应用统计学实践

电影票房与获奖因素分析

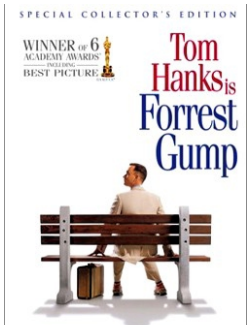
小组成员：叶森淼 黄金凤 刘昱雯 许晓晴 丁元俊

研究框架

1. 问题背景及研究目标
2. 数据收集及处理
3. 电影票房的影响因素分析（回归分析）
4. 电影获奖的影响因素分析（因子分析）
5. 我们的结论

背景介绍

电影是人类文明史上最伟大的发明之一，是近代科技大发展的产物。电影不但是现代人最重要的娱乐方式之一，而且已经成为体现现代人类最高精神活动和审美理想的重要艺术载体。



背景介绍

✓美国电影市场：1894年4月，爱迪生用他的电影放映设备——“电影视镜”开始了商业性放映活动，这是美国电影史的开端。从80年代中期直到今天，美国电影逐渐重新称霸世界影坛，并利用雄厚的资金，将高科技带入电影制片，为已有百年历史的电影带来了一种新的语言。

✓中国电影市场：电影发明以后，很快传入中国。1896年8月11日，上海徐园第一次放映了电影。60年代以来，为了繁荣电影事业，先后举办百花奖和金鸡奖的评选活动，还在长春、上海等地举办了国际电影节。其中有不少优秀影片在权威性国际电影节上获奖。

分析对象及其目标

✓近年来，世界电影向高科技、高投入的趋势发展，我们小组将用1981年至2006年美国每年票房排名前三和2003至2006年每年中国电影内地票房排名前十的电影说明电影的发展趋势，进而作出描述和分析。

✓我们还将针对上面提到的78部美国电影的奥斯卡获奖情况作出分析，说明不同因素的影响程度。

数据来源

电影行业具有一定特殊性，给数据搜集带来了很大困难。
我们的主要数据来源于以下几个网站：

●美国权威电影杂志《首映》网站：http://www.premiere.com

●北美票房统计专业网站：http://www.boxofficemojo.com

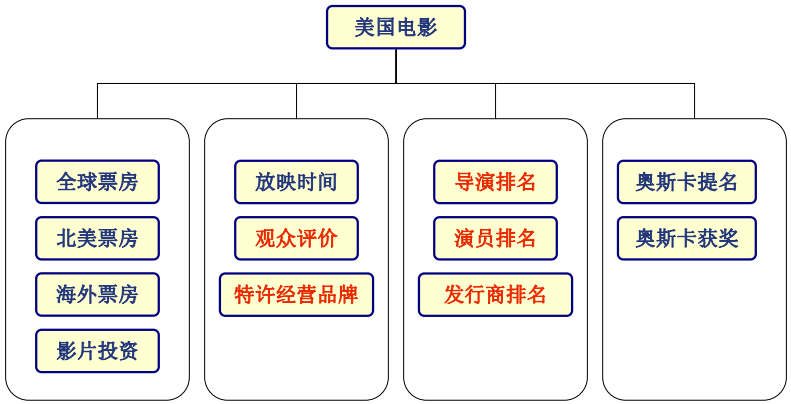
●美国演员、导演统计网站：http://www.the-members.com

前人综述：

Stanley Rosen，Chinese Cinema in the Era of Globalization: Prospects for Chinese Films on the International Market, with Special Reference to the United States，[J] Contemporary Cinema，2007

数据框架

在此仅以美国电影的数据框架为例：



框图中红色的表示分类数据，其他为数值型数据

数据处理

我们主要针对分类数据进行数据处理，大体分为两个步骤：将文本转换为可量化的数据，再将其转化为满足正态分布的数值型数据。

演员评价标准

演员排名	1—10	11—50	51—100	101—199	200
演员得分	5	4	3	2	1

注：1. 我们取影片演员阵容中排名最靠前的两位名次得分求和
2. 动画片的总分均为6分
3. 演员排名依照好莱坞权力榜，见附录1
4. 没有上榜的演员统一按第200名计

数据处理

我们主要针对分类数据进行数据处理，大体分为两个步骤：将文本转换为可量化的数据，再将其转化为满足正态分布的数值型数据。

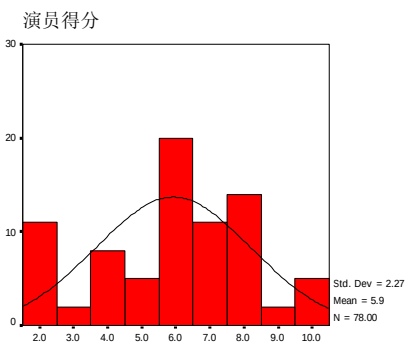
导演评分标准

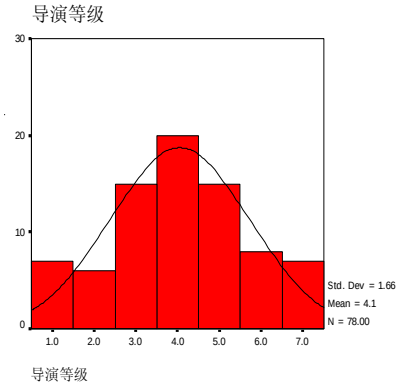
导演排名	1	2—3	4—9	10—30	31—90	91—149	150
导演得分	7	6	5	4	3	2	1

