

《数据统治世界》

图书基本信息

书名：《数据统治世界》

13位ISBN编号：9787300177540

10位ISBN编号：7300177549

出版时间：2013-9-1

出版社：Numbers Rule Your World: The Hidden Influence of Probability and Statistics on Everything You Do

作者：[美]冯启思（Kaiser Fung）

页数：248

译者：曲玉彬

版权说明：本站所提供下载的PDF图书仅提供预览和简介以及在线试读，请支持正版图书。

更多资源请访问：www.tushu000.com

《数据统治世界》

内容概要

《数据统治世界》

作者简介

书籍目录

| | |
|--------------------------------|--|
| 前言 让数据自己说话 / 1 | |
| 第1章 关注异常值，而非平均数本身：解决拥堵之害 / 001 | |
| 迪士尼，让游客牢骚的长队 | |
| 选择上下班线路的冒险 | |
| 匝道控制，反拥堵的利器 | |
| 知觉管理，快速通行卡让等候时间“变短” | |
| 适当放弃最佳，赢得支持 | |
| 消除变异，消除怒气 | |
| 第2章 相关比因果更重要：疾病侦测与信用评分 / 035 | |
| 污染的菠菜与大肠杆菌 | |
| 建模师为信用评分 | |
| 寻找疾病的罪魁祸首 | |
| 信用评分，相关创造商业奇迹 | |
| 抛弃脏数据 | |
| 统计建模的两大模式 | |
| 第3章 分层与同类比较：考试公平与保险风险 / 087 | |
| 黑人考生与白人考生之间通过率的巨大差距 | |
| 统计学家助力SAT题目诞生 | |
| 项目功能差异分析消除差异 | |
| 突然不可保的飓风 | |
| 被误解的“百年一遇” | |
| 将不同的组分分开 | |
| 第4章 假阳性与假阴性的博弈：药检与反恐 / 131 | |
| 不能给假阳性一丝机会 | |
| 统计学上的分界线 | |
| 用测谎仪证明自己 | |
| 难以把握的成本效益比 | |
| 宁可错杀三千，不可放过一个 | |
| 假警报，检测系统远非完美 | |
| 第5章 小概率的力量：航空安全与彩民信心 / 187 | |
| 夜空中的灾难 | |
| 4次灾难惊人的巧合 | |
| 27000年才有一次的中奖机会 | |
| 白点黑点，换个角度看数据 | |
| 在整个背景下评价数据 | |
| 精心选择的数字更丰富 | |
| 结论：像数据科学家一样思考 / 211 | |
| 译者后记 | |

