

魔鬼统计学

作者：伊恩·艾瑞斯

魔鬼统计学

[美]伊恩·艾瑞斯

刘清山 译

版权信息

书名：魔鬼统计学

作者：[美]伊恩·艾瑞斯

译者：刘清山

版权：后浪出版咨询（北京）有限责任公司

目录

1. [前言 超级数据分析师的崛起](#)
2. [第一章 谁在替你思考？](#)
 1. [回归方程帮你找到灵魂伴侣](#)
 2. [赌徒的痛点可以量化](#)
 3. [当公司比客户更理解客户](#)
 4. [消费者的反击](#)
 5. [让人信服的回归方程](#)
 6. [世界是个数据矿](#)
 7. [“坏蛋”是可以分析出来的](#)
 8. [寻找神奇数字](#)
3. [第二章 反直觉：随机化与最优解](#)
 1. [第一资本掷骰子](#)
 2. [你看到的可能是随机测试网页](#)
 3. [随机测试是对直觉的检验](#)
 4. [随机测试——应用广泛的利器](#)
 5. [随机测试为什么没有大范围普及？](#)
4. [第三章 概率治国](#)
 1. [为省钱而花钱](#)
 2. [州级思想实验室](#)
 3. [随机应用的随机测试](#)
 4. [概率的世界](#)
5. [第四章 医生该如何看待循证医学？](#)
 1. [拯救十万人的生命](#)
 2. [错误模型导致错误思维](#)
 3. [随查随用——统计结果的快速传输](#)
 4. [未来已来](#)
6. [第五章 传统专家 vs 算法方程](#)
 1. [“令人不安的小书”](#)
 2. [为什么人类不善于预测？](#)
 3. [数据分析与经验知识是否可以兼容？](#)
 4. [我们还能做什么？](#)
7. [第六章 世界正在迅速数字化](#)
 1. [从90到300万](#)
 2. [数据的商品化](#)
 3. [数据混搭](#)
 4. [技术还是方法？](#)
 5. [计算机可以像人类一样思考吗？](#)
 6. [哪一部电影会赚钱？](#)
 7. [超级数据分析革命来临](#)
 8. [惊心动魄的未来图景](#)
8. [第七章 我们找到乐趣了吗？](#)
 1. [老师是什么？花瓶吗？](#)
 2. [帝国的反击](#)
 3. [地位之争](#)
 4. [你会购买超级数据分析师的二手车吗？](#)
 5. [好莱坞的破坏者](#)
 6. [提防超级数据分析者的赠礼](#)
 7. [隐性歧视](#)
 8. [我知道你明年夏天会做什么](#)
 9. [约翰·洛特是谁？](#)
 10. [假如数据分析师犯错](#)
9. [第八章 未来需要的人才](#)
 1. [未来属于统计学家](#)
 2. [一个数字中蕴含的海量信息](#)
 3. [概率领跑者](#)
 4. [反向推导](#)
 5. [波拉克（Polak）的怀孕问题](#)
10. [结语](#)

11. [后记 关于革命的进一步解释](#)
12. [致 谢](#)
13. [出版后记](#)

前言 超级数据分析师的崛起

奥利·阿森费尔特（Orley Ashenfelter）非常喜爱葡萄酒。“你只要把优质红酒放上一段时间，它就会发生一些非常神奇的事情。”阿森费尔特不仅对于葡萄酒的品鉴很着迷，而且希望知道优质葡萄酒和普通葡萄酒背后的推动力量。

“当你购买优质红酒时，”他说，“你是在进行一项投资，因为它未来很可能会变得愈发诱人。你希望知道的不是它现在的价值，而是它未来的价值，尽管你可能不会卖掉它或者喝掉它。你可以通过推迟品尝获得多少快乐呢？这是一个非常有趣的话题。”过去25年，这个话题占据了他的很大一部分时间。