注：1. 导演排名依照http://www.the-members.com的导演总排名，见报告附录2
2. 没有上榜的导演统一按第150名计

数据处理

演员和导演的正态分布图





电影票房的影响因素分析（回归分析）

- 采用强制进入策略

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.835(a)	.698	.653	171.0969960

调整判定系数为0.653，说明电影票房变差中有65.3%可以用它与放映时间，特许经营，发行商，演员得分，上映季度，观众评价，投资，导演等级，获奖数，提名数的线性关系来解释

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4532065.162	10	453206.516	15.481	.000(a)
	Residual	1961370.197	67	29274.182		
	Total	6493435.359	77			

通过F检验

电影票房的影响因素分析（回归分析）

Coefficients[a]

没有通过t 检验

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-286.245	144.335		-1.983	.051
	投资	3.507	.447	.607	7.837	.000
	提名数	-12.545	10.759	-.156	-1.166	.248
	获奖数	59.982	16.165	.468	3.711	.000
	观众评价	42.327	30.491	.121	1.388	.170
	导演等级	14.229	14.243	.081	.999	.321
	发行商	6.372	15.263	.034	.417	.678
	演员得分	.146	9.147	.001	.016	.987
	特许经营	49.102	47.252	.084	1.039	.302
	上映季度	-25.370	23.955	-.078	-1.059	.293
	放映时间	14.816	4.577	.239	3.237	.002

a. Dependent Variable: 全球票房

电影票房的影响因素分析（回归分析）

- 采用逐步进入策略（消除多重共线性）

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.819(a)	.670	.652	171.2669439	1.940

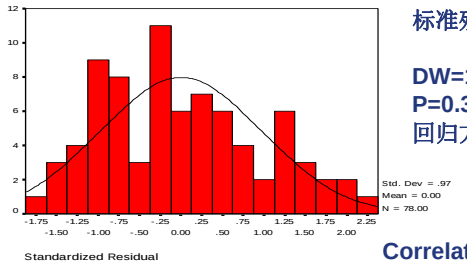
ANOVA [b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4352173	4	1088043.159	37.094	.000[a]
	Residual	2141263	73	29332.366		
	Total	6493435	77			

Coefficients [a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-304.422	115.408		-2.638	.010
	投资	3.717	.398	.644	9.343	.000
	获奖数	40.628	8.726	.317	4.656	.000
	导演等级	26.023	12.175	.149	2.137	.036
	放映时间	16.754	4.189	.271	3.999	.000

残差分析



标准残差值总体服从以零为均值的正态分布；（假设1满足）

DW=1.940，约为2；（假设2满足）

P=0.325>0.05，方差相同；（假设3满足）

回归方程：

$$Y = -304.422 + 3.717X_1 + 40.628X_2 + 16.754X_3 + 26.023X_4$$

Correlations

			Standardized Predicted Value	Standardized Residual
Spearman's rho	Standardized Predicted Value	Correlation Coefficient	1.000	-.113
		Sig. (2-tailed)	.	.325
		N	78	78
	Standardized Residual	Correlation Coefficient	-.113	1.000
		Sig. (2-tailed)	.325	.
		N	78	78

回归分析——探索（观众评价）

Model Summary(b)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.824(a)	.678	.656	170.3369700	1.885

a Predictors: (Constant), 放映时间, 特许经营, 获奖数, 投资, 观众评价

b Dependent Variable: 全球票房

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4404378.157	5	880875.631	30.360	.000(a)
	Residual	2089057.202	72	29014.683		
	Total	6493435.359	77			

回归模型——探索（观众评价）

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-306.469	115.190		-2.661	.010
	投资	3.679	.402	.637	9.146	.000
	获奖数	41.231	9.360	.321	4.405	.000
	放映时间	13.253	4.422	.214	2.997	.004
	特许经营	78.228	41.816	.134	1.871	.065
	观众评价	47.974	26.588	.137	1.804	.075

在a=0.1的显著性水平下，通过了检验。

回归方程如下：

$$Y = -306.469 + 3.679X_1 + 41.231X_2 + 47.974X_3 + 78.228X_4 + 13.253X_5$$

中国——回归分析

LOGO

Model Summary(b)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.896(a)	.804	.787	2524.098	1.641

a Predictors: (Constant), 投入（万元）
b Predictors: (Constant), 投入（万元）, 观众评价
c Predictors: (Constant), 投入（万元）, 观众评价, 导演级别

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	9384390904.874	3	312813031.625	49.099	.000(a)
	Residual	2293584709.501	36	6371068.875		
	Total	1167797574.375	39			

