

基于对数回归的电影票房影响因素分析

陈静羽, 李瑞雪, 黄月池, 史江兰

贵州大学数学与统计学院, 贵州 贵阳

收稿日期: 2024年9月20日; 录用日期: 2024年10月15日; 发布日期: 2024年11月28日

摘要

电影越来越成为一种潮流的象征, 日益丰富着我们的日常生活。此次案例分析的数据为截至目前中国大陆上映的100部电影的票房数据及其影响因素, 选取首日票房、上映首周票房、档期、题材、评分(猫眼)、平均票价、播放形式、电影时长、预告片播放量(亿)、想看人数(万人)、演员影响力、导演影响力、预算(万元)这13个指标作为电影票房的影响因素。首先, 我们对数据进行预处理, 对处理好的数据进行分类描述性统计分析, 再对自变量与因变量进行相关性分析, 通过分析发现, 想看人数和首日票房之间存在较强的相关性, 故在回归建模时删掉这两个变量; 其次, 我们使用逐步回归模型进一步筛选变量, 使用对数回归模型探索对电影票房具有显著影响的指标; 最后, 我们得到档期、题材、播放形式、上映首周票房、评分、平均票价、预告片播放量这7个变量对票房有显著影响。较高票房的两类电影是动作片和喜剧片; 导演影响力对于电影票房不存在很大的影响; 票房较高的电影得分也较高, 得分基本在9分左右, 但并不是得分越高的电影票房也就越高。

关键词

电影票房, 逐步回归模型, 对数回归模型

Analysis of Influencing Factors of Movie Box Office Based on Logarithmic Regression

Jingyu Chen, Ruixue Li, Yuechi Huang, Jianglan Shi

School of Mathematics and Statistics, Guizhou University, Guiyang Guizhou

Received: Sep. 20th, 2024; accepted: Oct. 15th, 2024; published: Nov. 28th, 2024

Abstract

Cinema is becoming more and more a symbol of a trend, enriching our daily lives day by day. The data of this case analysis is the box office data and its influencing factors of the 100 films released in Chinese mainland so far, and selects 13 indicators such as the box office on the first day, the box

office of the first week of release, the schedule, the theme, the score (Maoyan), the average ticket price, the broadcast form, the movie duration, the number of trailers played (100 million), the number of people who want to watch (10,000 people), the influence of actors, the influence of directors, and the budget (10,000 yuan) as the influencing factors of the movie box office. Firstly, we preprocessed the data, performed a categorical descriptive statistical analysis on the processed data, and then performed a correlation analysis between the independent variable and the dependent variable, and found that there was a strong correlation between the number of people who wanted to watch and the box office on the first day, so these two variables were deleted in the regression modeling. Secondly, we used a stepwise regression model to further screen the variables, and used a logarithmic regression model to explore the indicators that had a significant impact on the box office. Finally, we get that the 7 variables of schedule, theme, broadcast format, box office in the first week of release, rating, average ticket price, and trailer playback have a significant impact on the box office. The two categories of films that gross higher are action and comedy; The director's influence does not have a great impact on the box office; Movies with higher box office scores are also higher, with a score of around 9 points, but it is not that movies with higher scores will also have higher box office.

Keywords

Movie Box Office, Stepwise Regression Model, Logarithmic Regression Model

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 背景介绍

随着现代化建设的发展,中国城乡居民收入水平显著提高,精神文化生活日益成为消费者的重要追求。曾经少数人能享受的影视作品,如今已成为大众消费品,遍布各大中小城市,成为人们精神生活的寄托和谈资,电影作品逐渐成为潮流的象征,丰富着日常生活。

近年来,我国电影产业快速发展,全产业链不断完善,制作和营销能力显著提升,为票房增长奠定了基础。政府的扶持政策及经济的快速发展也为电影产业创造了良好的环境。2012年,中国电影票房首次超过日本,成为全球第二大电影市场,仅次于美国。然而,电影产业仍属于高风险行业,投资亏损时有发生。因此,研究票房影响因素对影视行业的繁荣至关重要,也有助于增强中华文化的国际竞争力和国家文化软实力。随着影视作品数量的快速增长,竞争日益激烈,制片公司必须关注票房表现。了解票房影响因素、把握市场偏好、制作大众喜爱的电影,已成为制片公司的当务之急。

2. 数据分析

2.1. 指标设计

通过文献查找,我们提炼出13个可能与电影票房相关的因素[1]-[3]:首日票房、上映首周票房、档期、题材、评分(猫眼)、平均票价、播放形式、电影时长、预告片播放量(亿)、想看人数(万人)、演员影响力、导演影响力、预算(万元)。

其中,档期、题材和播放形式为分类变量,采用单因素分析来判断其对电影票房的影响。其余10个因素为数值型变量,我们将首先通过相关性分析来评估每个因素对电影票房的影响程度,随后运用逐步回归法进一步筛选变量,最终建立对数-对数回归模型。