阿森费尔特的日常工作是分析数字。他用统计量提取隐藏在大型数据集中的信息。作为普林斯顿的经济学家，他曾考察同卵双胞胎的工资，以估计多上一年学所带来的影响。他曾考察限速差异，以估计各州对于统计寿命的重视情况。他曾在多年时间里担任美国顶级经济学期刊《美国经济评论》的编辑。

阿森费尔特个子很高，留着浓密的白色络腮胡子，他那洪亮而友好的嗓音往往会成为会场上的主旋律。他并不胆小。你可能认为数字分析师是一些懦弱而腼腆的人，但阿森费尔特会使你迅速消除这种成见。我曾见过阿森费尔特在教室里踱步，用和蔼而热情的语气阐述一篇研讨会论文背后的原理。当他以高度赞扬的语气开始他的评论时，你就要当心了。

阿森费尔特通过分析数字评价波尔多葡萄酒品质的做法使他遇上了很大的麻烦。他不是使用罗伯特·帕克（Robert Parker）等葡萄酒大师“痛饮和回味”的方法，而是用统计量寻找与拍卖价格高低相关的年份酒特征。

“这是显而易见的，”他说，“葡萄酒是农业产品，它会受到不同年份天气的极大影响。”根据法国波尔多地区数十年的天气数据，阿森费尔特发现，较少的收获期降水量和较高的平均夏季气温可以给人们带来品质最佳的葡萄酒。正如彼得·帕塞尔（Peter Passell）在《纽约时报》中所说，阿森费尔特的统计方程与数据吻合得相当完美。

当葡萄成熟、汁液浓缩时，波尔多酒是最好的。在夏季特别热的年份里，葡萄会充分成熟，其酸度会降低。在降水量低于平均水平的年份里，水果汁液会浓缩。所以，你往往会在炎热干燥的年份得到具有传奇色彩的年份酒。成熟的葡萄可以酿造出口感柔和的（酸度低的）葡萄酒。汁液浓缩的葡萄可以酿造出风味浓郁的葡萄酒。

阿森费尔特大胆地将他的理论简化成一个公式：

$$\text{葡萄酒品质} = 12.145 + 0.00117 \times \text{冬季降水量} + 0.0614 \times \text{生长季平均气温} - 0.00386 \times \text{收获期降水量}$$

没错。通过将当年的天气数据代入这个公式，阿森费尔特可以预测任何年份酒的总体品质。通过另一个更加复杂的公式，他可以更加准确地预测100多个庄园的葡萄酒品质。“这看上去可能有一点数学成分，”阿森费尔特说，“但这正是法国人在著名的1855年分类中为葡萄酒厂排名时使用的方法。”

传统葡萄酒评论家并不接受阿森费尔特基于数据的预测。英国《葡萄酒》杂志说：“这个公式显然很可笑，不值得尊重。”纽约葡萄酒商威廉·索克林（William Sokolin）说，在波尔多葡萄酒从业者眼里，阿森费尔特的工作“介于野蛮和歇斯底里之间”。阿森费尔特时常遭到葡萄酒交易者的轻视，当他在佳士得葡萄酒部门发表关于葡萄酒的演讲时，会场后排的交易商公然对他的演讲发出嘘声。

罗伯特·帕克也许是世界上最有影响力的葡萄酒评论家，他也是《葡萄酒先锋》的出版人。他生动地将阿森费尔特形容为“彻头彻尾的骗子”。虽然阿森费尔特是世界上最受尊重的计量经济学家之一，但是在帕克看来，他的方法“是尼安德特人看待葡萄酒的方式。它太荒谬了，简直不值得嘲笑”。帕克否认了数学公式有助于鉴别优质葡萄酒的可能性，“我不想去他家做客，喝他家的葡萄酒”。

