

# Data-driven Individual Influence Analysis: A Case Study of Chinese Film Industry

Wenxuan Li<sup>a</sup>, Yinghong Ma<sup>\*a</sup>, Lixin Zhao<sup>a</sup>

<sup>a</sup>*School of Business, Shandong Normal University, Jinan 250014*

<sup>\*</sup>*Corresponding author: yinghongma71@163.com*

## ABSTRACT

With the development of China's economy and the prosperity of cultural industry, the development of film, as one of the representative products of cultural industry, has also shows an unstoppable trend. The box office of a film is a key indicator to measure the influence of a film. Therefore, a comprehensive analysis of the relevant factors of film influence can guide the healthy development of the film industry. This is of great significance to enhance the soft power of Chinese culture, carry forward Chinese culture and meet the growing cultural needs of the people. Through the statistical analysis of domestic (non animated) films with an annual box office of more than 100 million yuan released from 2015 to 2019, this paper studies the role of many factors on the influence of films by using the method of multiple linear regression. Finally, based on the experimental results, we put forward some suggestions to promote the development of individual influence in the film industry.

**Keywords:** movie box office; influencing factors; Individual influence, regression analysis

## 1. INTRODUCTION

China's film industry has achieved very rapid growth since 2003. Especially in recent years, the development momentum of China's film industry is very rapid[1]. All signs show that China's film is moving towards a deeper level of industrialization. As the core and most representative entertainment product of the cultural industry, film has the dual attributes of culture and economy. The box office is an important indicator to measure the individual influence of the film. So, what factors will boost the success of the film at the box office and improve its influence? This is the purpose of this paper.

Gaenssle s[1] Studied the influencing factors of film box office, and found that the success factors of different cultural countries are different. Zhang L[3] analyzed the influencing factors of film box office and predicted the box office by using BP neural network. Some scholars have constructed a Bayesian belief network to study the causal relationship between various film attributes in box office success prediction. The results show that the prediction accuracy of Bayesian belief network model is improved compared with artificial neural network and decision tree[4].

Although the above studies have achieved certain results, there are still some shortcomings such as lack of timeliness of sample data and incomplete factor analysis. Therefore, this article will combine domestically produced (non-animated) movies with annual gross box office above 100 million yuan released in 2015-2019, select 7 variables to study their impact on the total box office of the movie, And establish a multiple regression model to study the main factors affecting the total box office of the movie, try to make up for the deficiencies of the above studies.

## 2. DATA COLLECTION

This article selects domestic (non-animated) movies with a total box office of 100 million yuan or more in the Chinese mainland movie market from 2015 to 2019. A total of 188 movies are collected, including their total box office and first day box office, Douban score, Douban evaluation number, movie duration, director influence, starring influence, release schedule. The influence of directors and starring is measured based on the number of nominations they have received in the five major Chinese film awards. The five major Chinese awards are the Chinese Film Golden Rooster Award, Popular Film Hundred Flowers Award, Chinese Film Huabiao Award, Hong Kong Film Award and Taiwan

The Golden Horse Film Awards, they are the highest awards for audience recognition and artistic standards in the Chinese film industry. The screening period is divided into New Year's files, holiday files, and non-holiday files. The values are 2, 1, and 0, respectively. New Year's files refer to the movie periods in December and January and February of the following

year. Holiday files include summer holiday files, May Day files and National Day files.

The data comes from China's influential movie information websites, Douban.com (<https://movie.douban.com>) and China Box Office ([www.cbooo.cn](http://www.cbooo.cn)). Table 1 shows descriptive statistics for the study variables.

**Table 1.** Descriptive statistics of variables.

	Min	Max	Average	Median	Standard deviation
Total box office (ten thousand)	1	56.79	6.43	3.46	8.2
First day box office (ten thousand)	266.8	54685.2	7617.2	4767.1	8024.5
Douban score	2.8	9	5.7	5.6	1.4
Douban evaluation number	6476	1354371	198654.5	118120	212578.5
Movie duration (minutes)	86	147	112	110	12.5
Director influence	0	27	4	1	6.5
Starring influence	0	64	11	5	13.3