符号说明如表 1 所示。

Table 1. The explanation of symbols
表 1. 符号说明

符号	定义	类型
Y	累计票房	数字
X_1	上映首周票房	数字
X_2	评分(猫眼)	数字
X_3	平均票价	数字
X_4	电影时长	数字
X_5	预告片播放量	数字
X_6	演员影响力	数字
X_7	导演影响力	数字
X_8	预算	数字
X_9	想看人数	数字
X_{10}	首日票房	数字
X_{11}	档期	字符串
X_{12}	题材	字符串
X_{13}	播放形式	字符串

2.2. 数据来源

通过猫眼专业版(网址：<https://piaofang.maoyan.com/rankings/year>)选取截止到目前在中国大陆上映累计票房前 100 名的电影。

2.3. 分类变量描述性统计分析

档期、题材、播放形式为分类变量，将通过单因素分析来判断其是否对电影票房产生影响。

2.3.1. 档期

根据档期对 100 部电影的数据进行分析，发现大多数电影集中在普通时段上映，春节档和其他节日档上映的电影数量较多，而剩余节日的电影仅有 14 部。从均值来看，春节档的平均票房最高，其次是其他节日档和暑假档，普通档的平均票房最低。如表 2 所示。

2.3.2. 题材

根据影片类型对数据进行拆分和分析。题材分布数据显示，动作类电影占据 30%，剧情和喜剧类型分别占 27%和 26%。相对而言，科幻、悬疑、犯罪的影片占比较少。在“想看人数”这一变量中，科幻类影片的关注人数最多，其次是悬疑类影片。详见表 3。

2.3.3. 播放形式

对播放形式这一因子进行分析时可以看出，大多数电影的播放形式是 IMAX-2D、IMAX-3D。如表 4 所示。

影响票房的影响因素中，档期、题材、播放形式这三个变量属于分类变量，对这些变量进行多因素方差分析，研究不同档期、不同题材、播放形式是如何影响电影票房的，结果如表 5。

Table 2. Schedule descriptive statistics
表 2. 档期描述统计

档期	变量	N	最小值	最大值	均值	标准偏差
普通档	累计票房	34	11.2700	42.500	16.757941	7.1913487
	评分(猫眼)	34	7.9	9.6	9.012	0.4212
	预告片播放量	34	0.01387	1.89000	0.6120779	0.47114963
	想看人数	34	2.8	185.0	46.626	45.8821
	有效个案数(成列)	34				
暑假档	累计票房	14	12.4300	59.159686	25.159686	13.6921748
	评分(猫眼)	14	7.5	9.7	9.114	0.5723
	预告片播放量	14	0.14970	6.17000	1.4657757	1.90578790
	想看人数	14	16.5	62.1	38.686	14.3120
	有效个案数(成列)	14				
春节档	累计票房	26	11.9800	54.1372	27.100662	12.9545034
	评分(猫眼)	26	7.7	9.6	8.954	0.5715
	预告片播放量	26	0.04930	5.27000	0.9037662	1.27365604
	想看人数	26	1.8	446.9	67.292	85.0036
	有效个案数(成列)	26				
节日档	累计票房	26	11.6900	57.7524	19.880862	10.4272849
	评分(猫眼)	26	8.3	9.8	9.192	0.3475
	预告片播放量	26	0.07436	2.42000	0.6251158	0.59756707
	想看人数	26	2.7	142.5	40.162	32.7353
	有效个案数(成列)	26				

Table 3. Descriptive statistics of the subject matter
表 3. 题材描述性统计

题材	变量	N	最小值	最大值	均值	标准偏差
剧情	累计票房	27	11.2700	57.7524	20.832311	11.3783605
	评分(猫眼)	27	2.7	114.6	42.204	28.6885
	预告片播放量	27	0.01387	3.84000	0.6795137	0.81918656
	想看人数	27	8.2	9.8	9.215	0.4400
	有效个案数(成列)	27				
动作	累计票房	30	11.4900	56.9456	18.525520	9.7069364
	想看人数	30	3.7	185.0	43.150	39.3140
	预告片播放量	30	0.07436	6.17000	0.7859767	1.09498176
	评分(猫眼)	30	7.5	9.7	8.997	0.4796
	有效个案数(成列)	30				
喜剧	累计票房	26	11.6000	54.1372	23.082585	11.1147912
	想看人数	26	1.8	9.6	56.242	84.3466
	预告片播放量	26	0.04930	5.27000	0.9532335	1.18212995
	评分(猫眼)	26	7.7	446.9	8.885	0.5182
	有效个案数(成列)	26				