帕克说，阿森费尔特“就像一个从不看电影、仅仅根据演员和导演判断电影好坏的影评人一样”。

帕克的观点有一定的道理。既然通过观影得到的判断更加准确，通过品尝葡萄酒得到的判断不也应该更加准确吗？但是，这里有一个问题：波尔多和勃艮第葡萄酒需要在橡木桶里存放18到24个月，然后封装在酒瓶里。像帕克这样的专家在葡萄酒装进酒桶4个月以后才能开封品尝。即便此时，他们喝到的仍然只是正在发酵的非常难闻的混合物。我不知道品尝这种无法饮用的早期葡萄酒能否为品酒师带来关于葡萄酒未来品质的非常准确的信息。例如，巴特菲尔德拍卖行葡萄酒部门前主管布鲁斯·凯泽（Bruce Kaiser）说过：“早期葡萄酒变化很快，在其存放至少10年甚至更长时间以前，没有一个人可以对葡萄酒做出准确评价。”

与之形成鲜明对比的是，阿森费尔特通过分析历史数据找到了天气和价格之间的关系。他通过这种方式发现，冬季降水量每增加1厘米，预期价格往往会增加0.00117美元。当然，这只是一种趋势。不过，通过分析数字，阿森费尔特可以在葡萄收获时立即预测出未来的年份酒品质——比首次品尝提前几个月，比首次销售提前几年。由于葡萄酒期货交易频繁，因此阿森费尔特的预测为葡萄酒收藏者带来了巨大的竞争优势。

20世纪80年代后期，阿森费尔特开始在半年度简报《流动资产》上发布他的预测。他首先在《葡萄酒观察家》上用小篇幅广告宣传他的简报，逐渐积累了大约600名订阅者。订阅者来自世界各地，包括许多百万富翁和品酒专家——其中大多数人来自接受计量经济学方法的葡萄酒收藏者小圈子。罗伯特·帕克的简报《葡萄酒先锋》有3万订阅者，年费为30美元。相比之下，阿森费尔特的订阅群体显得不值一提。

1990年初，《纽约时报》头版发布了一篇关于阿森费尔特新型预测机器的文章，使更多的人接触到了他的思想。他公开批评帕克对于1986年波尔多葡萄酒的评估。帕克认为1986年葡萄酒“非常好，简直称得上出类拔萃”。阿森费尔特不同意这种观点。他认为这种酒的生长季平均气温低于平均水平，收获期降水量高于平均水平，因此该年份酒一定很平庸。

不过，这篇文章中真正的重磅炸弹与阿森费尔特对1989年波尔多葡萄酒的预测有关。这些葡萄酒在酒桶里只存放了3个月，还没有机会得到评论家品尝，但阿森费尔特认为它们将成为“这个世纪的葡萄酒之选”。他表示，这种酒一定“非常好”。根据他的评分标准，如果上好的1961年波尔多葡萄酒是100分，那么1989年波尔多葡萄酒将达到惊人的149分。阿森费尔特公然表示，它们的“售价不会输给过去35年酿造的任何一款葡萄酒”。

葡萄酒评论家被激怒了。帕克此时称阿森费尔特的定量估计“荒谬可笑”。索克林说，人们的反应夹杂着“愤怒和恐惧。阿森费尔特的确惹恼了許多人”。在几年时间里，《葡萄酒观察家》不再为阿森费尔特（和其他人）的简报发布任何广告。

传统专家团结在一起，试图诋毁阿森费尔特及其方法。他们说，阿森费尔特的方法存在缺陷，因为它无法精确预测未来的价格。例如，《葡萄酒观察家》品尝主管托马斯·马修斯（Thomas Matthews）抱怨说，阿森费尔特对于27种年份酒的价格预测只有3次是完全准确的。虽然阿森费尔特的“公式得到了特别设计，以拟合价格数据”，但他的“预测价格不是高于实际价格，就是低于实际价格”。不过，对于统计学家（以及其他任何稍做思考的人）来说，拥有时高时低的预测是一件好事，它是无偏估计的象征。实际上，阿森费尔特指出，帕克最初对于年份酒的评价存在系统性的上偏趋势。帕克常常需要对他最初的评价进行下调。

