

共进专栏：心中有数

- 我们统计工作主要的精神便是从纷繁复杂的数据之中摸索出有用的讯息，并赋予其生命意义。
- 在教授统计分析课程时，我会特别强调对数据背景的了解，这和统计分析方法几乎是同等重要。
- 一个成功的案例应该有上、中、下游。上游收集背景资料以提出有意义的问题；中游利用统计方法来解读、推敲答案；下游给予统计结果赋予生命，提供有意义的结论以方便领导决策。



（美国）林共进 / 文

数字工作者经常犯的一个毛病，就是常把实际数据当作代数问题来处理。反正所有数据丢入电脑软件包中，跑出一堆有用没用的统计量，再胡乱解释一通。需知道“数字本身是没有意义的”。我们统计工作主要的精神便是从纷繁复杂的数据之中摸索出有用的讯息，并赋予其生命意义。所以数据本身的背景资料是非常重要的。

在教授统计分析课程时，我会特别强调对数据背景的了解，这和统计分析方法几乎是同等重要。毕竟我们是在解决问题，而不是在作代数习题。铁达尼号(Titanic)沉船资料常常是我的第一个案例分析。学生必须上网去查询所有可能得到的背景资料，然后提出问题。最后数据才呈上台面，使用统计方法来解答这些问题。当然统计并非万事皆通，有些问题可以解答，有些则不行。以铁达尼号为例，学生搜集的背景资料包括：造船历史背景、船体资料、乘客背景、航线与沉船经过、以及事后发展。有了这些资料，这才进一步着手了解数据、根据数据，正式开始统计分析。其基本架构详述如下。

从铁达尼号 沉船资料谈起

前言

铁达尼号搭乘的乘客详细资料当中分别针对每个乘客的属性作记载，我们在其中看到了一些特别有趣的现象，我们针对这笔数据提出这样的问题：乘客资料的各项属性（年龄、性别、所在舱等）中，何者是影响生存与否最重要的因素，是否有特定的属性下生还机率明显不同的情况，我们认为这应该也是历史学家极欲了解的资讯。

历史背景

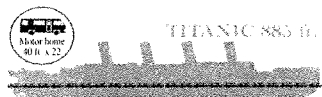
铁达尼是十九世纪人类科技与工艺的登峰造极之作，同进也是为了满足当时人类的自信与奢华而打造的昂贵产

物：她可以说是一个图腾，一个因工程科学的进步而给人类带来幻想与希望的巨大海上图腾，如同巴黎圣母院、埃及金字塔以及中国万里长城那样的伟大与不朽。在她长眠大西洋底近一世纪后的今日，人们依旧歌颂着她所谱写出的传奇乐章，遥想铁达尼号给予当时各社会阶层乘客所带来的美梦与幻想，牢记着她带入海底昂贵的历史教训，缅怀着长伴她左右超过一千五百名的罹难乘客不朽的精神。

英国的白星航运公司的超级豪华邮轮，由汤姆士·安德鲁处理设计，她的船长882英尺9英寸，宽度为92.5英尺，吃水线以上至甲板的高度为60.5英尺。排水量6万6千吨，三个推进器，最大航速

24~25节。有20个救生艇,可容纳1178人。总重46329吨(净重21831吨),动力3000匹马力,有16个防水隔舱,双层船底,因此被认为是不沉之船。船上装潢极具堂皇,可称为融汇了当时科技及精湛艺术结晶的“海上浮宫”。1911年5月31日下水,经过数月的试航后,1912年4月10日作首次航行,船长为E.J.史密夫,搭载了当时社会上不同阶层的乘客共二千多人,于是日正午离开了南汉普顿港,预定渡过北大西洋,直达美国纽约。

铁达尼号属英国白星海运公司所有,当时为了与对手吉娜海运公司的摩尼速尼亚轮竞争,花了750亿美元,由William Pirrie's Belfast firm设计建造完成。铁达尼号于1912年建造完成,成



为全世界最大的游轮。至少以其头等舱来看,也可称得上是最豪华的游轮了。它的头等舱能容纳905名旅客,二等舱能搭乘564名,而三等舱则有1124名旅客的容量,另外加上900名船上工作人员。不过这次首航旅程上,乘客并未载满。铁达尼号与现在的运输工具大小的比较大略如下。

船体内部的资料

长度 882尺8寸/268公尺(约等于3个草地足球场长度)

甲板 9层(包括最下层甲板)分为A、B、C、D、E、F、G及G下的锅炉房

横梁 92.5尺/28公尺

高度 入水深60.5尺,水面至烟囱175尺(相等于11层楼高)

乘客 329(头等),285(二等),710(三等)

全体机员 899人

引擎 2个直接启动转化引擎,4个气缸:30000hp 75rpm,1个低压涡轮机:16000hp 165rpm

螺旋桨 3个:16尺(中间);23尺6寸(左右)

总重量 46328吨

实际排水量 24900吨

排水量 每日14000加仑

锅炉 29个

燃料 每日825吨煤

气压 215P.S.I.

