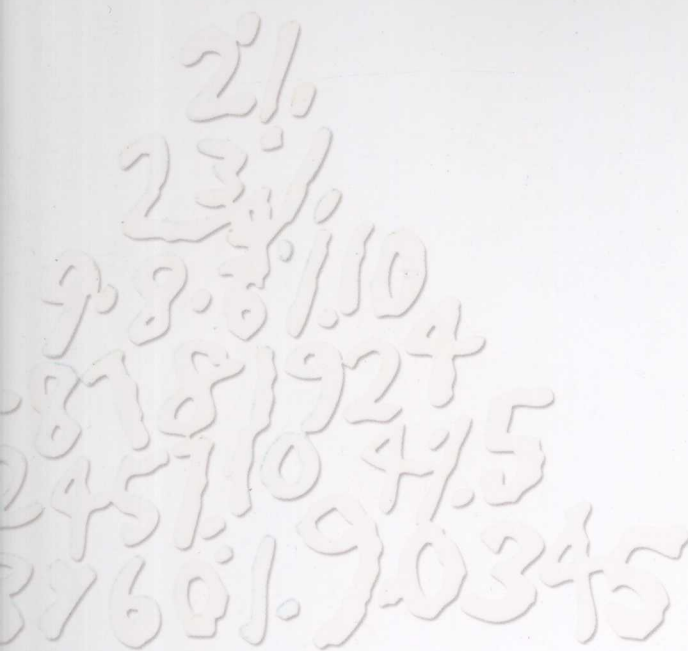


How to lie with statistics

像“故事书”一样精彩的数据造假曝光宝典



[美] 达莱尔·哈夫◎著

廖颖林◎译

统计数字会撒谎

美国统计专家畅销50年的伟大时代经典



中国城市出版社

统计数字会撒谎

(美)达莱尔·哈夫 著
廖颖林 译

中国城市出版社

· 北 京 ·

北京版权局著作权合同登记

图字:01-2008-3707

图书在版编目(CIP)数据

统计数字会撒谎/(美)哈夫著;廖颖林译.-北京:
中国城市出版社,2009.2

ISBN 978-7-5074-2085-2

I. 统… II. ①哈…②廖… III. 经济统计—统计数据
IV. F222

中国版本图书馆CIP数据核字(2009)第015708号

Copyright©1954 by Darrel Huff and Irving Geis

This edition arranged with POLLINGER LIMITED

through BIG APPLE TUTTLE-MORI AGENCY, LABUAN, MALAYSIA.

Simplified Chinese edition copyright©2009 by CHINA CITY PUBLISHING HOUSE

ALL rights reserved.

策	划	王 立
责	任 编 辑	王月芳 唐 浒 郑良苹 曾纪洲
封	面 设 计	蒋宏工作室
责	任 技 术 编 辑	张建军 阮中强 杨冬梅
出	版 发 行	中国城市出版社
地	址	北京市海淀区太平路甲40号 (邮编 100039)
网	址	www.citypress.cn
电	话	(010)63275378 (营销策划中心)
传	真	(010)63489791 (营销策划中心)
总	编 室 信 箱	citypress@sina.com 电话: (010) 52732057
投	稿 信 箱	world66@263.net (营销策划中心)
经	销	新华书店
印	刷	北京中科印刷有限公司
字	数	80千字 印张5.5
开	本	880x1230 (毫米) 1/32
版	次	2009年3月第1版
印	次	2009年3月第1次印刷
定	价	28.00元

版权所有, 盗版必究。举报电话: (010)52732057



译者的话

“自从使用了某某牌牙膏，我们的蛀牙减少了 23%。”或许你刚被这样的广告宣传攻陷，对这款新牙膏的“神奇功效”深信不疑，但是美国统计专家达莱尔·哈夫告诉你：“没有比这更无聊的广告了。”没错，问题就在这个“23%”上。

在一个用事实说话的社会，我们接触到了越来越多的统计数据和资料，例如各种经济数据、证券信息、投资可行性研究报告、公司财务报告等。但是却有不少销售员、公关公司、广告撰稿人等在滥用书中所揭露的“数据造假方法”来蒙蔽对数据知识不甚了解的客户、消费者和上司。面对这些良莠不齐、真伪并存的数据或资料，我们需要去粗取精、去伪存真的过程。我们又该如何进行鉴别？