精彩短评

- 1、思想早就在土地保护和风险分析课上介绍过了，今天还是坚持把它看完了。里面至少还是有些有趣的案例的。
- 2、这是一本很“现实”的统计学方面的书籍，案例很形象，贴近现实而容易引导读者思考。
- 3、里边的例子还行，不过这个作者似乎逻辑比较差，以至于目录也没有归纳出重点。三星，不能再多了。
- 4、这其实是一本讨论统计式思维的书，用了大量的具体事例分析来讨论一些比较简单的道理，给人感觉作者不像他在书中推崇的那种数据应用专家，而像是数据理论专家，废话过多。当然，如果你喜欢读案例的话，这本书还是非常值得一看的。
- 5、看完之后发现只看目录就够了。如果正在学习统计学，同时配合看看这个，有助于概念的理解。
- 6、从数据到正确的决策之间还差了十万八千里。很容易被名字给误导，不过可以很好培养统计学思维。比如假阳性一章比书上的一类错误二类错误讲的明白多了。
- 7、:无
- 8、感知管理
- 9、行文啰嗦，内容还有点意思
- 10、Crowd control (<http://www.tansuocn.com/detail/2634.html>)
- 11、2015.8.11起读。图书馆。
- 12、就那样吧，翻译的不好，感觉被茅茅骗了~
- 13、例子都详细有趣，也很注意总结，翻译有些地方不清楚。希望有空拜读原作。适合外行闲时读读
- 14、从生活中的衣食住行中的实例去思考，来科普所谓的“统计式思维”，运用统计式思维，不被生活中的各种数据蒙蔽眼睛。但Daniel Kahneman说过：“具有统计式思维是很困难的”。最后，吐槽下某些章节的翻译，有待提高。
- 15、大数据科普类书籍，适合我这种小白~~
- 16、简单的原理配上应用案例来说明而已，当小说看吧。
- 17、所有模型都是错误的，只不过有一些是有用的。比其他一些大数据类的书要深入一些，更多探讨了统计思维的负面，包括之前意识到的多次博弈问题。假阴/阳性一章的讨论涉及到how to draw a line, 很有启发
- 18、重要的不是统计数据，而是统计思维。一本小书，一本好书。
- 19、看看
- 20、这本书是挺好读的，很轻松也很有趣，说了要具有统计式思维是很困难的，要像数据科学家一样思考，透过数据，看到这个与眼睛看到并不太一样的世界。如果在日常生活中，你知道如何使用数据作出决策时，你就掌控了自己的世界。
- 21、案例生动易懂，再次激起了我对统计学的热爱；全书没有一个数学公式，但是使人对其中的数理思想深有感触。的确是一本好书。
- 22、十个故事讲统计思想，随便看看
- 23、翻译有一些硬伤，比如原文是A比B低15个百分点，译文变成A是B的15%
- 24、略扫了一遍，我把它当作统计科学史，作为数据分析师有必要了解，有空再拿出来细读一遍，现在想学习大数据的应用案例没有什么收获
- 25、这本书里的例子没用，思路扁平，很一般
- 26、没看完，统计思维，异常值，可参考能给些直观的概念理解
- 27、作为一本讲故事和概念的数还是不错的，但如果想从里面学点实用的东西，恐怕就要失望了
- 28、说了统计中最普遍的几个问题：相关性分析；分层和组间差异；假设检验。作者把这些统计概念放到实际生活中去解释，适合去阅读加深对概念的理解。
- 29、美剧的现实中更精巧的监测民众的是the machine
- 30、不要被标题副标题骗了 其实看目录就可以了【喂
- 31、对于从事数据工作的我来说，这本书有用并且值得向大家推荐，虽然会像某一位同事浏览后形容道：“内容就是讲故事而已嘛”。是的，就是一本用生活案例讲述数据如何统治世界的故事书。
- 32、抽象出来的五个观点还是不错的，这本书更适合作为一篇长文来写，作者的逻辑性和行文思路太

《数据统治世界》

差了，如果你想读这本书就看最后一章《像数据科学家一样思考》就足够了，前面废话太多了，而且写的不明不白的。

33、比较深入地呈现五大统计思维

1、书的内容比较简单，但是很值得一读。书主要讲述统计和数据分析在日常生活中的应用，读起来比较有趣，同时也能启迪思维。个人来说，最大的启迪有几点：1.人性、心理学的重要性迪斯尼深谙人性的洞察，可感知的等待时间和实际的等待时间是两码事。应用科学时，不能仅仅关注数学上的效用，也要关注人们的心理感受。2.权衡折中现实生活是灰色的，黑白两个极端只存在于理论之中。统计和数据分析应用当中，也存在很多类似的权衡折中。假阴性和假阳性往往是一对跷跷板，此起彼伏。