1990年，阿森费尔特做出了更加冒险的举动。在将1989年葡萄酒称为“世纪年份酒”之后，他发现，根据数据，1990年的葡萄酒将会表现得更好。他宣布了这一结果。事后来看，我们知道《流动资产》的预测是非常准确的。1989年葡萄酒成了非常优秀的年份酒，而1990年葡萄酒表现得更好。

怎么能连续两年拥有“世纪年份酒”呢？原来，自从1986年以来，每一年的生长季气温都要高于平均水平。法国天气已经温暖了20多年。这是种植柔和的波尔多葡萄的好时机，对葡萄酒爱好者来说也是一个好消息。

传统专家现在对天气的重视大大提高了。许多人从未公开承认阿森费尔特的预测威力，但他们自己的预测与阿森费尔特那个简单公式得到的结果更加吻合了。阿森费尔特仍然在维护他的网站 www.liquidasset.com，但他不再制作简报了。他说：“现在和过去不同了，品酒师不会再犯可怕的错误了。坦白地说，我是在自掘坟墓。我再也没有像之前那么高的附加值了。”

阿森费尔特的诋毁者将他看作异端。他揭开了葡萄酒的神秘面纱，对他们产生了威胁。他回避了华丽而荒谬的术语（“强健”“紧实”“轻快”），为他的预测给出了理由。

葡萄酒行业毫不妥协的态度不仅与审美有关。“葡萄酒交易商和作家不想让公众获得阿森费尔特提供的那种信息，”凯泽评论道，“事情始于1986年陈酿。阿森费尔特称之为骗局，因为那一年很糟糕，下了很多雨，气温也不够高。不过，当时所有葡萄酒作家都在唱赞歌，称之为伟大的年份酒。阿森费尔特是正确的，但正确并不总是受人欢迎。”

通过维持对于葡萄酒品质的信息垄断，葡萄酒交易商和作家可以从中获利。交易商通过长期高估的初始评价稳定价格。《葡萄酒观察者》和《葡萄酒先锋》维持着葡萄酒品质主要评判者的地位，并以此获取数百万美元收入。正如厄普顿·辛克莱（以及现在的阿尔·戈尔）所说：“当一个人的工资取决于他对某件事情的不理解时，你很难让他理解这件事情。”同样的道理也适用于葡萄酒。“许多人的生计取决于葡萄酒饮用者对于这个公式的不信任，”阿森费尔特说，“他们突然之间变得有些过时了，这使他们感到愤怒。”

你可以看到一些变化的迹象。伦敦佳士得国际葡萄酒部门主席迈克尔·布罗德本特（Michael Broadbent）以外交家的口吻表述这件事：“许多人认为阿森费尔特是个怪人，我认为他在许多方面的确如此。不过，我发现他的思想和研究每年都与事实相符。他所做的事情对于希望购买葡萄酒的人相当有帮助。”

棒球界的奥利·阿森费尔特

葡萄酒品鉴这一高端行业似乎与面向大众的棒球运动相去甚远。不过，从许多方面来看，阿森费尔特为葡萄酒所做的事情与比尔·詹姆斯（Bill James）为棒球所做的事情是相同的。

詹姆斯在简报《棒球摘要》中对于棒球专家通过观察判断球员天赋的观点提出了疑问。迈克尔·刘易斯（Michael Lewis）在《魔球》中指出，詹姆斯将数据驱动型决策引入了棒球领域。詹姆斯的观点简单而有力，他认为基于数据的棒球分析优于专业观察：