20 世纪 50 年代，美国的各大媒体和宣传机构就已经开始越来越重视利用统计——“这个神秘的语言”——说话，然而大量的统计数据、统计资料由于主、客观的原因被滥用，很难起到描述事实、传递信息的作用。相反，还往往对读者形成误导。

达莱尔·哈夫 (Darrell Huff)——一位具有深厚统计背景的新闻记者——发现了这一现象。他在广泛调查的基础上，从报刊、杂志、书籍中，从美国统计学会一些统计学家提供的实例中，

收集了大量案例,并在 1954 年写下了《How to Lie with Statistics》一书。该书一经出版,便畅销美国,成为美国 20 世纪 50 年代的畅销书之一,并受到了当时美国各种书评杂志的好评,至今依然常常被美国不少权威媒体所引用。

《管理评论》认为:“哈夫先生用如此生动的,充满人情味的方式来论述统计这个干巴巴的课题,真是一剂灵丹妙药……我们太需要这本书了,它虽然娱乐性强、浅显易读,却十分具有说服力。”

《图书期刊》如此评价:“作者和制图者倾注了全力,给大家提供了一本十分轻松活泼的读物和卡通画。它们能给你带来娱乐,又能引发思考,而且还揭穿了许多统计方法的谎言。”

《大西洋》评价道:“这是一本具有善意破坏性的书,读完它后,你对于‘万能统计’的信任将大大降低。”正是由于这本书融娱乐性和知识性为一体,使它成为一本具有影响力的著作。

《How to Lie with Statistics》一书还列选英国作家利奥·高夫撰写的《25 本投资经典:透视所有时代最伟大的投资名著》(金融时报出版社,1999 年版)。该书出版至今,多次重印并被译成多国文字。

虽然这本书的写作距今已有将近 50 年,但书中所体现的思想和方法仍然相当实用。如果上网查询,输入关键字“How to Lie with Statistics”,我们将发现:相匹配的查询结果不是几条,几十条,而是成百上千条。事实上,目前在美国几乎所有知名的网上书店,仍然可以购买到好几个版本的《How to Lie with Statistics》。而且,由于该书的近 50 年来的畅销,使得“编造虚假信息”这一命题受到了人们持续的普遍的关注。

哈夫先生独具匠心地重构了全书的行文。从第1章到第9章,作者想象自己是一个技艺高超的惯骗,正在向后起之秀面授各种行骗的技巧。协助行骗的工具很多,包括:有偏的样本、精心挑选的平均数、遗漏某些重要的数据、样本的误差、统计图、平面图、不匹配的资料、混淆相关关系与因果关系以及不正确地使用资料。对每一种工具的功能和使用场合,他都一一做了详细的介绍。第10章是一个总结性的章节,在此章中,作者摒弃了前9章使用的描述手法,直接正面地阐述了“对统计资料”应该“提出的五个问题”:(1)“谁说的”;(2)“如何知道的”;(3)“是否遗漏了什么”;(4)“是否偷换了概念”以及(5)“资料是否有意义”。通过寻找这5个问题的答案,读者能初步判断资料是否真实可信。

像“故事书”一样有趣的案例是该书的又一特点。哈夫先生精选读者生活中最常见的案例,再以幽默、活泼的方式叙述深奥的统计知识,不知不觉中读者已经被书中的“故事”所深深吸引。在序言中,作者将伪装统计资料的人比喻成一个在昏暗的灯光下没日没夜工作的“白衣英雄”,将本书描述成“骗子的行骗宝典”,类似手法的描述散布在书中,十分形象。虽然书中也出现了一些统计术语和统计方法,但在作者重说明轻证明的描述下,即使你从来没有接触过统计,也能很好地掌握其中的思想。

回到国内,最近几年,关于“平均工资”的讨论十分热烈。早在2004年,广州市人大代表就曾对广州市政府工作报告中的“全市职工平均工资为28237元”提出了质疑,他们指出在平时接触的市民中,普遍反映年工资收入基本达不到这个水平,有的甚至

