

# 灾难面前，人性何去何从？

## ——基于泰坦尼克号乘客生存数据分析

strength library and muscle laboratory（第七组）

陈立国 樊可 韩思越 黄永晟 杨远琨 俞铖昊

### 一、背景介绍

灾难，是人性与文明的试金石。

1912年泰坦尼克号沉船事件轰动了一整个世纪。美国票房纪录数据显示，电影《泰坦尼克号》全球累计票房高达21.87亿美元，仅烂番茄影评平台上，该电影影评量便超过3500万条。同时，谷歌学术不完全统计显示，涉及泰坦尼克号沉船事故人性分析的伦理学社会学研究论文超过11.6万篇。尽管人们纷纷津津乐道于这场灾难中，揭示的人性与文明的善恶，但受到信息局限，评论者往往并未或很难充分透彻地了解泰坦尼克号真实灾难情况，使得论断与事实真相常常大相径庭。

在《泰坦尼克号》影评领域，影评者常常乐于对电影情节与现实事件发表真知灼见，然而他们面对着海量亦真亦假的网络转述转载信息，真假掺杂的信息往往易引导无知的评论者误入歧途，层层掩盖事件真相。在伦理学研究领域，研究者也常常使用泰坦尼克号真实事件，来论说灾难面前人道与人本原则，刻画灾难面前人性善恶，但却往往苦于没有确切的第一手资料，亦或难以在第一手数据资料密密麻麻的表格记录中，提取有效信息辅助观点佐证。

因此，我们对泰坦尼克号乘客信息与生存状况数据进行分析。以是否幸存于泰坦尼克号沉船灾难为因变量，乘客性别，年龄，家庭成员数量等因素为自变量，建立模型进行分析。力图挖掘冰冷数据背后的情感，意义与温度，由客观数据入手尽可能还原泰坦尼克号沉船事件真相。《泰坦尼克号》影评者得到的，是一份具备真实数据佐证的泰坦尼克号事件分析，由此能够基于事件真相发挥合理影评见地。伦理学，社会学研究者与文学家能够得到的，是一份基于泰坦尼克号事件第一手资料分析提炼而成的有效信息总结，由此便可基于事件真相提供更具说服力的观点论证。广大受众也都能够从这份基于历史真相的分析解读中，观得泰坦尼克号灾难与绝望面前，人性与文明所彰显的善恶与真实。

## 二、数据介绍

本案例选取 891 条 1912 年泰坦尼克号遇难乘客信息用于描述与模型分析。为便于后续分析，我们提取姓名变量中“Mr. Mrs.”等专有特殊称谓作为定性变量，定义乘客“在船兄弟姐妹数+配偶数+乘客在船父母+子女数”为乘客“家庭成员数量”作为自变量。

预处理后用于下文描述分析与模型分析的变量情况如表 1 所示。

表 1 变量表

变量类型		变量名	详细说明	取值范围与备注
因变量		是否存活	定性变量 共 2 个水平	是，否
自变量	个人信息	性别	定性变量 共 2 个水平	男，女
		年龄	连续变量	0 ~ 80（单位：岁）
		称谓	定性变量 共 5 个水平	Master, Miss, Mr, Mrs, Other
	亲属信息	家庭成员数量	定性变量 共 4 个水平	1（没有亲属陪伴），2（一位亲属），3，4，>4
	船票信息	客舱等级	定性变量 共 3 个水平	1（一等舱）， 2（二等舱）， 3（三等舱）
		船票价格	连续变量	0 ~ 512 （单位：美金）
		登船港口	定性变量 共 3 个水平	Q(Queenstown), C(Cherbourg), S(Southampton)

三、描述分析

本次案例分析关注的因变量是泰坦尼克事故中是否最终存活。由图 1 事故存活情况柱状图中可以看出，本数据集中超过半数的人未能够逃脱死亡的厄运。尽管数据集内并未涵盖所有泰坦尼克号船员信息，但与泰坦尼克号事故实际死亡人数比例相近。可以如此理解，泰坦尼克号事故是一场“仅允许部分人能够生存”的灾难，那么在这场灾难下，人性与人类文明又将做出什么样的抉择呢？