2、虽然我们现在热议大数据时代，开始到处谈论“大数据”。但其实，在我们的生活中，数据无处不在，当我们买彩票时，彩票的中奖概率组成了数据，当听到飞机失事时，我们又会看到相关的统计数据，更不用说经常听到的国民统计数据，还有一些以据调查显示开头的新闻了，甚至还会有关于“人们是否相信统计数据”的调查数据。由此，你是否想起了最近听到的统计数据呢？我们先以你我熟悉的彩票故事为例，人们买彩票时，是否有想过自己中大奖的概率？理论上说，每次彩票抽奖都是随机抽取的，与上次的结果和上上次的结果毫无关系，但还是有不少人在拿着以往的数据在预测未来，期待自己能成为下一个幸运儿。但不幸的是，实现这样的梦想总归也没有几人，让我们再来看一下乘飞机出行，每当我们看到飞机失事的时候，总是想飞机太不安全了，当下次有机会在飞机和地面交通中选择的时候，都义无反顾地选择了地面交通。这样做对吗？这个时候，统计科学家出现了，他们明确地告诉我们，飞机失事的概率和买彩票中的概率差不多。人们都在妄想买彩票中大奖，获得巨额财富，而又惧怕坐飞机，怕自己成为受害者，但两者的概率并无太大不同。事实上，这正是小概率的力量，在《数据统治世界》中，作者冯启思用这样一个故事为我们讲述一项统计式思维，“小概率的事件的发生性几乎为0”。其实，作者在讲述买彩票的例子的时候，是用这样一个故事来引出的，在加拿大，因为一桩彩票店店员将一名老人的中奖彩票据为己有的案例，让应用统计学家发现，彩票店店员的中奖概率要高出普通人的4倍！这不科学，因为发生这样的事情的概率只有10的十六次幂分之一（已经非常非常小的一个数字）。你应该明白了，在这个问题上，两名专家分别就彩票店中奖概率和飞机失事概率进行了研究，他们使用到的统计学思维便是“小概率事件不会发生”。因此，对他们来说，彩票对他们毫无吸引力，而且也十分信任自己乘坐的飞机。在《数据统治世界》一书中，包括“小概率事件不会发生”，作者总共介绍了5条统计式思维：“留意事件中的变异性”“从事实中发现有用信息”“同类进行比较”“两种错误之间的此消彼长”“小概率事件不会发生”。而他们又分别对应了十则精彩动人的故事，迪士尼的等候队列与明尼苏达州高速公路匝道，美国大肠杆菌流感与信用分评分，SAT考试测试与佛罗里达州的飓风保险，兴奋剂检测与测谎仪，异常的彩票中奖率与十分巧合的飞机失事事故。这十则精彩的故事，在每章中通过交叉叙述，事件发展跌宕起伏，作者带领着读者一步步探寻事情的真相，初次读来，本以为事情应该就此打住时，作者却又从新的线索或角度，将我们拉到另外一条精彩的故事线上，由此产生极好的脑力激荡。当你读完本书后，你在面对数据时，就可以通过新的角度，利用这些统计式思维重新审视我们身边发生的故事，和那些浮现在表面的数据，来揭示隐藏在数字背后的奥秘和数据的神秘力量。

3、这本书应该算是一本统计思维的科普读物，虽然看上去不算薄，但读起来很快。建议快速浏览前5章，精读后记，有什么不明白的可以回去前5章看看。另外，出版单位在书尾附了一篇小文章《商业书籍的阅读方法》很值得一看。

4、标题的名字有点哗众取宠，不过书当然是好书。是很好的统计学思维入门书。第一章：强调某些事物的特征有类似于木桶原理：运行效率取决于所有发生的因素中的最小值，而不是其均值。探明事物发生的性质本身，找到其实的依赖关系，而不要仅仅把思维局限于改变均值，关注变异，更能够解决问题。想到武汉的交通堵车，有些人认为不是因为车多，而是因为有些不守规矩的行为导致行车速度变化大，拖累后面所有行车人员，从而总行车时间增加。如果控制开车行为，减少使车速降至极低事件的发生概率，车多造成的拥堵问题没那么大。第二章：相关关系和因果关系在运用上不同。有些重大的事件影响深远，必须严密地探察出其原因，才能在解决问题的同时带来的伤害最小。而另外一些事件，原因众多纷纭，找到真实原因即使可能用处也不是那么大，在此时找到相关关系足够运用于实际决策，那么相关分析不需要真正探明因果关系。真正的因果可能会在逐渐寻找相关中显现出来。第三章：相互比较得出结论需要其他条件一致，这是一种基本的科学方法。在统计中做决策也是如此，不能把苹果和橘子进行比较。因此需要比较不同的组之间的差别。只有在组间其它因素都相同时