续表

科幻	累计票房	4	14.6100	46.8700	29.687500	16.2937910
	想看人数	4	25.2	145.1	80.975	61.2479
	预告片播放量	4	0.30411	2.16000	0.8190025	0.89847330
	评分(猫眼)	4	8.6	9.3	9.025	.3096
	有效个案数(成列)	4				
悬疑	累计票房	2	35.2300	45.4400	40.335000	7.2195602
	想看人数	2	32.3	109.9	71.100	54.8715
	预告片播放量	2	0.29961	0.34154	0.3205750	0.0264899
	评分(猫眼)	2	9.2	9.4	9.300	.1414
	有效个案数(成列)	2				
犯罪	累计票房	2	12.7300	38.4800	25.605000	18.207996
	想看人数	2	12.5	114.0	63.250	71.7713
	预告片播放量	2	0.26819	0.41336	0.3407750	0.10265069
	评分(猫眼)	2	8.9	9.1	9.000	0.1414
	有效个案数(成列)	2				

Table 4. Descriptive statistics of the playback format
表 4. 播放形式描述统计

档期	变量	N	最小值	最大值	均值	标准偏差
2D	累计票房	11	11.2700	54.1372	19.889745	12.0441032
	评分(猫眼)	11	8.2	9.8	9.000	0.5459
	预告片播放量	11	0.04930	2.42000	0.9781582	0.81009817
	想看人数	11	2.3	116.2	51.500	40.2438
	有效个案数(成列)	11				
3D	累计票房	6	11.7300	59.9456	24.552600	17.8529788
	评分(猫眼)	6	8.4	9.7	9.050	0.5648
	预告片播放量	6	0.13456	6.17000	1.7126250	2.33109182
	想看人数	6	2.8	61.4	30.300	22.3571
	有效个案数(成列)	6				
IMAX2D	累计票房	48	11.2700	57.7524	22.782758	11.5102661
	评分(猫眼)	48	8.2	9.7	9.194	0.3322
	预告片播放量	48	0.07436	5.27000	0.6198035	0.82396017
	想看人数	48	5.5	446.9	52.558	65.5442
	有效个案数(成列)	48				
IMAX3D	累计票房	35	11.5800	50.3500	19.538571	9.8877456
	评分(猫眼)	35	7.5	9.6	8.891	0.5559
	预告片播放量	35	0.01387	5.19000	0.8656109	1.08205048
	想看人数	35	1.8	185.0	47.131	44.8756
	有效个案数(成列)	35				

Table 5. Analysis of variance
表 5. 方差分析

源(因变量: 累计票房/亿)	III 类平方和	自由度	均方	F	显著性
修正模型	7665.358 ^a	45	170.341	1.760	0.024
截距	19306.935	1	19306.935	199.441	0.000
档期	1448.174	3	482.725	4.987	0.004
题材	1573.009	7	224.716	2.321	0.038
播放形式	806.833	3	268.944	2.778	0.050
档期*题材	1711.223	11	155.566	1.607	0.123
档期*播放形式	109.025	7	15.575	0.161	0.992
题材*播放形式	943.238	7	134.748	1.392	0.228
档期*题材*播放形式	267.039	4	66.760	0.690	0.602
误差	5227.486	54	96.805		
总计	58839.847	100			
修正后总计	12892.844	99			

由于题材、档期、播放形式的概率 P 值小于显著性水平 α ，可认为档期、档期对累计票房产生显著影响，播放形式影响次之。

2.4. 回归建模变量描述性统计分析

从 2.3 节可以看出题材、档期、播放形式对累计票房具有显著影响，本小节我们对用于回归建模的剩余 10 个变量进行描述性统计分析。

2.4.1. 变量直方图

从图 1 我们可以看出，累计票房、首日票房、上映首周票房和预告片播放量(亿)、想看人数(万人)的分布呈右偏形式，评分(猫眼)、平均票价、电影时长近似正态分布，而演员影响力、导演影响力、预算(万元)的分布较无规律，特别是预算几乎集中在一个值，说明我们搜集的数据大部分是合理的，少部分存在瑕疵。

2.4.2. 变量描述分析

从表 6 可以看出，累计票房的最大值为 57.75 亿元，最小值为 11.27 亿元，总体标准差为 11.41，前 100 名的票房差距较为明显；评分最高的有 9.8 分，最低的有 7.5 分，但均值为 9.0 分，说明前 100 的电影评分都比较高；电影时长的均值在 127 分钟左右，此外，导演影响力和演员影响力的最大值和最小值之间的差距也比较大。

2.4.3. 相关性分析

1) 自变量与因变量的相关性

将累计票房按中位数划分为高低两组，从图 2 可以看出，累计票房高的电影在首日票房、首周票房、评分(猫眼)、预告片播放量(亿)和平均票价上均较高。而在剩余的 4 个变量(电影时长、演员影响力、导演