密室 16个,伸展至F甲板

总运载量 3547人(全船 fully loaded)

救生艇 20艘—16艘木制及4艘折叠(可运载人数1178人)

救生设施 3560件救生衣及49个救生圈。

全速 24海里

乘客背景

1912年,铁达尼号建造代表着一个航向新世界的梦想,2208位来自社会各阶级的各流仕绅乃至无名小卒登上这艘号称永远不会沉没的巨轮,一趟由英国横跨大西洋的旅程,前往人们当时心目中的新大陆—美国。

铁达尼号建立在一个社会阶级分明的时代。从铁达尼号的船舱等级可以发现,其将等级分为三个层级,即头等舱、二等舱、三等舱。头等舱是最高级的舱别,搭乘的大多为当时的贵族阶级、富翁,带着他们的仆人一同登上船,其中不乏有时常往来大西洋两岸的商贾和他们的管家们一同参与这趟旅程;也包含了许多著名的美国名流,富人,因此他们通常会有随从跟着,例如他们的仆役、护士、女婢,甚至司机等等。

至于二、三等舱的乘客,背景大多来自不同的国家以及各种不同的社会阶级,对他们的大多数而言,这趟旅程几乎都是他们一生中第一次搭船长途旅程的经验,甚至是第一次的旅行经验,这些人大部分都向往着美国这块新大陆,怀着希望和梦想,期待能够在那块土地上,再创一个新的人生;另外有些人则

单纯是属于返乡的旅客,一些前往欧洲旅游而有幸在回程搭乘这趟具有历史性的旅程。

航线与沉没点

铁达尼的处女航从英国的南汉普顿先航向法国与爱尔兰,目的地是美国的纽约。预定单程是七天,她最后靠岸是在爱尔兰女王街港。四月十四日深夜,铁达尼号以每小时二十三里高速航行于大西洋上。深夜十一时四十分,当铁达尼航行于北纬四十一度四十六分、西经五十度十四分,在新凡兰岛南部400英里,忽然撞上游离冰山。冰山像锋利的钢刀在客轮左舷撕开一个大裂口,5个防水隔舱顿时破裂,一片混乱。到了4月15日凌晨2点20分,这艘巨型客轮悲哀地结束了短促的生命,在距离纽约港东北2575公里的冰海沉没,1500多名乘客葬身在海底。

沉船经过

铁达尼号这次的擦撞,不是水线以上的船身,而是船头下右舷的底舱部分,擦撞并没有撞出破洞,而是船身擦撞处的几块钢板凹陷,板端铆钉崩脱而向外张开,形成了长达百公尺的一道口子,占全船长的三分之一,涵盖了六舱,前五舱都有水密舱,一舱又一舱灌满,虽有五万匹马力的海上巨无霸,也随着海水涌入而下沉,十一点四十分擦撞,凌晨两点十八分全船沉没,它只在海面上支持了两小时四十二分。

这座海上皇宫首航之日,乘客满载,连船员共达两千一百五十四人。而船上仅仅只有十四艘救生艇和四艘可折小艇,救生艇每艘可载四十人,可折小艇每艘可载七十五人,总共能载八百六十人(实际上只救了六百五十一人,有些救生艇上的人数才四成),其他的一千两百九十四人(实际为一千五百零三人),就注定了要眼睁睁随船下沉淹死!船从中间断成了两截,船头沉的很快,然后船尾似乎有站起来一分钟,便悄然沉下海底。

附近的Cunard liner Carpathia,得知铁达尼的困境,立即转向铁达尼,以19海里全速前进,她费时4小时行驶58英里漂着冰山的水域。铁达尼号上2200多人中仅有705人生还。那些搭上救生艇

的生还者在早上时被 Cunard liner Carpathia 号搭救。

在1912年4月14日号,发生整个铁达尼号的撞冰山事件,以下我们就时间的经过来叙述整个事件:

11:40PM 发现冰山,船员通知舰桥左转以避开右侧冰山的撞击,经过了37秒以后整艘船才开始向左边转向,然而已来不及避开与冰山的碰撞,冰山水面下的部分撞击铁达尼号并在与其右侧的船壳上发生总计长248英尺的摩擦,造成接下来在这个部分的船壳出现破洞与裂痕,并开始在前方的四个船舱进水,Titanic的主要设计者Edward Wilding估计在发生撞击后的40分钟内,约16000立方英尺的海水灌进铁达尼号的船舱,并判断铁达尼号即将沉没,当Smith船长了解船的进水状况之后,随即发出求救信号,Frankfurt与Carpathia这两艘位于附近的船马上回应的求救的讯号,但是即使距离最近的Carpathia也必须在4小时后才能赶抵现场,以铁达尼号的进水速度已经来不及赶到现场救援,Smith船长了解在救生艇只有一半乘客的承载量的状况下,船上半数的人在华氏30度的海水中终将罹难。

00:05AM 船员开始放下各艘救生艇。

00:25AM 乘客开始纷纷登上救生艇。

00:45AM 第一艘救生艇Lifeboat7出发,然而就在此时,在北方的海平面上出现一道船舰的光线,Californian正从铁达尼号附近经过,然而尽管船员施放求救的信号火箭,Californian终究没有看到这个信号并逐渐驶离。

1:15AM 随着进水情形更趋严重,乘客纷纷争相登上救生艇。

1:40AM 在大部分救生艇都已驶离的情况下,船上的乘客只好朝甲板上较安全的地方移动。

2:05AM 最后一艘救生艇驶离铁达尼号。

2:17AM 铁达尼发出SOS call,也是她最后一通求救讯号。

2:18AM 船上的灯光熄灭。

2:20AM 船上锅炉的部分由于进水逐渐沉入水中,造成前半部船身几乎垂直于水面,最终沉入水中,至此铁达尼号终于沉没。

事后发展

科学家推测,可能是冰山撞凹了约3公尺×9公尺的钢板,使铆钉因此而迸脱,海水便由接缝处灌入。但原因似乎并不如此单纯。比较对钢板边缘的接缝处和船沉没前船员所看到的受损状况纪录,令人觉得难以理解。而且船不到3个小时就沉没,速度也太快了。1985年海洋地质学家拜勒德在约3844公尺深的海底,找到了TITANIC,令人惊讶的是,在海底找到的铁达尼号已断成两截,船头和船尾的距离相当远,各种碎片散置各处,经研究结果发现TITANIC的钢板太脆了,以至于碰到冰山后,钢板没有弯曲,却断裂了,因当时的人没有脆性断

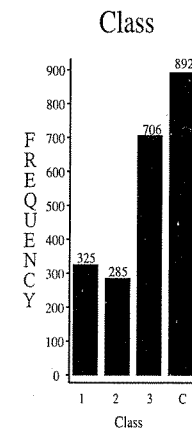
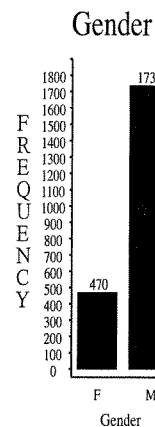
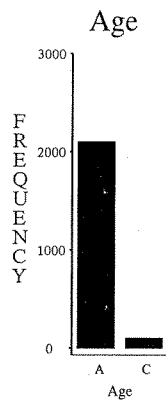
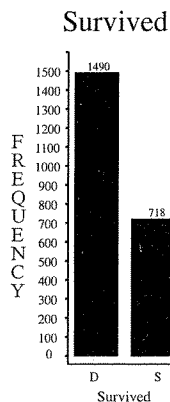
生死	年龄	性别	舱等
D	A	F	1
D	A	F	1
D	A	M	1
D	A	M	1
S	A	M	1
S	A	M	1
S	A	F	1
S	A	F	1
S	A	F	1
S	A	F	1
S	A	F	1
S	A	F	1
S	A	F	1
S	A	F	1

裂的概念,通常只要通过抗张强度的测验就可以了,因而造成此一重大悲剧的发生。

资料描述

本组资料具有生死、年龄、性别、舱等四个属性(栏位)共计2208笔,下面列出部分资料:其中符号说明如下:生死栏:S表示存活,D表示死亡;年龄栏:A表示成人,C表示小孩;性别栏:F表示女性,M表示男性;舱等栏:1表示头

单一变数直条图



等舱,2表示二等舱,3表示三等舱,C表示船员。

统计分析方法

首先我们采用简单的单一变数直条图,可以看出各栏位大略之分布情形。大略而言:生死比率约2:1;成人小孩比率约20:1;男女比率约3.5:1;三等舱人数为头等舱与二等舱之人数总和。