肉眼无法获得评价球员所需要的知识。想想吧，你显然无法通过观察判断打击率为0.300的击球手和打击率为0.275的击球手之间的差异。这种差异相当于每两个星期相差一次打击……如果你观看两个人在一年中的15场比赛，那么打击率为0.275的击球手的击球数超过打击率为0.300的击球手的概率是40%……优秀击球手和普通击球手之间的差异是肉眼无法看到的——它只能在数据中体现出来。

和阿森费尔特类似，詹姆斯相信公式。他说：“你应该用击球手试图取得的成功衡量他，而击球手试图取得的成功是制造得分。”于是，詹姆斯提出了一个新的公式，用于衡量击球手对于得分的贡献：

$$\text{制造得分} = (\text{击球数} + \text{保送数}) \times \text{总垒数} / (\text{打数} + \text{保送数})$$

这个公式重点强调了球员的上垒率，为那些经常获得保送的球员给出了特别高的评分。球探特别讨厌詹姆斯的数字分析方法。像罗伯特·帕克那样的葡萄酒评论家通过感受口感和气味生存，而球探的眼睛则可以决定他们的生死。这是他们的附加价值。正如刘易斯所说：

在球探看来，要想发掘大联盟球员，你需要开车行驶10万公里，在100家劣质汽车旅馆里住宿，无数次在丹尼餐厅就餐，以便在4个月时间里观看200场高中和大学棒球比赛，其中199场对你毫无意义……你会走进球场，在捕手正后方第四排铝制长椅上找到一个座位，看到其他人看不到的一些事情——至少没有人知道这些事情的意义。你只要看到这个球员一次就够

了。“只要你一看到他，你就明白了。”

球探和像罗伯特·帕克那样的葡萄酒评论家的共同点不仅仅是喜欢回味和吐痰。帕克相信，他能根据一次品鉴评估庄园年份酒的质量。类似地，棒球球探相信，他们可以根据一次观赛评价高中潜力球员的质量。

在这两个领域，人们试图预测未经检验的不成熟产品的市场价值，不管它们是葡萄还是棒球选手。到底应该依靠专家的观察还是定量数据呢？这是这两个领域的核心争议。

和评论家类似，棒球球探常常使用无法证伪的委婉说法，比如“他是真正的选手”或者“他是工具型球员”。

在《魔球》中，当奥克兰运动家队总经理比利·比恩（Billy Beane）想要买入杰里米·布朗（Jeremy Brown）时，数据和传统专业知识的冲突达到了顶点。比恩读过詹姆斯的文章，决定根据数字选秀。比恩喜欢杰里米·布朗，因为他的保送频率是大学球员中最高的。球探讨厌他，因为他很胖。一名运动家队球探嘲笑说，如果他穿着灯芯绒服装跑步，“他就会引发火灾”。球探认为，像他这种体型根本不可能在大联盟打球。比恩完全不在乎球员的外表。他的选秀准则是“我们不是卖牛仔裤的”。比恩只想赢得比赛。球探似乎想错了。在运动家队当年征召的新人中，布朗的进步是最快的。2006年9月，他代表运动家队在大联盟首次出场，取得了0.300的打击率（上垒率为0.364）。

阿森费尔特和詹姆斯最初宣传数字分析结果的方式具有惊人的相似性。和阿森费尔特类似，詹姆斯起初为他的第一份简报《棒球摘要》发布小广告（他将《棒球摘要》适度修饰成一本书）。第一年，他一共卖出了75份。阿森费尔特遭到《葡萄酒观察者》封杀。类似地，当詹姆斯请求分享数据时，他遭到了埃利亚斯体育中心的排斥。

不过，詹姆斯和阿森费尔特在其行业里永远留下了自己的印迹。《魔球》所记载的奥克兰运动家队的长期成功以及波士顿红袜队在西奥·爱泼斯坦（Theo Epstein）的数据管理下首夺世界大赛冠军的壮举都被归功于詹姆斯的持续影响。现在，即使是传统葡萄酒评论家的预测也会考虑到天气因素，而且进行相对优化，这是对阿森费尔特所做贡献的无声致敬。