影响力和预算(万元))中, 高低票房之间没有明显差异。

2) 自变量间的相关性

根据自变量间相关性检测, 发现想看人数、首日票房与其他变量之间存在较强相关性, 故在回归建模时, 排除这 2 个变量, 用剩余 8 个变量建立回归模型(图 3)。

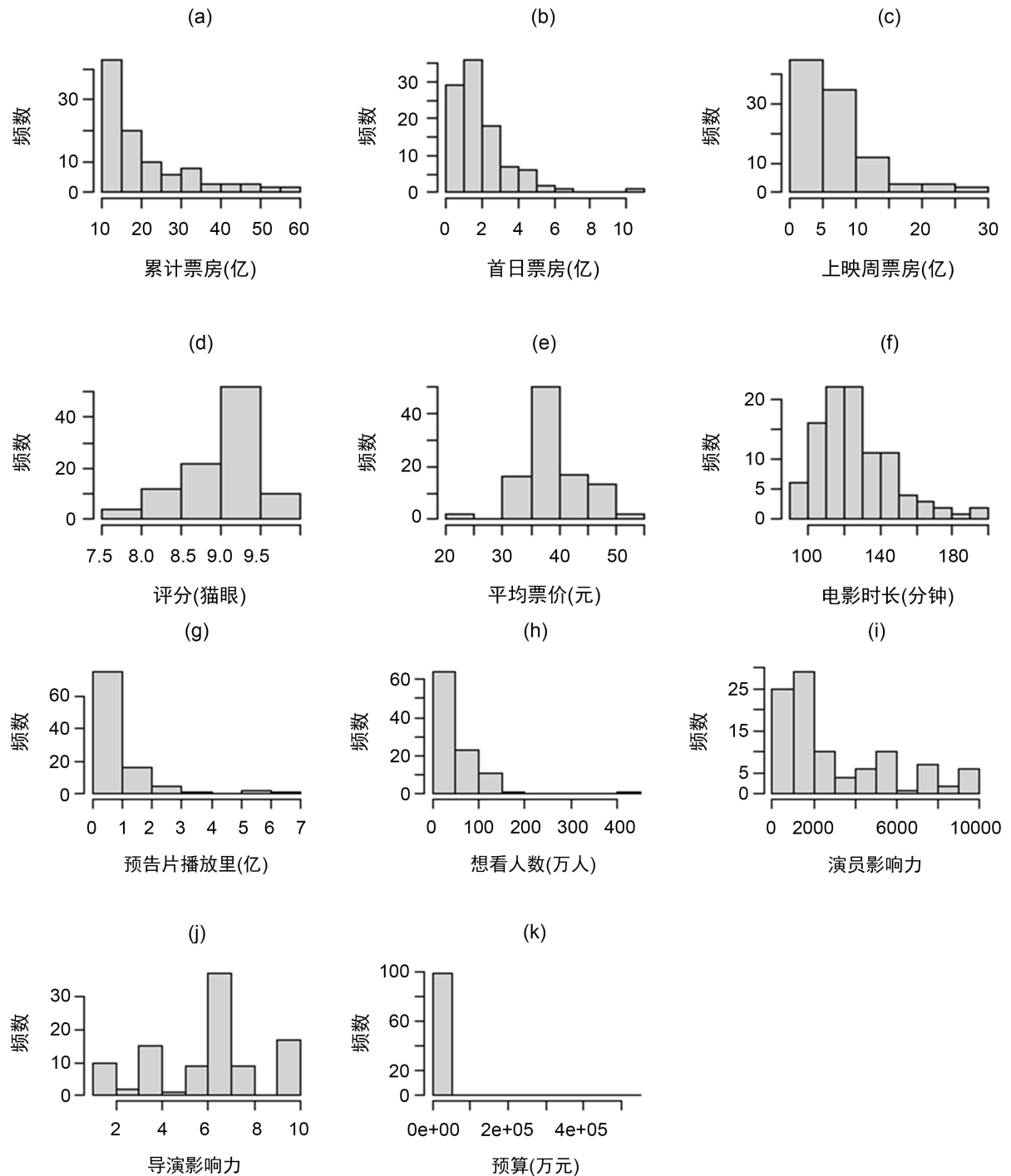


Figure 1. Histograms of each variable
图 1. 各个变量直方图

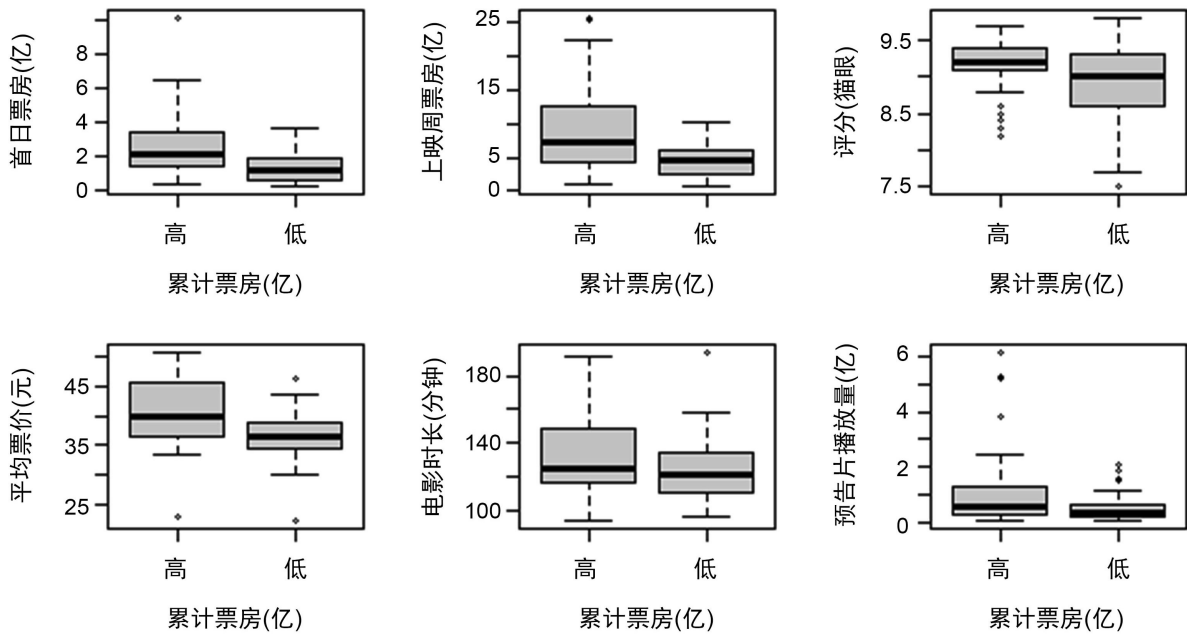
Table 6. Descriptive statistical analysis
表 6. 描述性统计分析

变量	N	最小值	最大值	均值	标准偏差
累计票房	100	11.2700	57.7524	21.435252	11.418688
首日票房	100	0.12792	10.11000	1.9077433	1.53615495
评分	100	7.5	9.8	9.058	0.4732
上映首周票房	100	0.54824	25.79000	6.9765330	5.28008305
平均票价	100	22.056992	50.791748	38.65400415	5.289758527
电影时长	100	93	194	127.79	20.906
想看人数	100	1.8	446.9	49.207	54.3066
预告片播放量	100	0.01387	6.1700	0.8108244	1.06597350
导演影响力	100	1	10	6.36	2.584
预算	100	140	503500	10884.55	50168.219
演员影响力	100	0	9652	3032.39	2815.408
有效个案数	100				

3. 回归模型理论

3.1. 逐步回归模型(Stepwise Regression)

逐步回归是一种在构建回归模型时，通过逐步选择或删除自变量的统计方法。其目的是找到对因变量有显著影响的自变量，从而构建最优回归模型。该方法逐个引入自变量，每次选择对因变量影响最显著的变量，并对现有自变量进行显著性检验，剔除不显著的变量。最终，回归方程中既包含对因变量显著影响的自变量，又剔除不显著的自变量。主要有三种方法：前进法(FORWARD)、后退法(BACKWARD)和逐步筛选法(STEPWISE) [4]。