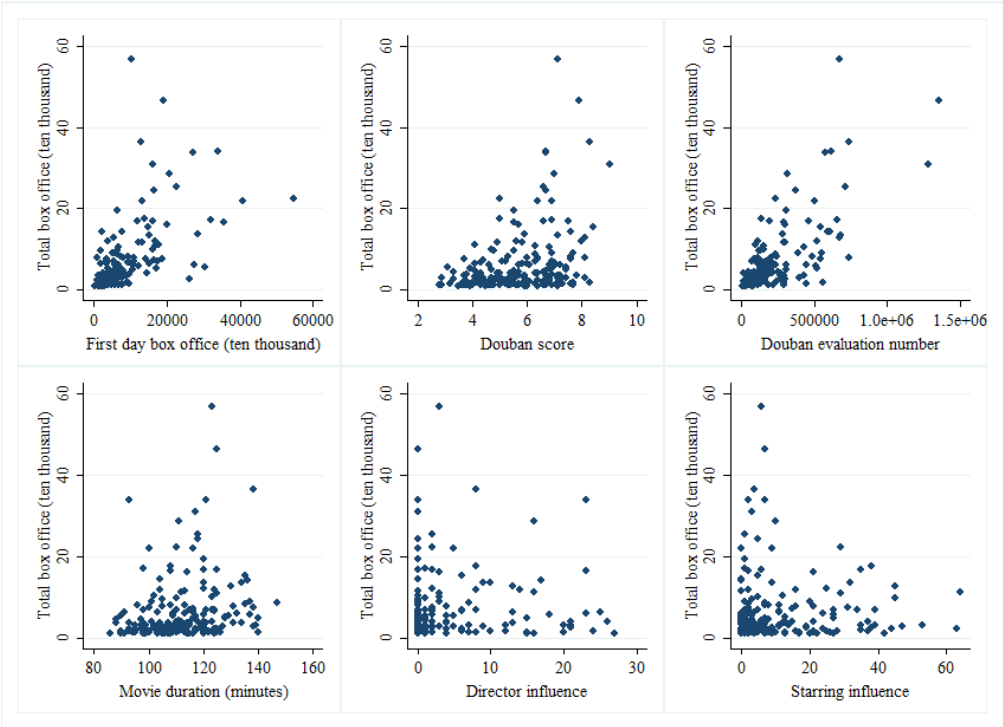
**Table 2.** Number of films per year.

Year	2015	2016	2017	2018	2019
Number of films	42	31	38	37	39

**3. VARIABLES ANALYSIS**

Draw a scatter plot of first day box office, Douban score, Douban evaluation number, movie duration,

director influence, starring influence, release schedule and total box office, and analyze the relationship between them, as shown in Figures 1 and 2.



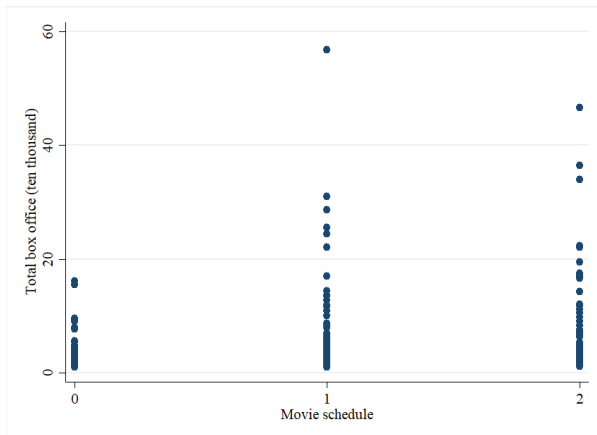
**Figure 1.** The relationship between the first day box office, Douban score, number of reviewers, movie duration, director influence, starring influence and total box office.

First day box office of most movies range from 0 to 200 million yuan. On the whole, movies with high box

office on the first day have higher total box office. There are very few movies with high box office on the first day, but their total box office is not high; or the box office on the first day is not high, but their total box office is high. The former situation may be due to consumers' strong publicity for movies, so they don't feel good after watching them. The latter situation may be that with the watching of a wavelet audience, the movie is getting more and more attention, and the total box office is rising in the middle and late stages of the movie.