生死与年龄之交互关系表格

	大人	小孩	总数
存活	661	57	718
死亡	1438	52	1490
总数	2099	109	2208

生死与性别之交互关系表格

	男性	女性	总数
存活	374	344	718
死亡	1364	126	1490
总数	1738	470	2208

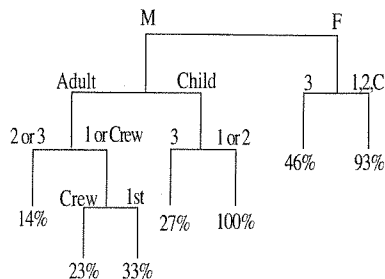
生死与舱等之交互关系表格

舱等	1	2	3	C	总数
存活	203	118	178	219	718
死亡	122	167	528	673	1490
总数	325	285	706	892	2208

接着我们利用交叉分析表来分析其他栏位与生死栏之关系。这是常见的表格。如何看这交互关系呢?我们拿生死与性别之表格为例。在全部2208人之中有718人生存149人死亡比率约1:2。如果生死栏与性别无关,那么不管男性或女性其生死比率都应该在1:2左右。事实上在470位女性之中有344人生存126人死亡,比率为2.7:1。这代表着生死栏与性别栏有着密切的关系。

另外比较先进的方式则是采用树状分类法。如下所示,决定生死最重要变数是性别。所以树状分类将所有数据一分为二:男性在左,女性在右。在右方的女性资料之中,决定生死最重要的变数是舱等:三等舱在左方生存率只有46.6%。而头等舱,二等舱以及女性的船员则有93%的高生存率。男性方面最重要变数是年纪。其中决定男性儿童的最主要因素是舱等:头等舱与二等舱的男性儿童全部生还(100%),而三等舱的男性儿童则只有27%的生存率。总而言之,女性与孩童的确有比较高的生存率。这与当时的急救措施应该有一定的关系(所以可以大胆假设当时的确是妇孺优先登上救生艇)。另一值得注意的是三等

舱。三等舱的生存率似乎不如其他舱等高(女性只有46%,男性儿童则只有27%),所以也可以合理的推断当时船舱裂洞进入处很可能是在三等舱。简单的统计工具,便可以看出当时的沉船轮廓。当然还有许多先进的统计方法可以使用,在这里仅就几样简单的统计工具作个初步说明。



相对铁打尼号沉船资料分析,另一个极端的例子是“大陆情书”。这一个例子是标准的八股文。报告本身的架构,一次搞定。尔后每一回换汤不换药,照本宣科。就我个人了解,国内许多单位均采用之。我换个比较温和的题目叫“大陆情书”,但用在一般统计报告上,倒也十分传神。此类统计报告,没头没脑,对数字背景资料不求甚解。原文如下:

亲爱的:

我们的感情,在组织的亲切关怀下、在领导的过问下,一年来正沿着健康的道路蓬勃发展。这主要表现在:

(1) 我们共通信121封,平均3.01天一封。其中你给我的信51封,占42.1%;我给你的信70封,占57.9%。每封信平均1502字,最长的达5215字,最短的也有624字。

(2) 约会共98次,平均3.7天一次。其中你主动约我38次,占38.7%;我主动约你60次,占61.3%。每次约会平均3.8小时,最长达6.4小时,最短的也有1.6小时。

(3) 我到你家看望你父母38次,平均每9.4天一次,你到我家看望我父母36次,平均10天一次。

以上充分证明一年来的交往我们已形成了恋爱的共识,我们爱情的主流是互相了解、互相关心、互相帮助,是平等互利的。当然,任何事物都是一分为

二的,缺点的存在是不可避免的。我们二人虽然都是积极的,但从以上的数据看,发展还不太平衡,积极性还存在一定的差距,这是前进中的缺点。相信在新的一年里,我们一定会发扬成绩、克服缺点、携手前进,开创我们爱情的新局面。

因此,我提出三点意见供你参考:

1. 要围绕一个爱字; 2. 要狠抓一个亲字; 3. 要落实一个合字。

让我们弘扬团结拼搏的精神,共同振兴我们的爱情,争取达到一个新高度,登上一个新台阶。本着我们的婚事我们办,办好婚事为我们的精神,共创辉煌!

你的小惠

综合而言,一个成功的案例应该有上、中、下游。上游收集背景资料以提出有意义的问题,中游利用统计方法来解读、推敲答案,下游给予统计结果赋予生命、提供有意义的结论以方便领导决策。很遗憾,一般统计教育只重视那既辛苦又得不到重视的中游工作。拿到数据,对资料背景不作深入了解便草草下手。统计方法任意套用,当作代数习题解答。错误的方法或使用不当,造成下游工作毫无着力点,得到统计结果只会写八股文——“大陆情书”只是套用轻松的一则美丽错误。数字工作者千万不要把统计当代数习题。面对现实、实事求是,让我们开拓另一个统计新格局。看看下面这段法官与犯人的对话,您作何感想?

法官:“你为什么要印假钞?”

被告无辜地说:“因为我不会印真钞呀!”