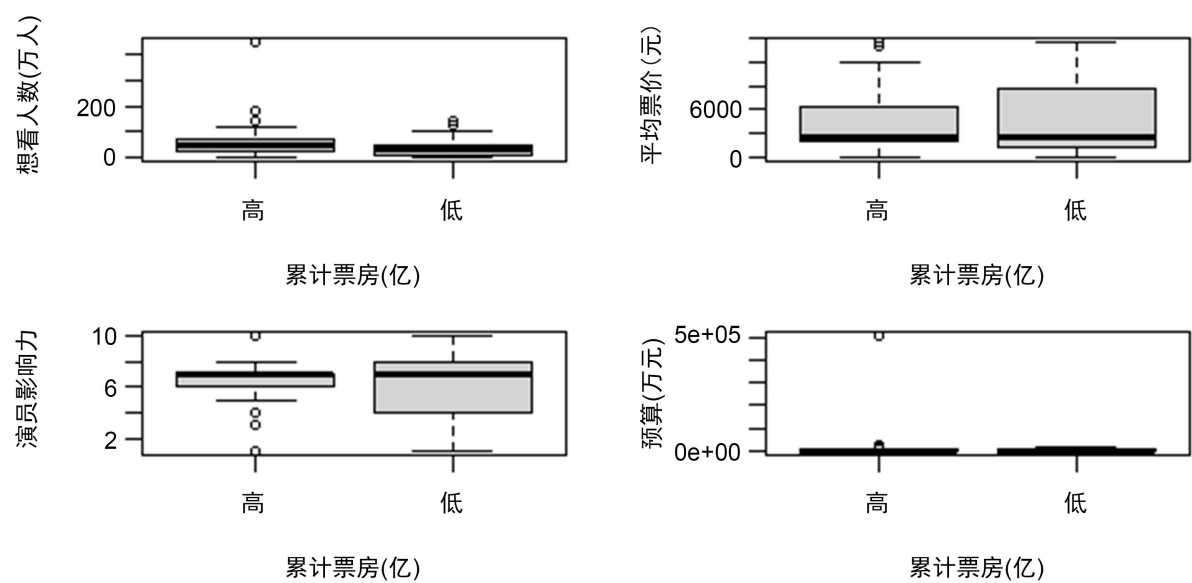


Figure 2. Box plot of independent and dependent variables
图 2. 解释变量与因变量箱线图



Figure 3. Correlation relationship among independent variables
图 3. 自变量间相关关系

3.2. 对数回归模型

对数回归模型是一种通过对因变量或自变量进行对数变换以建立变量之间线性关系的回归分析方法。简单的二元线性模型可以有四种涉及对数的变换组合[5]:

- 1) 线性情况：没有对任何变量进行变换，采用标准的线性回归模型；

- 2) 线性 - 对数模型: 自变量保持线性形式, 而因变量进行对数变换;
- 3) 对数 - 线性模型: 因变量保持线性形式, 而自变量进行对数变换;
- 4) 对数 - 对数模型: 对因变量和自变量均进行对数变换。

3.3. 模型检验

3.3.1. QQ 图检验

学生化残差 $r_i (i=1, 2, \dots, n)$ 的正态 QQ 图的基本步骤如下[6]:

- 1) 将学生化残差 r_1, r_2, \dots, r_n 由小到大排序为 $r_{(1)}, r_{(2)}, \dots, r_{(n)}$, 则各 $r_{(i)}$ 是数据的 i/n 分位数;
- 2) 对每个 $i=1, 2, \dots, n$, 计算标准正态分布 $N(0,1)$ 的 i/n 分位数 $q_{(i)}$, 即

$$q_{(i)} = \Phi^{-1}\left(\frac{i-0.375}{n+0.25}\right), i=1, 2, \dots, n \quad (1)$$

其中 $\Phi(x)$ 表示标准正态分布的分布函数, 而 0.375 和 0.25 是常用的修正量;

3) 在直角坐标系中描出每个点 $(q_{(i)}, r_{(i)}) (i=1, 2, \dots, n)$, 则此散点图被称为学生化残差的正态 QQ 图; 基此 QQ 图, 若散点 $(q_{(i)}, r_{(i)}) (i=1, 2, \dots, n)$ 明显不在一条直线上, 则有理由怀疑残差(从而误差)服从正态分布的合理性; 否则, 则没有理由拒绝误差分布的正态性假设。进一步, 我们可以基于数据点 $r_{(i)} (i=1, 2, \dots, n)$ 和 $q_{(i)} (i=1, 2, \dots, n)$ 之间的相关系数:

$$\hat{\rho} = \frac{\sum_{i=1}^n (r_{(i)} - \bar{r})(q_{(i)} - \bar{q})}{\sqrt{\sum_{i=1}^n (r_{(i)} - \bar{r})^2 \sum_{i=1}^n (q_{(i)} - \bar{q})^2}} \quad (2)$$

来度量二者之间线性关系的强弱, 其中 $\bar{r} = \frac{1}{n} \sum_{i=1}^n r_{(i)}$, $\bar{q} = \frac{1}{n} \sum_{i=1}^n q_{(i)}$ 。若 $\hat{\rho}$ 比较接近于 1, 则说明散点 $(q_{(i)}, r_{(i)}) (i=1, 2, \dots, n)$ 大致在一条直线上。

3.3.2. Cook 距离

著名的统计学家库克(D.R. Cook)想出了一个非常巧妙的办法。他说可以为每一个样本打分, 给影响力特别大的样本打高分, 给影响力不大的样本打低分, 然后就可以根据分值高低判断谁是“捣蛋鬼”。该打分后来被人们称作 Cook 距离(Cook's distance), 其定义如下[7]:

$$D_i = \frac{(\hat{\beta} - \hat{\beta}_{(-i)})' (X'X) (\hat{\beta} - \hat{\beta}_{(-i)})}{(p+1)\hat{\sigma}^2} \quad (3)$$