两个人都促成了分析团队的诞生，这些团队拥有自己的数据分析品牌。在詹姆斯的影响下，美国棒球研究协会得以成立。棒球数据分析现在甚至拥有了自己的名字，叫作赛伯计量学。2006年，在阿森费尔特的帮助下，《葡萄酒经济学期刊》得以创刊。现在，美国甚至有了一个葡萄酒经济学家协会。阿森费尔特不出意外地成为协会首任主席。顺便一提，事后看来，阿森费尔特最初的预测非常准确。我查询了拉图庄园最近的拍卖价格。显然，1989年葡萄酒的价格是1986年葡萄酒的两倍多，1990年葡萄酒的价格还要更高。听到了吗，罗伯特·帕克？

酒中的真相

本书的核心观点是，数据分析在葡萄酒和棒球领域的兴起不是孤立事件。实际上，葡萄酒和棒球的案例是本书核心主题的缩影。我们正处于马车和火车相互竞争的历史性时刻，我们的直觉和经验知识正在一次又一次地输给数据分析。过去，许多决策仅仅依赖于经验和直觉的某种结合。专家凭借几十年的个人试错经历获得了神圣地位。我们相信他们知道最好的行事方式，因为他们之前做过几百次同样的事情。经验型专家的角色被保留下来，而且发展得很好。如果你想知道某件事情应该怎样做，你应该询问银发老人。

现在，一些事情正在发生变化。商业和政府专业人士正在越来越多地用数据库指导他们的决策。对冲基金的故事实际上是新一代数据分析师的故事——我称之为超级数据分析师——他们分析大型数据集，以便在看似无关的事物之间发现实证关联。想为大量购买欧元避险吗？事实上，你应该卖出由26种其他股票和商品精心搭配而成的投资组合，其中可能包括沃尔玛股票。

什么是超级数据分析？它是影响现实决策的统计分析。超级数据分析公式预测通常是大小、速度和规模

的某种组合。首先，从观测值的数量和变量的数量来看，数据集合通常很大。其次，分析的速度正在加快。我们常常可以看到与数据生成同时进行的数据分析。最后，影响的规模有时是很大的。这不是几个书呆子在制造戏谑式的期刊文章。超级数据分析有些是由决策者执行的，有些是为决策者执行的，这些决策者希望寻找更好的行事方式。

当我说超级数据分析师在使用大型数据集合时，我指的是非常大的数据集。企业和政府的数据集合正在越来越多地用太字节甚至拍字节（1,000太字节）衡量，而不是用兆（M）字节或吉（G）字节衡量。1太字节（terabyte）相当于1,000吉字节。前缀tera来自希腊语，表示怪兽。一太字节的确十分庞大，像怪兽一样。整个国会图书馆大约有20太字节文本。本书的一个目标是让大家开始习惯这个前缀。例如，沃尔玛数据库存放了超过570太字节数据。谷歌拥有大约4拍字节存储量，它在不断对其进行分析。太字节挖掘不是巴克·罗杰斯（Buck Rogers）的幻想——而是正在发生的事情。

在一个又一个领域，“直觉主义者”和传统专家正在对抗超级数据分析师。在医疗领域，关于“循证医学”的激烈争论可以归结为是否根据统计分析选择疗法。直觉主义者不会束手就擒。他们说，数据库永远无法体现出临床专家通过一生的经历培养出来的知识，回归永远无法和拥有20年经验的急诊室护士相比，后者可以判断一个孩子看上去是否有问题。

我们往往认为，国际象棋大师加里·卡斯帕罗夫（Garry Kasparov）之所以输给深蓝计算机，是因为国际商用机器公司的软件更加聪明。那个软件实际上是一个为不同局面排序的大型数据库。计算机的速度很重要，但是最具决定性的是计算机访问70万盘大师棋局数据库的能力。卡斯帕罗夫的直觉输给了基于数据的决策。