Most movies have a Douban score range from 3 to 8 points. Douban scores will affect the box office of movies to a certain extent, which reflects that consumers will refer to the views of others when deciding whether to watch a movie, and will watch some higher-rated movies because of herd mentality. The number of movie evaluations is concentrated within 500,000, and Douban evaluation number is directly related to the number of people watching the movie. This result is also directly reflected in total box office of a movie. There is no special rule for the length of domestic movies. The length of most movies is distributed between 80-140 minutes, and there is no case where the movie screening time is too long or too short. The relevance of director influence, starring influence and movie total box office is not high.

There is a certain correlation between movie schedule and total box office. Of these 188 movies, the number of movies shown on the New Year's Day, Holiday, and Non-Holiday are 62, 74, and 52. The average box office in these three periods are 823 million yuan, 690 million yuan, and 364 million yuan, respectively. Most of the box office movies are released on New Year's day. The current New Year's day is considered to be the most complete, fiercely competitive, most expensive ticketing property in the country, and the strongest consumer capacity. Many movies have chosen to release in this period.



**Figure 2.** The relationship between movie schedule and total box office.

## 4. EMPIRICAL ANALYSIS

### 4.1. Multiple linear regression models

Multivariate linear regression model is a regression analysis that uses the given values of multiple eigenvectors as conditions[5]. Through the calculation of mathematical formulas, the relevant data is trained to establish a model that reflects the relationship between the input and output variables. Then, the model is used to predict according to the obtained model.

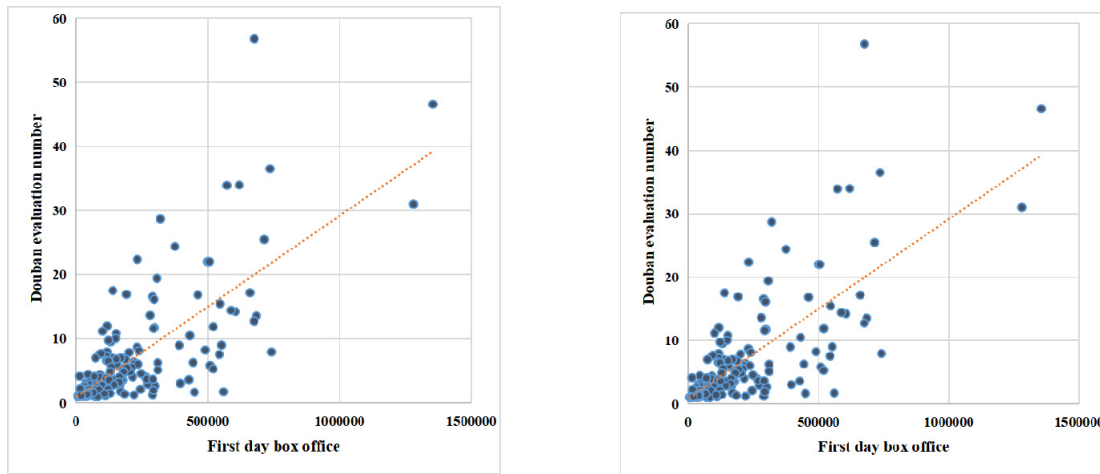
The multiple linear regression model is generally expressed by formula (1):

$$Y = X_1\beta_1 + X_2\beta_2 + K + X_n\beta_n + \varepsilon \quad (1)$$

Here,  $Y$  is the dependent variable;  $\beta_1, \beta_2, K, \beta_n$  is the regression coefficient;  $X_1, X_2, K, X_n$  is the independent variable;  $\varepsilon$  is the random disturbance term[6]. In this article, total box office (ten thousand) is the dependent variable.

### 4.2. Regression analysis

We use the Pearson correlation coefficient[7] to calculate correlations between all variables. First day box office, Douban evaluation number, and total box office have strong correlations. The correlations between Douban score, movie duration, movie schedule and total box office are weak; the relationship between director influence, starring influence and total box office are not significant. For the relationship between the independent variables, we can see that there are significant positive correlations between first day box office and Douban evaluation number, director influence, movie schedule; there are significant positive correlations between Douban score and Douban evaluation number, movie duration; there is also significant positive correlation between Douban evaluation number and movie duration. There are correlations between the independent variables, and stepwise regression[8] is used to exclude variables that cause multicollinearity [9]. Finally, we use first day box office and Douban evaluation number as independent variables to fit total box office.



