the total gross revenue of movies.

2 Part 2: Exploring the Data

Although basic CGI was introduced as early as the 1950's, it became very popular in the 1990's. Specifically after the release if Jurassic Park in 1993, which utilized CGI to provide stunning visuals. Using 1993 as a threshold year where before 1993 movies didn't use CGI and after 1993 movies did use CGI, we will conduct some statistical tests to see if the usage of CGI corresponds with the increase in gross revenue. First let's look at **Figure 2** which is a box plot showing summary statistics of both samples: gross revenue before 1993 and gross revenue after 1993. We can see from the plot that the median gross revenue is around the same. However, looking at the interquartile range and outliers we can see there is a large difference in

gross revenue. The mean gross revenue before 1993 is around $50 million while the mean gross revenue after 1993 is $87 million. This apparent difference between the samples indicates that there may be an underlying cause for the increase. Looking at **Figure 1** again we can see in between the years of 1990 and 2000 there is also an increase in number of movies, which may have been aided by the introduction of CGI during that time.

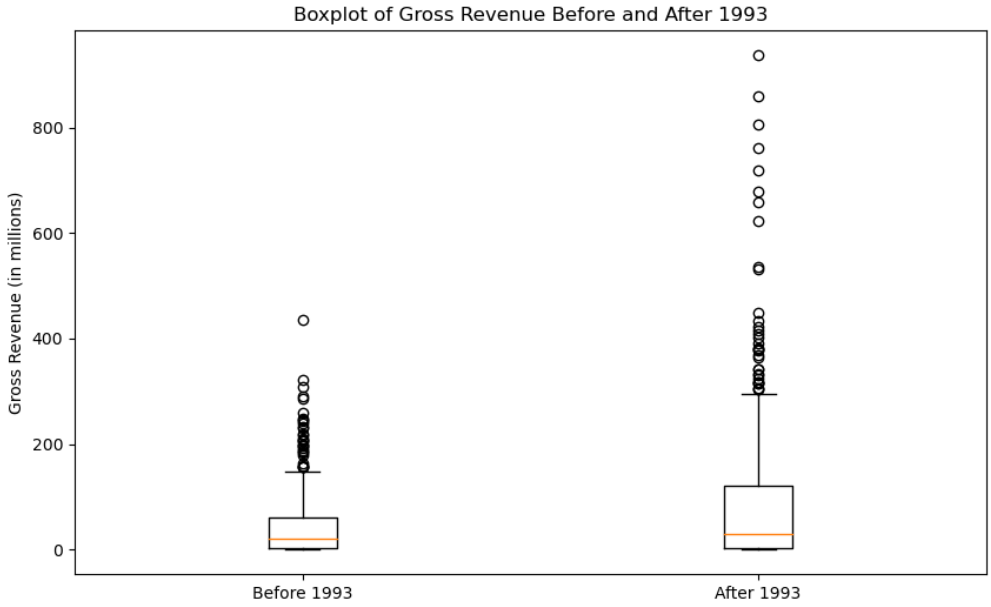

Figure 2. A boxplot showing summary statistics of gross revenue before 1993 and after 1993.

3 Part 3: Building Statistical Model

Since we have two separate samples from a population, a two-sample t-test will be used. The two-sample t-test function, "ttest_ind" from the SciPy library will be used. We will pass the data of gross revenue before 1993 and gross revenue after 1993 as two of the parameters. For the third parameter we will pass false because the variances of the samples are not equal. This will use a special test called Welch's t-test to conduct the test. Our **null hypothesis** or the default statement will be CGI does not have a significant impact on gross revenue of movies. Our **alternative hypothesis** or the statement we are trying to prove will be CGI did have an impact on the gross revenue of movies. If the resultant p-score is less than a significance level of 0.05 then we reject the null, otherwise we fail to reject the null.

After executing the code to conduct the test, these are the results:

t-statistic: -5.19

p-score: 2.58x10^-7 = 0.000000258

Mean gross revenue before 1993: $50 million

Mean gross revenue after 1993: $87 million

4 Part 4: Interpreting Results

The statement $p < 0.05$ is true after running the test. This means that we reject the null or we reject that CGI did NOT have a significant impact on gross revenue of movies. This proves our hypothesis true or that CGI did have a significant impact on the gross revenue of movies. This is also reflected in the mean gross revenue before and after 1993, with the movies being after 1993 generating more revenue.

5 Part 5: Communicating Results

The result of this analysis shows that the introduction and adoption of CGI in the film industry did have a significant impact on the gross revenue of movies. This indicates that perhaps the audience enjoys more visually appealing or interesting films that can be created using CGI. CGI also made filming much easier by eliminating the need to always have fancy sets and costumes. It gave much more freedom to directors as they were not blocked by the physical limitations of the world. By Making it easier to make movies and attracting more viewers CGI allowed the gross revenue to increase as soon as it was introduced into mainstream filmmaking. This highlights the importance of technological advancement not just in the film industry, but other industries as well to continue growing and generating revenue.

6 Resources used to achieve this goal

Python Libraries: NumPy, matplotlib, pandas, SciPy

Canvas Home Page: Homework template and Presentation on Epicycles of Analysis

7 References

- inductive_anks. (2023). *Top 1000 IMDb Movies Dataset*. Kaggle.com. https://www.kaggle.com/datasets/inductiveanks/top-1000-imdb-movies-dataset?resource=download
- FOTW. (2023, February 3). *Everything You Need to Know About CGI (Plus Examples)*. Fall off the Wall. https://falloffthewall.com/what-is-cgi/