超级数据分析师不仅在入侵传统专家的地盘并取代他们，而且正在改变我们的生活。他们不仅在改变决策方式，而且正在改变决策本身。棒球球探之所以输给分析师，不仅仅是因为分析数据比搭乘飞机前往帕路卡维尔成本更低。最主要的原因在于，数据分析得到的预测结果更加准确。当然，超级数据分析师和专家并不总是存在冲突。数据分析有时会证实传统观点。世界不是绝对的，传统专家不会永远犯错，他们的表现至少比碰运气要好。不过，数据分析仍然可以引导决策者做出不同的决策。总体而言，这种决策比之前要好。

在一个又一个领域，统计分析在完全不同的信息之中发现了一些隐性关系。如果你是政客，希望知道谁最有可能为你投票以及哪种拉票方式最有可能成功，那么你不需凭空猜测、遵循经验规则或者信任头发花白的传统主义者。相反，你可以从不同特性中梳理出越来越多的可测效应，从而更好地知道哪种拉票方式效果最好。数据库搜索可以揭示出传统专家从未考虑过的内在因素。

在我们周围，基于数据的决策正在变得越来越多：

- 租车公司和保险商拒绝向信用分数不佳的人提供服务，因为数据挖掘显示，信用分数与较高的事故概率存在相关性。
- 现在，当航班取消时，航空公司会跳过常旅客，首先向那些最有可能选择其他航空公司的旅客提供开放席位。当然，这些旅客是通过数据挖掘确认的。航空公司不是遵循先后来后到原则，而是根据几十项与消费者有关的因素提供服务。
- 《不让一个孩子掉队》法案要求学校采用由严格数据分析支持的教学方法。在其帮助下，教师花费高达45%的课堂时间培训孩子们通过标准化考试的方法。超级数据分析甚至使一些教师转向了每句台词得到编排和统计审查的课堂模式。

直觉主义者要当心了。本书将会讲述一系列超级数据分析的故事，向你介绍执行这些分析的人物。数字分析革命不仅与棒球甚至整个体育界有关。它关乎我们生活中的方方面面。很多时候，这种超级数据分析革命对消费者有利，因为它可以帮助商家和政府更好地预测谁需要什么东西。不过，在另一些时候，消费者需要面对以统计为武器的对手。数据分析会使普通人处于极为不利的境地，因为商家可以更好地预测他们能从我们身上榨取多少利润。

史蒂文·D·莱维特（Steven D. Levitt）和斯蒂芬·J·都伯纳（Stephen J. Dubner）在《魔鬼经济学》中提到了几十个案例，展示了对数据库的统计分析是如何揭示隐秘因果关系的。莱维特和约翰·多诺霍（John Donohue，他是我的合著者和朋友，你在后面还会听到他的名字）指出，1970年堕胎率和1990年犯罪率这两个看似无关的数字存在重要关联。不过，《魔鬼经济学》并没有过多地谈论定量分析对于现实决策的影响程度。与之相比，本书会谈论数据分析的影响。商业和其他领域的决策者正在以你从未想象过的方式使用统计分析，以进行各种选择。

全球各行各业都在围绕现代计算机的数据库容量进行升级。万斯·帕卡德（Vance Packard）的《隐形说客》等作品展示了人们在20世纪50年代至60年代的预期（和恐惧）——由大政府和大公司实施的复杂的社会工程即将控制世界。现在，这个预期突然在这一代人心中复活了。过去，我们认为大政府可以通过指挥和控制解决人类所有问题。现在，我们看到，类似的事情正在以大规模数据网络的形式出现。

我与《魔鬼经济学》

我本人是数据分析师。虽然我在耶鲁教授法律，但我在麻省理工读博士时学过计量经济学。从保释保证书和肾移植，到枪支管制和冲动型犯罪，我曾为各种事情做过数据分析。你可能认为，象牙塔里的书呆子与现实世界中的决策没有任何关系。（是的，我属于那种粗心大意的教授。我曾在火车上专注于写作，忘记在纽黑文下车，直接坐到了波基普西。）不过，即使是书呆子的数据挖掘有时也会对世界产生影响。