相距甚远^①。在今年召开的全国两会上,有全国代表明确指出:“看职工工资不能只看平均数,还要关注那些平均线以下的人群。”^②那么,我们是不是不该使用平均工资?其实,关键的问题并不在平均数这个统计指标上,而在于人们如何理解并正确地使用这个统计指标。

处理海量数据的常用方法是计算平均数,以了解一组数据的平均水平。平均数的计算首先遇到的问题是这组数据的范围是什么?用统计术语来说,就是统计的口径是什么?例如,广州市统计局对“全市职工平均工资为 28237 元”统计口径的解释是:“职工统计中有七类人员没有列入范围,而这部分人正属于收入较低的群体。如乡镇企业从业人员,私营企业从业人员,城镇个体劳动者,离休、退休、退职人员,再就业的离、退休人员,民办教师以及其他按有关规定不列入职工统计范围的人员。”对统计口径的说明能够一定程度上解决市民关于平均工资过高的疑虑。

接下来,第二个问题是计算中采用了哪种平均数?平均数有不同的种类,最常见的有 3 种:均值、中位数和众数。关于这 3 种平均数的概念、计算和分析,Darrell Huff 在第二章给出了十分详细的解释和相当精彩的案例,我就不罗嗦了。我们平常所说的平均工资基本上都是均值,即所有人的收入相加除以人数。均值具有计算简单,易于理解的特点,但是它也有自身不足之处:当数据的分布呈现正偏态时,均值往往偏离一般水平,并且高于一般水平。虽然人人都梦想良田万顷、豪宅林立,但是达成此梦想的人

① 参见 <http://www.southcn.com/news/dishi/guangzhou/shizheng/200403260379.html>.

② 参见 http://www.china.com.cn/2008lianghui/2008-03/09/content_12082013.html.

毕竟是少数,因此就收入而言,往往高收入的人比例偏少,而中低收入的人比例很高,收入分布是典型的正偏态分布,这样一来,平均工资偏高就十分正常了。

为了避免平均数引起误解,提供更多的信息是较好的办法。例如,在给出平均数的同时罗列出收入的频数分布,即给出不同收入区间的职工比例,从而对职工收入情况有更全面的了解。又例如,在给出平均数的同时给出各组的组平均数,可以是不同行业职工的平均收入,可以是不同职位职工(管理层和职员层)的平均收入。再例如,在给出平均数的同时罗列出最低收入,以及享受最低收入职工的比例。

以上平均工资仅仅只是经济领域的一个例子。生活中我们还可能接触到各种各样的数据,虽然不是每个人都需要练就火眼金睛,或者武装到牙齿,成为数据打假的斗士,但是掌握书中的工具,至少不会轻易地受到“全国牙防所”的诱惑。

培根曾经说过:“如果一个人以种种肯定的立论开始,他必将终止于各种怀疑;但如果他愿意抱着怀疑的态度开始,那么他必将获得肯定的结论。”我想对数据的判断和接收也是如此。

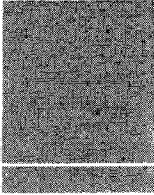
《How to Lie with Statistics》可以有多种译法。可直译为《如何利用统计撒谎》,这样基本体现了本文前9章的行文风格。但是为了更加贴近读者,承续文中活泼的文风,经过与出版社的商议,此版翻译将书名确定为《统计数字会撒谎》。书中,中国城市出版社的编辑为该书的出版和插画做了大量辛勤的工作,在此表示诚挚的谢意。

在我看来《How to Lie with Statistics》是一个“一顾倾人城，再顾倾人国”的绝色美人，而我这个译者，有点像化妆师，虽然水平不高，辜负了美人的姿色，但是好在美人自身的底子很好，化妆后的效果虽然不能让人惊艳，但我已经竭尽所能，希望她是个“第二眼美女”。敬请各位专家、读者手下留情，不要拍砖。剩下的嘛，当然是希望读者朋友们能够“抱得美人归”了。

廖颖林*

2009年2月于上海财经大学

* 廖颖林，上海财经大学经济学博士。统计学研究青年学者。现为上海财经大学应用统计研究中心研究人员，九三学社上海市委青年委员会委员。主要研究方向：顾客满意度指数研究、经济管理统计分析。



