


统计数据 会说谎

畅销美国50年，
经久不衰的
统计学经典
作品

让你远离数据陷阱

[美] 达莱尔·哈夫 (Darrell Huff) © 著
靳琰 武钰璟 © 译

世界上有三种谎言：
谎言、
弥天大谎
和统计数据。



HOW TO LIE
WITH STATISTICS

统计数据 会说谎

让你远离数据陷阱

[美] 达莱尔·哈夫 (Darrell Huff) 著
靳琰 武钰璟 译

HOW TO
LIE
WITH
STATISTICS

图书在版编目 (CIP) 数据

统计数据会说谎 / [美] 达莱尔·哈夫著; 靳琰,
武钰璟译. -- 北京: 中信出版社, 2018.1
书名原文: How to Lie with Statistics
ISBN 978-7-5086-8214-3

I. ①统… II. ①达… ②靳… ③武… III. ①经济统计—统计数据 IV. ① F222

中国版本图书馆 CIP 数据核字 (2017) 第 246586 号

How to Lie with Statistics by Darrell Huff

Copyright: © 1954 by Darrell Huff.

This edition arranged with Pollinger Limited through BIG APPLE AGENCY, INC., LABUAN, MALAYSIA.

Simplified Chinese edition copyright © 2018 CITIC Press Corporation

All rights reserved.

本书仅限在中国大陆地区发行销售

统计数据会说谎

著 者: [美] 达莱尔·哈夫

译 者: 靳 琰 武钰璟

出版发行: 中信出版集团股份有限公司

(北京市朝阳区惠新东街甲 4 号富盛大厦 2 座 邮编 100029)

承 印 者: 山东鸿君杰文化发展有限公司

开 本: 880mm×1230mm 1/32

印 张: 5.5 字 数: 100千字

版 次: 2018年1月第1版

印 次: 2018年1月第1次印刷

京权图字: 01-2014-0416

广告经营许可证: 京朝工商广字第 8087号

书 号: ISBN 978-7-5086-8214-3

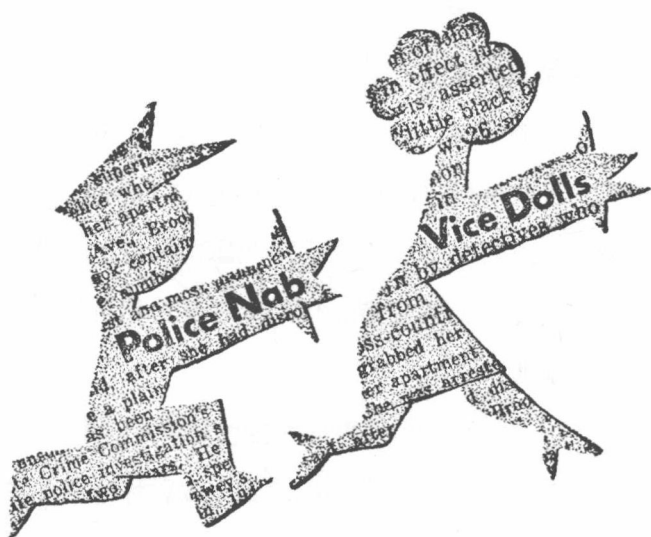
定 价: 39.00元

版权所有·侵权必究

如有印刷、装订问题, 本公司负责调换。

服务热线: 400-600-8099

投稿邮箱: author@citicpub.com



我的岳父从艾奥瓦州搬到加利福尼亚州没几天，就对我说：“这个地方的犯罪事件太多了！”他所读的报纸的确报道了许多犯罪事件。这份报纸从不放过报道当地的任何一起犯

罪事件，而且以注重报道谋杀案闻名，其详尽程度超过了艾奥瓦州的任何一家大型报社。

我岳父的这一结论属于非正式统计。这个统计基于一个明显带有偏差的样本。与其他许多较为规范的统计一样，这个统计也存在虚假的成分。因为这个统计认为报纸上报道犯罪事件版面的大小是衡量犯罪率高低的标

几年前，十几位调查人员分别发表了一份关于抗组胺剂药物的报告。每份报告都表明，服用该药物后感冒会明显好转。紧接着就是各种各样的渲染，至少广告商是这么干的，于是这种药物被大量生产。造成这一结果的原因是人们一直对药物有着大量需求，也从未越过统计学去了解自己早已知道的事实。正如亨利·G·费尔森（Henry G. Felsen，他是一位幽默作家，但绝不是医学专家）很久以前所说的，适当的治疗可使感冒在七天之内痊愈，但如果任其自然发展，感冒则会持续一个星期。