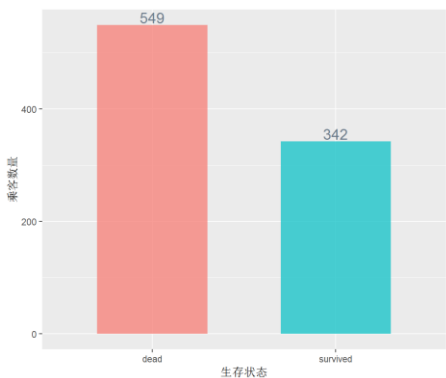


图 1 泰坦尼克事故存活情况柱状图

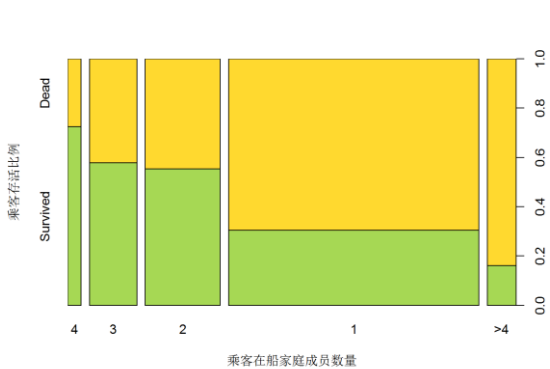


图 2 乘客家庭成员数量与存活比例棘状图

我们逐一看各个自变量对于泰坦尼克事故中最终存活的影响情况。先从乘客个人信息，乘客家庭成员数量信息入手。如图 2 所示，只有一位家庭成员（即没有亲友陪伴）的乘客，幸存率不到 40%，而家庭成员超过一位（即有亲友陪伴）的乘客，幸存率接近 60%。我们推测有亲友陪伴的乘客，当灾难发生时，多多少少会得到亲友一定程度的陪伴与支撑，从而坚定了生存下去的选择与信念。可见，灾难面前，来自亲友的支持与鼓励，能够给人以生存的信念与信心。

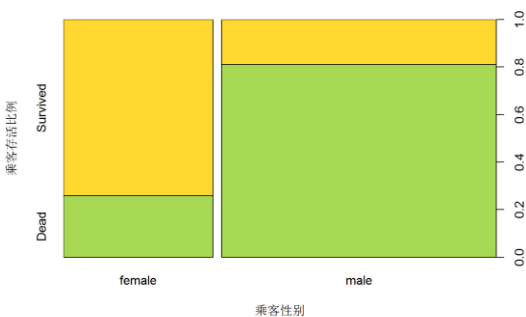


图 3 乘客性别与存活比例棘状图

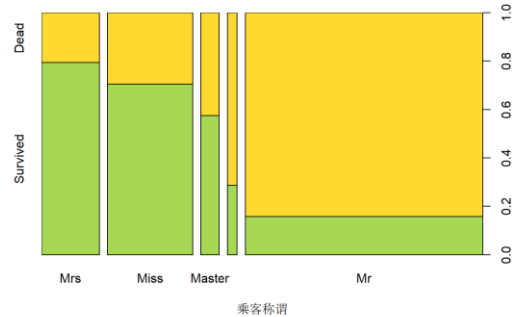


图 4 乘客称谓与存活比例棘状图

然后，我们考察性别与是否幸存的联系。从图 3 乘客性别与存活比例棘状图可以明显看

出，尽管男性在数量上多于女性，但女性的幸存率高达 80%，接近男性幸存率（大约 20%）的四倍。尽管一般情况下女性相对男性更为弱势，在这场灾难下，作为弱势群体的女性得到了更多出于人性与人道的优先救助保护。可见伦理学研究者常常论述的灾难面前，人类社会对于女性的人道主义关怀与关注，确有事实佐证。接着，我们深入考察图 4 中不同称谓的存活情况，可以发现社会地位较高男士（Master）幸存率居然超过 60%，远高于普通男性乘客（Mr.），一方面可能处于当时社会人普遍对于社会名流的尊重，使得社会名流优先获得了救助，另一方面也可能是这些达官贵人出于利己主义而不择手段地运用名利谋求获救的优先权。