几年前，我曾和史蒂文·莱维特合作，以研究路捷对汽车盗窃的影响。这是一件非常现实的事情。路捷是一种小型无线电发射器，可以隐藏在汽车内部的许多部位。接到失窃报告时，警方会远程激活发射器，然后用特殊装备的警车跟踪失窃车辆的确切位置。路捷是一种非常有效的汽车寻回设备。路捷公司知道这一点，并且自豪地宣布，路捷的寻回率为95%。不过，我和史蒂文希望测试路捷是否有助于降低整体汽车失窃率。许多汽车防盗设备的问题在于，它们可能只是把罪行转移到了其他地方。如果你在汽车上使用“俱乐部”这一产品，它很可能无法阻止犯罪。窃贼会沿着街道继续前进，盗窃下一辆汽车。路捷的优点在于其隐蔽性。在一座由路捷覆盖的城市里，窃贼并不知道某辆车是否安装了路捷。

这正是莱维特喜欢探索的那种反常现象。《魔鬼经济学》的评论家说，史蒂文看待事物的角度和常人不同，这话一点不假。几年前，我手里多出一张票，因此邀请史蒂文去看芝加哥公牛队迈克尔·乔丹的比赛。史蒂文认为，如果他对比赛投资，那么他会更加享受比赛过程。不过，他不太在乎公牛队的输赢，这一点和我完全不同。所以，在比赛即将开始时，他在网上下了很大的赌注，赌芝加哥获胜。现在，他为比赛做了投资，网上的赌注改变了他的观赛动机。

奇怪的是，路捷也是一种改变动机的设备。在路捷出现以前，许多职业窃贼几乎不会留下任何线索。路捷改变了这一切。有了路捷，警方不仅可以寻回车辆，而且常常可以捉住窃贼。仅在洛杉矶，警方就通过路捷发现了100多家汽车销赃店。如果你在安装路捷的城镇偷走100辆汽车，那么你几乎一定会偷到一些装有路捷的车辆。我们希望测试路捷能否震慑盗窃犯，使之不在当地行窃。如果是，路捷就创造出了经济学家所说的“正外部效应”。当你在车上安装“俱乐部”产品时，你很可能提高了其他人的汽车遭到盗窃的概率。不过，我和史蒂文认为，如果足够多的人安装路捷，他们也许可以震慑职业偷车贼，使之不敢在他们的街区行窃。

我们最大的问题是说服路捷和我们分享销售数据。我记得，我曾反复拨打电话，对他们说，如果我和史蒂文是对的，这件事就可以成为人们购买路捷的另一个理由。如果路捷可以降低窃贼盗窃其他车辆的概率，那么路捷公司也许可以说服保险公司向路捷用户提供更大的折扣。最后，一位低级别经理终于向我们发送了大量有用数据。说实话，路捷最初对于这项研究不太感兴趣。

当他们看到我们的论文初稿时，一切都变了。在考察了56座城市14年的汽车盗窃数据后，我们发现，路捷可以为其他人带来巨大的好处。在高犯罪率地区，一个人对于路捷的500美元投资可以使非路捷用户的汽车失窃损失减少5,000美元。我们按照年份和城市分析了路捷销量，对于行驶在路上的路捷车辆比例做出了非常准确的估计。（例如，波士顿拥有州内最大的保险折扣，该市超过10%的汽车安装了路

欢迎访问：电子书学习和下载网站 (<https://www.shgis.com>)

文档名称：《魔鬼统计学》伊恩·艾瑞斯.epub

请登录 <https://shgis.com/post/3646.html> 下载完整文档。

手机端请扫码查看：