致 谢

在多方的帮助下,经过广泛的收集,我得到了散见在本书各章当中的关于狡辩和欺诈的小例子。在我通过美国统计学会发出呼吁后,许多专业统计学家——请相信我,他们和其他人一样痛恨统计资料的滥用——从他们自身收集的资料中为我提供了大量的实例。这些人,我猜想,将乐意成为无名英雄。同样,我在许多书籍中发现了有价值的案例,仅列出其中主要的书籍:Martin A. Brumbaugh 和 Lester S. Kellogg 所著的《商业统计学》;Hadley Cantril 的《公共观点的测定》;Willard Cope Brinton 的《图表表示法》;Frederick E. Croxton 和 Dudley J. Cowden 的《实用商业统计学》;George Simpson 和 Fritz Kafka 的《基础统计学》以及 Helen M. Walker 的《基础统计方法》。

Darrell Huff

序 言

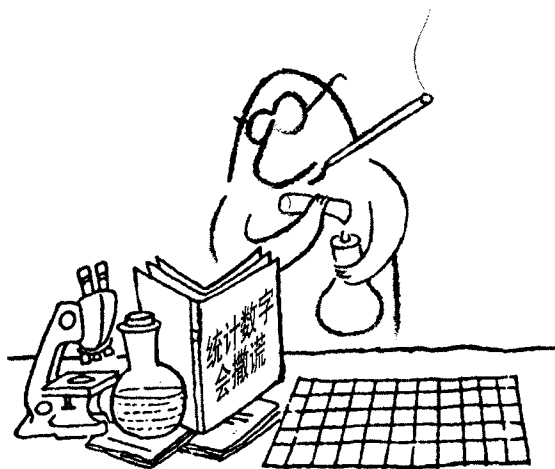
我的岳父从爱荷华州到加利福尼亚州不久便对我说：“你们这儿治安不好。”在他所阅读的关于加州的报道的确如此。但是，这些报道通常来自一份爱荷华州的报纸。这份报纸不会轻易忽略掉加州发生的任何犯罪行为，虽然它也报道本州的谋杀案，但看起来它更乐意大肆渲染加州出现的同类情况，而且还因此而闻名。

我岳父得出的这个结论是建立在明显有偏样本基础之上的，是一个随意的统计结论。类似于其他更为精致的统计结论，也存在着证据与结果不匹配的问题，因为这些结论都在假定：报纸专栏中对犯罪行为的报道是测量犯罪率的工具。

几年前，十来个调查人员独立地发表了关于抗组织胺药的试验数据。所有的数据都证明，在经过抗组织胺药物治疗后，相当高比例的感冒能够治愈。这一结论引起了传媒的大肆宣传报道，抗组织胺药的广告铺天盖地，医药界也掀起了此类药物的生产热潮。人们对健康永恒不变的追求造就了这种热潮，但奇怪的是，人们拒绝越过统计资料去注意一下早就了解的事实。正如一位幽默的非医学权威人士，亨利·G·菲尔森(Henry G. Felsen)不久前所指出的——正确的治疗的确能在7天内治愈感冒，但是即使不进行任何治疗，感冒也能在一个星期内痊愈。

事实往往在所见所闻之外。平均数、作用关系、趋势和图表总是与看上去的不一致。虽然经验告诉我们“眼见为实”，但眼睛告诉我们的“真相”或许隐瞒了部分事实，或许夸大了事实。

统计这种神秘的语言，在一个靠事实说话的社会里是如此地吸引眼球，但有时它却被人利用，并成为恶意夸大或简化事实、迷惑他人的工具。在报告社会经济趋势、商业状况、民意调查和普查的大量数据时，统计方法或者统计术语是必不可少的。但如果作者不能正确理解并恰当地使用这些统计语言，而读者又并不能真正了解这些术语的含义，那么，统计结果只能是废话一堆。



在科技主题的文章中，经常充斥着滥用统计资料的现象，每一个数据都很有力地打消你的疑虑，都张着嘴告诉你这是对的，但是人们很难联想到这样一种场景：灯光如豆的实验室中，“白衣英雄们”仍在不计报酬、不辞辛劳地包装这些数据。统计，