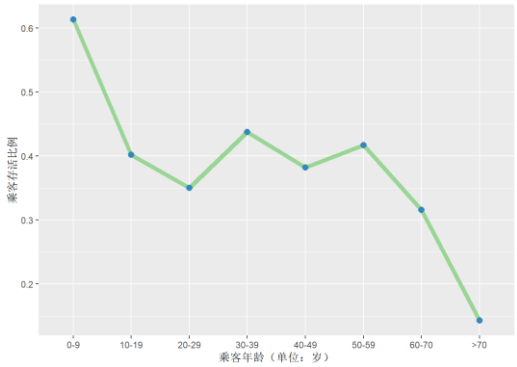


图 5 乘客年龄与存活/死亡数量折线图

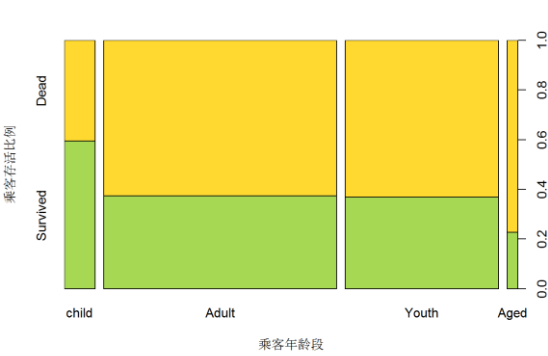


图 6 乘客年龄与存活比例棘状图

然后我们从乘客年龄信息着手。由图 5 图 6 可以看出，超过 60%的 0-9 岁年龄段的儿童乘客得到了救助，而其他 10-80 岁各个年龄段的乘客获救率普遍低于 50%。尽管数据集中唯一一位最高龄 80 岁的 Algernon Henry Wilson 获救，但高于 70 岁的年迈乘客获救率仍然不到 20%。我们推测年迈乘客本身更希望将求生的机会，留给拥有更多人生时光与可能性的年轻乘客，而儿童乘客，既作为具备最多人生生涯与希望的群体也作为弱势群体，得到了更多出于人性与人道的优先救助保护。这也表明，电影《泰坦尼克号》当中优先让妇女儿童上救生艇，优先考虑弱者生命的人道关怀情节，具备一定的事实依据。可见，无论是妇女还是儿童，人类文明社会，在灾难面前，弱者得到了更多的人伦与人道关怀。



图7 客舱等级与存活比例堆叠图

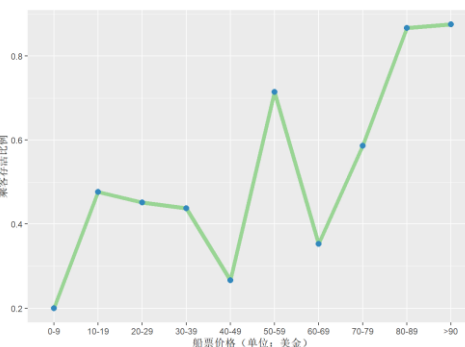


图8 船票价格与存活比例折线图

最后再看经济因素如何影响乘客存活情况。如图7所示，最昂贵的一等舱内的乘客幸存比例高达60%，幸存率最高，而最廉价的三等舱内的乘客存活比例却刚过20%。图8也反映出，船票价格小于9美金的贫困乘客，幸存比例仅20%，随着船票价格增加至超过80美金，乘客幸存率逐步上升并超过80%。可见，无论客舱等级还是船票价格，反映出较为富裕的乘客有着明显较高的幸存率。我们猜测，富裕人群具备高幸存率，并非出于人道与人本的关注与保护，而更可能的是富人为一己私利，而通过金钱手段为己谋生（例如泰坦尼克号灾难事件内，部分富人使用金钱预先购买救生艇座位，更有甚者担心救生艇超载而花钱额外购买救生艇空座，宁可少救人而确保自身生命安全）。可见，灾难面前，部分富裕人群仍然不免会显现出自私与利己主义的一面，利用名利自保而不顾他人生命饱受践踏。

#### 四、逻辑回归建模分析

考虑到许多《泰坦尼克号》影评者，喜欢滔滔不绝于猜测自己置身泰坦尼克号上的幸存情况，伦理学与社会学的研究，也常常需要模拟大型灾难面前，人类文明出于人性会做出的善恶抉择，现实主义文学艺术创作者，也希望构建出符合现实情境的大型灾难文学艺术作品。我们尝试使用逻辑回归模型，试图预测各类人群在泰坦尼克号上，或是在类似大型灾难面前的幸存情况，为《泰坦尼克号》影评者，伦理学研究者，文学艺术创作者提供模型预测依据。