中国——回归分析

LOGO

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-3496.537	1505.734		-2.322	.026
	观众评价	1439.238	460.304	.263	3.127	.003
	导演级别	1228.790	559.612	.215	2.196	.035
	投入（万元）	.390	.069	.607	5.620	.000

a Dependent Variable: 国内票房总收入(万元)

Y=-3496.537+1439.238X1+1228.790X2+0.39X3

电影获奖的影响因素分析(因子分析)

获奖分析思路

选取78部美国电影中获得奥斯卡奖的38部电影作为研究对象

采用因子分析的方法归纳出影响获奖的主要因素并算出影片综合得分

将因子分析的结果与获奖实际情况对比进而得出结论

电影获奖的影响因素分析(因子分析)

因子分析的合适性检验

通过Bartlett检验，输出结果如下。
可知Sig=0<0.001，适合做因子分析

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.479
Bartlett's Test of Sphericity	Approx. Chi-Square	121.185
	df	45
	Sig.	.000

电影获奖的影响因素分析(因子分析)

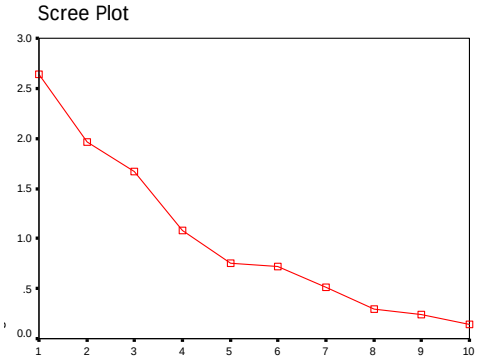
求解相关矩阵的特征值和方差贡献率

因子	特征值	贡献率(%)	累计贡献率(%)
1	2.635	26.354	26.354
2	1.959	19.588	45.941
3	1.668	16.677	62.618
4	1.085	10.854	73.473
5	.749	7.488	80.960
6	.715	7.154	88.114
7	.518	5.181	93.295
8	.296	2.960	96.255
9	.238	2.378	98.633
10	.137	1.367	100.000

从表中可以看出，有四个特征值大于1，因此取四个主成分，并且可以解释原数据10个变量总方差的73.5%，可以认为解释了原数据的大部分信息，基本符合要求。

电影获奖的影响因素分析(因子分析)

碎石图



从输出的碎石图中，也可以直观地看到，即前4个因子对解释变量的贡献最大，所以因子分析中提取4个因子最合适。

电影获奖的影响因素分析(因子分析)

旋转后的因子载荷矩阵

	Component			
	1	2	3	4
导演等级	**.821**	-.100	.101	-.334
演员得分	.699	.105	-.021	.028
观众评价	.684	.209	.102	.439
特许经营	.653	-.227	.267	.114
提名数	.044	**.859**	.020	.115
发行商	-.231	-.727	.303	.086
上映季度	-.341	.707	.198	.028
投资	.055	-.072	**.944**	-.091
全球票房	.247	.071	.831	.334
放映时间	-.008	.012	.094	**.942**

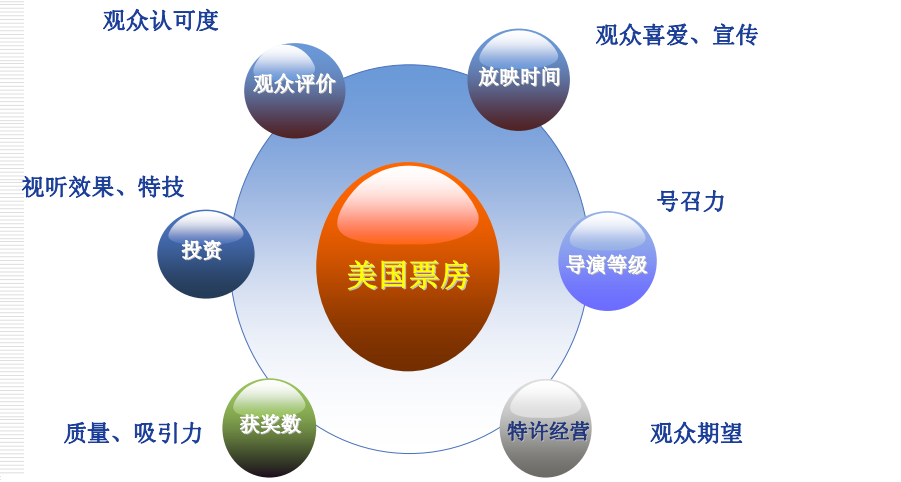
重新命名的四个因子为：
影片特质
影片影响
影片投入与收益
影片宣传

电影获奖的影响因素分析(因子分析)

因子得分

	Component			
	1	2	3	4
全球票房	.018	.053	.433	.124
投资	-.070	.021	**.580**	-.215
提名数	.019	**.456**	.036	.006
观众评价	.291	.081	-.065	.295
导演等级	**.375**	-.012	.024	-.298
发行商	-.144	-.387	.156	.098
演员得分	.322	.057	-.083	-.005
特许经营	.267	-.116	.066	.052
上映季度	-.174	.390	.190	-.066
放映时间	-.041	-.078	-.088	**.740**

我们的结论



Thank You !