就像最神奇的化妆术,只要略施粉黛,东施尤胜西施。巧妙伪装的统计资料胜过希特勒的弥天大谎,虽然它也会引起人们的误解,但制造它的人却能巧妙脱身。

这本书是一本如何利用统计瞒天过海的入门读物。看上去,它很像骗子的行骗宝典。但或许我可以模仿一个已退休的窃贼——如何神不知、鬼不觉地撬开一把锁,他的回忆录达到了研究生课程的水平——替这本书说句公道话:毕竟,骗子对于行骗的技巧早已胸有成竹,而诚实的人出于自卫也应该掌握它。

Darrell Huff

目 录

序 言

Chapter 1 内在有偏的样本 (1)

我们来看一则以前的新闻报道,“1924 级的耶鲁毕业生平均年收入为 25111 美元。”好家伙,这些人干得真不赖!大家都把孩子送去耶鲁和牛津吧,那就是高薪的代名词,年纪轻轻就能享受幸福的生活。

可是,等一等,在充满怀疑的惊鸿一瞥后,关于该数字的两个疑点凸现出来:它居然惊人的精确,它也大得令人难以置信。

Chapter 2 精心挑选的平均数 (21)

房地产商费尽心思大力宣传,你所居住的小区是高档小区,你邻居的年均收入是 10000 英镑;而该区域的纳税人委员会却反复向政府强调,这里居民的平均年收入只有 2000 英镑,是该减减税赋了。

到底谁在说谎?事实上,他们都是诚实的。两个数字都是正规的平均数,来自相同的居民,根据相同的收入,计算方法也完全正确,奥妙何在?

Chapter 3 没有披露的数据 (33)

“自从使用了多克斯牌牙膏,我们的蛀牙减少了 23%。”

让我再引用一篇文章的标题——《现在就来预测孩子将来长多高》。“只需要利用现有的身高,再查表中的比例即可。”标题如是写道。

没有比这个“23%”和“身高表”更无聊的广告和报道了!可他们看起来是如此的煞有介事,问题出在哪?

Chapter 4 毫无意义的工作 (51)

彼得和琳达做了公认最好的智力测验——斯坦福-比内测验,结果是琳达的智商为 101,彼得的只有 98。专家告诉我们智商的平均数是 100,即 100 意味着“正常”。于是进一步推断,琳达是比较聪明的孩子,彼得是个笨孩子。

相信我,任何类似的结论纯粹都是胡说。

Chapter 5 令人惊奇的图形 (61)

一张图告诉你朴实的 10% 的增长,而另一张却看起来是让人振奋的 100% 的增长,别怀疑你的眼睛,截然不同的两幅图说的可是同一回事!

Chapter 6 一维图形的滥用 (69)

数字是 2:1,但视觉效果却是 8:1。嘴上说的是

1.5 倍,看起来却是 3 倍……或许你正在被这些图形所振奋着。

Chapter 7 不完全匹配的资料 (79)

一篇来自著名实验室的报告:在 11 秒钟内仅仅半盎司该药的剂量就杀死了试管中 31108 个病菌。

随处可见某种榨汁机的广告:“经过实验室的证明”该榨汁机的“榨汁功能增强了 26%。”……

听起来真不错,这是货真价实的“挂羊头卖狗肉”。

Chapter 8 相关关系的误解 (93)

抽烟与大学成绩;独身与上大学;身上的跳蚤与健康;房屋顶上白鹤鸟巢的个数与荷兰某个家庭中已出生孩子的人数;马萨诸塞州长老教会会长的收入与哈瓦那朗姆酒的价格……是风牛马不相及? 还是亦步亦趋、息息相关? 真实的结果一定让你大跌眼镜。

Chapter 9 如何进行统计操纵 (107)

怎样在一年内获得 22500 美元的总收入? 你只需 1 个妻子(或丈夫)和 13 个孩子。

“现在就购买圣诞礼物,你将节省 100%。”精明的商家居然成了圣诞老人,开始免费馈赠了。

有些人很害怕你知晓其中的“技巧”。

Chapter 10 如何反驳统计资料 (129)

谁说的？他是如何知道的？遗漏了什么？是否有人偷换了概念？这个资料有意义吗？

提这 5 个问题，凭双眼就识破并揭穿虚假的统计资料；更为重要的是在具有欺骗性的数据海洋中找出可靠有用的资料，不再让你的关键决策南辕北辙。