式中, $X = (X_1, X_2, \dots, X_n)'$ 是设计矩阵(design matrix)。可以通过对比不同 D_i 的相对大小, 对各个样本的影响力做出大概判断。如果发现, 有少数一两个样本的 Cook 距离特别大, 那么应该考虑将此类样本剔除后, 重新拟合回归分析。将新的结果同全样本结果做详细对比, 再结合经验做出主观判断。

4. 模型应用与结果

在描述分析的基础上, 将进一步分析因变量对电影票房之间的关系。首先将数据按 80% 训练集和 20% 测试集划分, 在训练集上进行逐步回归, 以挑选合适的变量[8]。

逐步回归结果如表 7 所示。根据 AIC 准则[9], 我们发现某些变量对 AIC 的贡献不大, 因此决定剔除变量(电影时长)和(导演影响力)。尽管 P 值显示的某个变量的重要性较低, 但在逐步回归中未被自动筛选出来, 因此我们将手动删除该变量, 以优化模型。

Table 7. Stepwise regression process
表 7. 逐步回归过程

步骤	自由度	RSS	AIC	选择的变量
0	8	2912.9	305.59	$X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8$
1	7	2914.1	303.62	$X_1, X_2, X_3, X_4, X_5, X_6, X_8$
2	6	2988.3	303.64	$X_1, X_2, X_3, X_5, X_6, X_8$

在训练集上建立对数 - 对数线性回归模型：

$$\log(y) = 0.234 \log(X_1) + 2.756 \log(X_2) + 0.870 \log(X_3) + 0.101 \log(X_5) + 0.101 \log(X_8) - 6.604 \quad (4)$$

相关参数及检验结果如表 8 所示。模型整体的 F 检验高度显著(P 值 < 0.001)，表明至少有一个解释变量与因变量显著相关，同时调整后的 R² 为 0.523。在 T 检验结果中，除了最后一个解释变量预算，其余系数在 1% 的显著性水平下均显著不为零，预算的 P 值为 0.982，显示高度不显著。

Table 8. Log-Log regression model results
表 8. 对数-对数回归模型结果

变量名称	回归系数	标准误差	P 值	方差膨胀因子
截距项	-6.604	1.657	<0.001	----
上映首周票房	0.234	0.053	<0.001	1.516
评分(猫眼)	2.756	0.676	<0.001	1.030
平均票价	0.870	0.279	<0.001	1.291
预告片播放量	0.101	0.032	<0.001	1.062
预算(万元)	0.101	0.028	0.982	1.217
模型全局检验	P 值 < 0.001	----	调整的 R ²	0.523

此外，五个解释变量的 VIF 值均小于 2，这意味着不存在多重共线性问题，因此可以排除共线性导致预算不显著的可能性。模型检验结果如图 4 所示，表明模型具有良好的拟合效果。

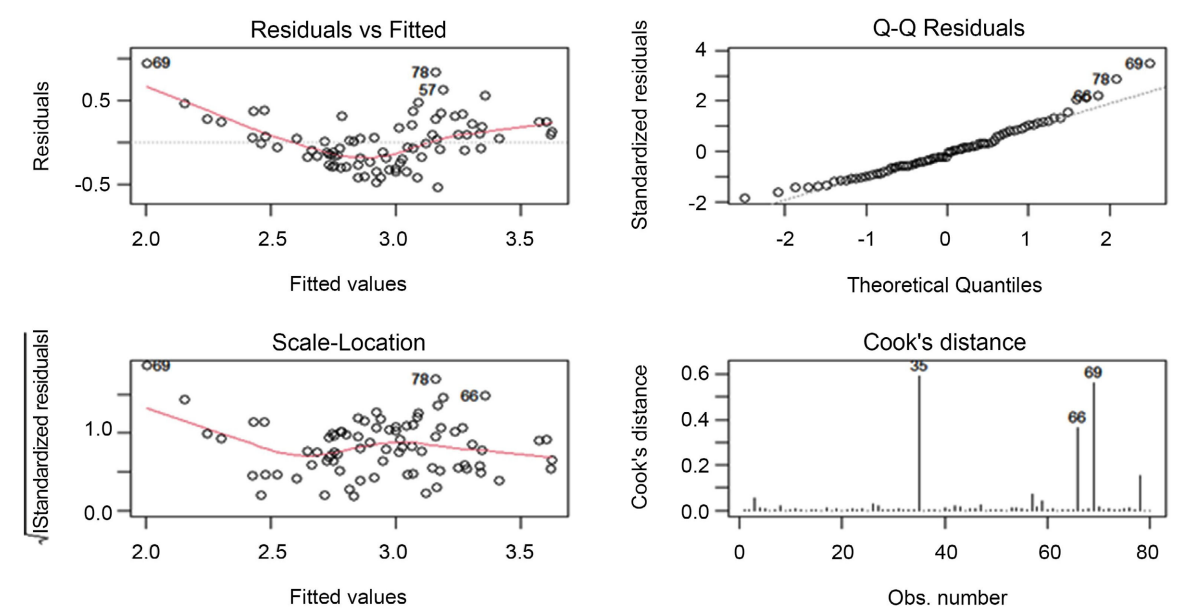


Figure 4. Test results
图 4. 检验结果

最后在测试集上，根据 AIC 和 BIC 准则做模型选择，结果如表所示模型：

$$\log(y) = 0.234\log(X_1) + 2.767\log(X_2) + 0.870\log(X_3) + 0.100\log(X_5) - 6.601 \quad (5)$$