经过 AIC, BIC 变量选择，九折交叉验证操作后，绘制的 ROC 曲线如图9所示。其中 ROC 曲线与坐标轴围成的面积（AUC）为0.86。同时模型显示，在假正例率（FPR）为0.20的情况下，对应真正例率（TPR）为0.81。可见，逻辑回归模型能够较好地反映出灾难面前，各类型

人群的生存情况，即灾难面前出于人性的抉择。

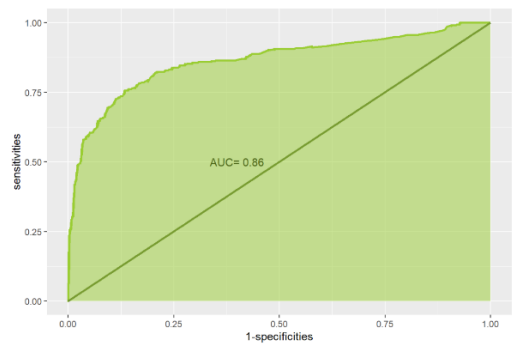


图 9 泰坦尼克号数据逻辑回归 ROC 曲线图

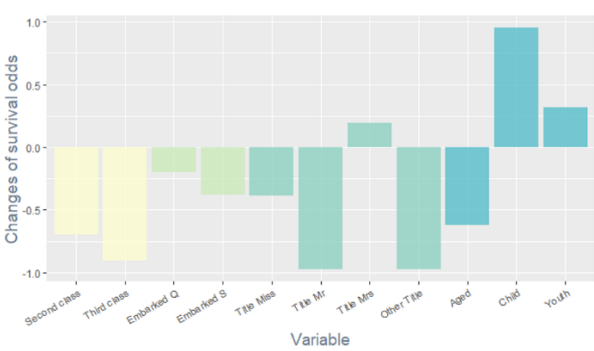


图 10 逻辑回归自变量回归系数柱状图

于此同时，图 10 显示了该逻辑回归模型中各变量的系数比较，由此我们发现：

1、年龄：婴幼儿年龄段（child）乘客生存率最高，相比中青年（adult）年龄段高了超过 75%，而老年年龄段乘客生存率最低，相比中青年（adult）年龄段低了超过 50%。模型表明，普遍的大型灾难面前，年迈人群考虑到本身人生将近，会倾向于最先让渡出自己的一份求生机会，人类文明会主动优先选择，将生存希望托付于最具人生可能性的婴幼儿群体。

2、客舱等级：二等舱乘客生存率比一等舱低了超过 50%，而三等舱乘客生存率比一等舱低了超过 75%。由此显示，大型灾难面前，仍然存在相当一部分的利己主义与自私面。至少很大一部分达官贵族，仍会优先自我中心，利用不合人道的金钱赎买等方式优先确保自身生存，而置其他受害者于不顾。

## 五、回归树建模分析

接着，我们再尝试使用的 CART 决策树模型，随机森林模型以及提升树模型对该数据集进行建模，模型预测结果如表 2 所示。

表 2 CART 决策树、随机森林、提升树、逻辑回归的预测对比表

	CART 决策树	随机森林	提升树	逻辑回归
真正例率（TPR）	0.81	0.80	0.80	0.81
假正例率（FPR）	0.17	0.20	0.23	0.20

对比逻辑回归可以发现，逻辑回归能够在真正例率大于随机森林与提升树预测结果的情况

下，具备更低的假正例率，因此逻辑回归优于随机森林与提升树。而在真正例率为 0.81 的情况下，逻辑回归的假正例率却高于 CART 决策树，因此逻辑回归在该问题上的预测结果不如 CART 决策树模型。最终比较下来，CART 决策树模型预测效果最佳。

同时，借助提升树模型，我们能够看出各个自变量对生存情况的影响程度。如图 11 所示，与性别因素高度相关的称谓变量对是否生存的影响最明显。可见灾难面前人类文明仍然优先出于共同善考虑弱势群体的女性，而非出于利己的自私因素，这部分人性伦理善的一面，超越了与利己主义自私的一面，而成为灾难面前文明的优先抉择。