**Figure 3.** Fitting diagram.

The data is sorted into stata, and the regression results are shown in Table 3:

**Table 3.** Regression results.

	Coef.	Std.Error	St.Coefficients Beta	t	Sig.
First day box office (ten thousand)	3.79	0.474	0.371	7.99	0.000
Douban evaluation number	0.23	0.018	0.594	12.80	0.000
Constant	-10045.37	5324.392		-1.89	0.061

The regression equation is summarized as follows according to Table 2:

$$Y = 3.79X_1 + 0.23X_3 - 10045.37 \quad (2)$$

Through statistical analysis of domestic (non-animated) movies with annual gross box office of more than 100 million yuan released in 2015-2019, the first day box office, Douban score, Douban evaluation number, movie duration, director influence, and starring influence are studied. The results show that there is a strong positive correlation between first day box office, Douban evaluation number, and total box office of the movie. By observing the coefficients of the regression equation after normalization, it can be found that the coefficient of the box office on the first day is 0.371 and the coefficient of the number of reviewers is 0.594. Among the above two factors, the biggest factor affecting the total box office of the movie is the number of reviewers, followed by the box office on the first day.

Douban evaluation number is directly related to the number of people watching the movie, and this result is directly reflected in the total box office of the movie. The box office on the first day of a movie's release has a positive impact on the total box office of the movie. The selection of movie schedules has strong market operation skills. It is more important for a movie to be successful at the box office.

The Douban score will affect the total box office of a

movie to a certain extent, which reflects that consumers will refer to the views of others when deciding whether to watch a movie, and will also watch some higher-rated movies because of herd mentality. However, this impact is not absolute. The audience's standards for a good movie are not exactly the same. There is no significant correlation between director influence, starring influence, and the total box office of the movie.

## 5. CONCLUSION

In view of the above analysis, in order to further increase the box office of domestic films in China and better promote the development of the film industry, domestic movies should combine contemporary movie technology to make reasonable and bold innovations to the plot, improve the structure of the story, and avoid "following the trend"; meanwhile, it is necessary to dig more themes from the real society and make more movies that reflect reality and really touch people's hearts, instead of just catering to the needs of the public and pursuing commercial interests[10]. It should be pointed out that this paper does not consider factors such as movie genre, movie ticket price, and number of movie screens. Future research will consider more variables to make the model more complete.

## REFERENCES

- [1] Hennig-Thurau T, Houston M B, Walsh G. Determinants of motion picture box office and

- profitability: an interrelationship approach[J]. Review of Managerial Science, 2007, 1(1): 65-92.
- [2] Gaenssle S, Budzinski O, Astakhova D. Conquering the box office: Factors influencing success of international movies in Russia[J]. Review of Network Economics, 2018, 17(4): 245-266.
- [3] Zhang L, Luo J, Yang S. Forecasting box office revenue of movies with BP neural network[J]. Expert Systems with Applications, 2009, 36(3): 6580-6587.
- [4] Lee K J, Chang W. Bayesian belief network for box-office performance: A case study on Korean movies[J]. Expert Systems with Applications, 2009, 36(1): 280-291.
- [5] Tranmer M, Elliot M. Multiple linear regression[J]. The Cathie Marsh Centre for Census and Survey Research (CCSR), 2008, 5(5): 1-5.
- [6] Zhang G, Liu X, Lu S, et al. Occurrence of typical antibiotics in Nansi Lake's inflowing rivers and antibiotic source contribution to Nansi Lake based on principal component analysis-multiple linear regression model[J]. Chemosphere, 2020, 242: 125269.
- [7] Benesty J, Chen J, Huang Y, et al. Pearson correlation coefficient[M]/Noise reduction in speech processing. Springer, Berlin, Heidelberg, 2009: 1-4.
- [8] Hornby T G, Henderson C E, Holleran C L, et al. Stepwise regression and latent profile analyses of locomotor outcomes poststroke[J]. Stroke, 2020, 51(10): 3074-3082.
- [9] Daoud J I. Multicollinearity and regression analysis[C]//Journal of Physics: Conference Series. IOP Publishing, 2017, 949(1): 012009.
- [10] Gao W, Lam K M, Chiu D K W, et al. A Big Data Analysis of the Factors Influencing Movie Box Office in China[M]//Intelligent Analytics With Advanced Multi-Industry Applications. IGI Global, 2021: 232-249.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