才能说明想要的结论。这也是计量经济学中控制变量和匹配的想法了。分组不当会引起所谓的道德风险问题，无端招致更多的风险使总体缺乏效率，同时也有风险低的补贴风险高的不公平现象。第四章：弃真和取伪是一个跷跷板的关系，不改变估计方法或者样本容量，是无法同时减少二者的。这个在计量经济学中的运用是在做检验的时候，会有一个原假设和一个备择假设。以谁为原假设会对其有保护作用，而对备择假设比较挑剔，除非有很强的证据证明，否则不予接收。而做计量回归中，我们希望得到原假设不成立，以接受备择假设的意思是，我们希望接受的备择假设是在有充分证据的基础上，因此是（从可得到的数据来说是）可靠的。在实际生活运用中，保护哪个，苛责哪个，取决于各自的收益和成本对比。如同食品药品监督管理局所承担的风险一样，冤枉了一个好的固然不好，可是没什么大问题，放过了一个坏的，可就备受批评了。因此原假设就是东西是坏的也就可以理解了。同时也可以说明，无罪假设需要对违规分子处于严厉的处罚，因为这是认为，一个人违规被抓只是众多违规行为的冰山一角，平时就干了不少坏事才会“天网恢恢，疏而不漏”。这样才能制止违规行为。第五章，小概率事件不可能在一次随机试验中发生。这几乎算是统计学家的信念了。但是小概率事件有吸引人的故事性，从而博得人们的眼球。同时，如果理论上小概率事件的确很凑巧地发生了，除非经过第四章取伪和弃真严密地考察排除一切可能，否则我们就要怀疑它是我们假设的随机发生的性质了，或者说我们要怀疑它的样本空间是不是原来那个。普通人没有所谓的统计思维或许是因为他们没有机会看见众多的正常事件或者忽略掉了它们。

章节试读

1、《数据统治世界》的笔记-04. 假阳性与假阴性的博弈

这章还是提到了几个涨姿势的点的。

*

假阳性/真阳性-宁可错杀三千不可放过一个

假设测谎仪准确率为90%，

情况A：甄选

10000个雇员（1:1000）

10个特务中成功找出9个错放1个；

9990个好员工中错抓999个。

每抓到1个特务，会有111个好职员遭到错误指控。

情况B：队列指认（Police Lineup）

100个嫌疑人（1:5）

20个谋杀犯中正确识别18个错放2个；

80个无辜嫌疑人中错抓8个。

每抓到9个谋杀犯，要有4个无辜市民被错误指控。

两者的成本效益比（cost-to-benefit ratio）差别之大，是由于两种情况中有罪和无辜的受测者之比不同。用测谎仪来进行对绝大部分是无辜的受测群体的筛选时，代价是非常大的。

*

待测目标很少发生，或者进行的是间接测量，那么检测系统的准确率就会骤然下降。

*

（假阳性和假阴性）两种类型的错误，通过不可避免的消长关系互相联系着。

*

准绳测试法（Control Question Test，CQT）

相关问题、无关问题、准绳问题（与一些宽泛模糊的事有关，设计得使完全诚实的人也感到不快，而骗子应该对相关问题产生更大焦虑，说真话的人预期对准绳问题更加紧张）。

*

司法心理学家索尔·卡斯恩（Saul Kassin）-“招供神话”（confession myth）

-无辜的人更可能放弃自我保护的权利，更可能自愿接受做谎言测试等。

-接受过审讯技术专业训练的所谓专家在辨别谎言方面并不比没受过训练的人更杰出（而只会增添自信）。

*

数据挖掘技术在准确性上超不过测谎仪。

《数据统治世界》

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:www.tushu000.com