其中 AIC 和 BIC 的回归结果一致，都只保留了上映首周票房、评分(猫眼)、平均票价和预告片播放量(亿)，而 P 值过高的预算则被抛弃。这个结果与前面的 T 检验结果一致，均表明上映首周票房、评分(猫眼)、平均票价和预告片播放量(亿)都是高度显著的变量，而预算不显著。在控制其他变量不变的情况下，可以得出结论：上映首周票房越高、评分(猫眼)越高、平均票价越高和预告片播放量(亿)越高的电影，其票房也越高。这进一步印证了这些因素对票房的积极影响(表 9)。

Table 9. AIC and BIC regression model results
表 9. AIC 和 BIC 回归模型结果

变量名称	AIC 回归系数	P 值	BIC 回归系数	P 值
截距项	-6.601	<0.001	-6.601	<0.001
上映首周票房	0.234	<0.001	0.234	<0.001
评分(猫眼)	2.767	<0.001	2.767	<0.001
平均票价	0.870	<0.001	0.870	<0.001
预告片播放量	0.100	<0.001	0.100	<0.001
预算(万元)	-----	-----	-----	-----
模型全局检验	P 值 < 0.001		P 值 < 0.001	
调整的 R ²	0.529		0.529	

5. 结论与建议

5.1. 结论

综上所述，我们发现档期，题材，播放形式，上映首周票房、评分(猫眼)、平均票价、预告片播放量(亿)这 7 个变量对票房具有显著影响。题材方面：累计票房前 100 的电影中动作和喜剧电影居多；播放形式方面：3D 和 2D/IMAX 电影有票房优势；档期方面：春节档居多且票房高。导演粉丝数对于电影票房不存在很大的影响，不知名的导演一样可以拍出票房过亿的电影；票房较高的电影得分也较高。

5.2. 建议

合适的档期选择至关重要，避免与其他大片竞争，并利用节假日和观影热情高涨的时段以提升票房。题材方面，根据观众口味和市场需求选择热门题材或与当前社会热点相关的内容可更容易吸引观众。播放形式上，考虑 IMAX、3D 等选项可以吸引更多观众，进一步提高票房。首周票房是评判电影口碑和热度的重要指标，需通过宣传和营销提升。合理控制平均票价，以兼顾吸引观众和保证收入，过高的票价可能影响观众选择。此外，制作精良的预告片并广泛传播可增加观众关注，提高期待感。最后，提升影片质量，确保剧本、制作和演员水平，并利用社交媒体及影评人促进正面评价，及时回应观众反馈以解决负面评论，对于票房的成功至关重要。

参考文献

- [1] 于兰婷. 影响国产电影票房的因素分析[J]. 中国电影市场, 2021(10): 17-23.
- [2] 刘志新. 中国电影票房影响因素分析[J]. 合作经济与科技, 2019(17): 114-116.

-
- [3] 程粮君. 电影票房影响因素分析——以 2016-2017 年票房过亿元的国产电影为例[J]. 声屏世界, 2018(4): 37-41.
- [4] 杜久升, 赵贝贝, 侯争. 基于逐步回归的学习行为与成绩评估模型研究[J]. 测绘通报, 2023(S2): 148-151.
- [5] Benoit, K. (2011) Linear Regression Models with Logarithmic Transformations. *London School of Economics, London*, **22**, 23-36.
- [6] 魏艳华, 王丙参, 张艺馨. 利用蒙特卡罗方法对 QQ 图检验的改进与比较[J]. 统计与决策, 2020, 36(16): 13-17.
- [7] 霍伟光, 曹静杰, 陈雪, 等. 基于 Cook 距离的阻尼多道奇异谱分析分离绕射波[J]. 石油地球物理勘探, 2024, 59(4): 771-781.
- [8] 王鹏, 李斌, 李佳伦, 等. 基于对数函数的岩石三轴强度回归预测模型[J]. 矿业研究与开发, 2023, 43(4): 103-109.
- [9] 田密, 熊自民. 基于 MARS 与 AIC 准则的泥石流冲出距离数据驱动预测方法[J/OL]. 武汉大学学报(工学版), 2024: 1-11. <http://kns.cnki.net/kcms/detail/42.1675.T.20230828.0924.002.html>, 2024-07-01.