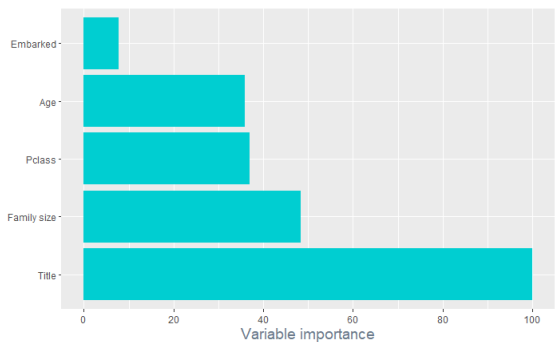


图 11 泰坦尼克号数据提升树变量重要性图

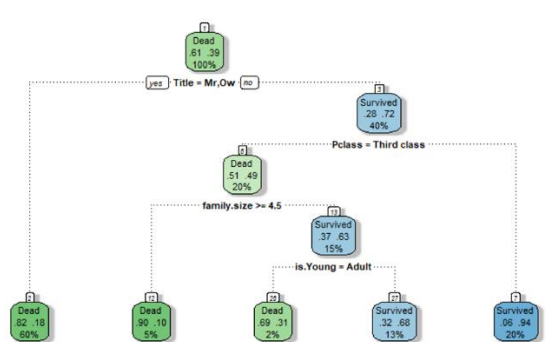


图 12 泰坦尼克号数据 CART 决策树结果图

接着，我们进一步利用我们训练得到的 CART 决策树模型，预测类似大型意外灾难发生时，一等舱的 Rose 女士，和三等舱的 Jack 先生的生存可能性有多高。如图 12 所示，通过 CART 决策树模型，我们预估在总生存率为 40%左右的情况下，Rose 女士的生存率可以高达 94%，而 Jack 先生的生存率仅为 18%。由此，我们可以为每一位《泰坦尼克号》影评者，预测自己在泰坦尼克号上的情形结局，并且为伦理学，社会学研究者，文学家等模拟大型灾难场面上，文明社会出于人性抉择可能情景提供参考。

## 六、结论与建议

为尽可能还原泰坦尼克号沉船事件历史真相，探讨灾难与绝望面前，人性与文明所彰显的善恶本质，为《泰坦尼克号》影评者，伦理学研究者等提供第一手资料分析，我们对泰坦尼克号乘客信息与生存状况数据使用了逻辑回归模型，CART 决策树模型，随机森林模型和提升树模型分别进行比较分析。

分析结果显示，泰坦尼克号灾难当中，女性相比男性，儿童相比成年人往往生存率更高，有家庭成员陪伴的乘客一般情况下生存率也更高，这两者分别体现了人类文明对于弱势群体的入道主义关怀，与亲友鼓励陪伴带来的生存信念，体现人性与文明的光辉面。而富人的生存率往往高于其他人，则也显示出富出于利己动用金钱，而不顾他人生命的人性丑陋面。

由此，我们能够為《泰坦尼克号》影评者提供如下建议：

- 1、《泰坦尼克号》电影中刻画妇女与儿童优先上救生艇的情节场景并非虚构，现实事件中确实存在对于弱者人伦与人道关怀的共识。
- 2、《泰坦尼克号》电影中富人通过金钱霸占救生艇座位，为一己私利，不顾他人死活的情节设定，同样有现实数据印证。
- 3、影评者同样可以构想，如果自己与另一半在泰坦尼克号上，会是如何的生存结果，由此为影评发挥提供进一步思路。

我们也为伦理学，社会学研究者，文学家提供如下参考：

- 1、基于泰坦尼克号沉船事件中妇女生存率与儿童生存率的第一手资料的数据描述分析，由此更易于基于我们的描述分析结果对入道主义与道德准则的深入挖掘与探索。
- 2、基于泰坦尼克号沉船事件中富人与相对贫困者的生存情况第一手资料的数据描述分析，由此为基于利己主义等人性面的深入挖掘提供资料分析参考。
- 3、泰坦尼克号类型事件灾难面前，文明社会不同人的可能生存情况预估，为模拟大型灾难场面上文明社会出于人性抉择可能情景提供参考。

同时，对于广大读者，也能够从中得到一定的启发：人性在灾难面前，既存在善的一面，存在普遍与共识的人文主义与人道关怀，但也存在部分恶的一面，仍有少数群体存在利己主义，倾向一己私利而践踏他人生命。

附：小组分工情况

变量预处理：全员讨论变量处理方式，俞铖昊负责代码

描述分析：全员讨论呈现形式，陈立国、黄永晟负责代码

模型分析：杨远琨负责逻辑回归模型，樊可负责回归树模型

分析报告：全员讨论整合细节，韩思越负责最终整合