那么，你读到和听到的事情大多也是如此。那些平均数、各种关系、趋势图以及图表并不总是一致。你所看见的未必是真实情况，事实上这些数据要么被过于夸大，要么被隐瞒。

在如今用事实说话的社会中，统计这种神秘的语言是如此诱人，但它却被人弄成了耸人听闻、华而不实、迷惑不清且过分简单的东西。在报道社会和经济趋势、商业状况、民

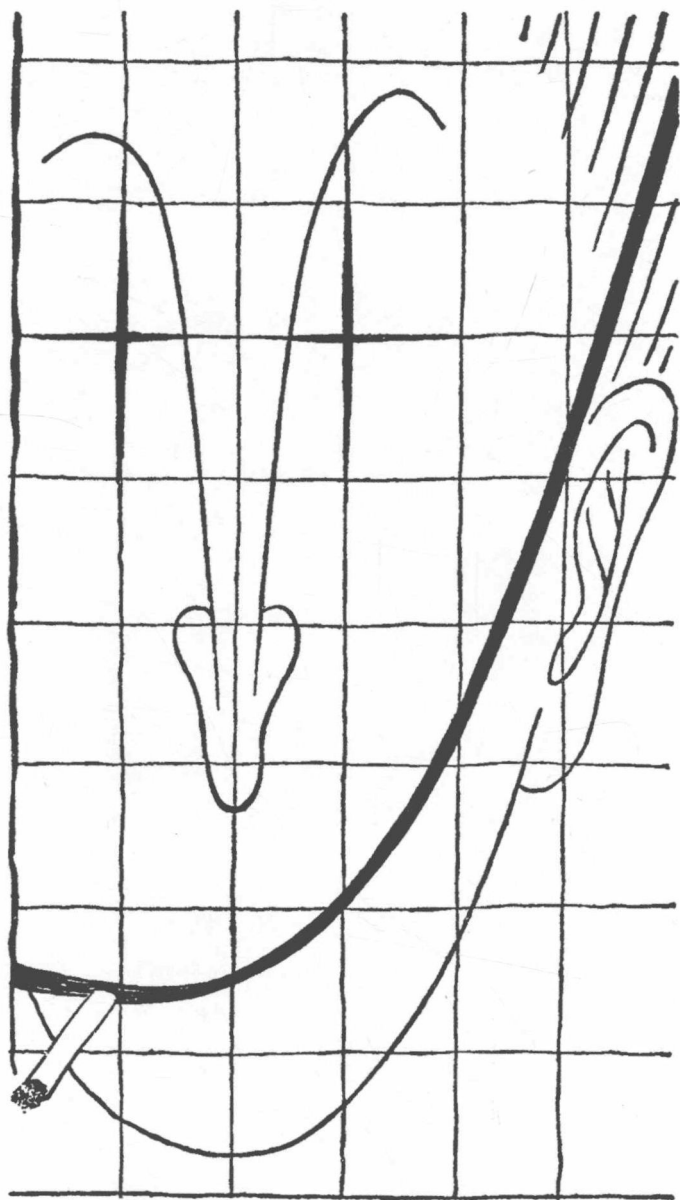
意调查、普查时要用到大量数据，此时统计方法和术语就不可或缺。但是，如果作者们不能诚实报道甚至根本没有理解这些统计词汇，读者也就无法明白作者所说的内容，那么这些统计结果就只能是无稽之谈。

如今科普读物很受欢迎，但这些书籍往往滥用统计数据，这让人联想到这样一幅场景：灯光昏暗的实验室里，一个穿着白大褂的人忙得天昏地暗，还没有加班费，正如“加一点粉末，再加一点颜料”，这样的统计数据把许多重要的事实搞得面目全非。一个精心包装过的统计结果比希特勒的弥天大谎还要厉害，因为它虽然误导了你，但你还没法去指责它。



在教人如何用统计“行骗”的同类书籍中，这本书只能

算是初级读本。它看起来像是一本“骗子指南”。也许我可以为这本书稍做辩解：这就好比一个洗手不干的窃贼出了一本回忆录，书中描述了如何能不出声地撬开门锁，其水平之高让研究生都佩服。既然“骗子”都已经熟练掌握了这些诡计，老实人又怎能不学来用于自卫呢？



引 言 // III

第一章 带有偏差的样本 // 001

第二章 精挑细选的平均数 // 019

第三章 没有透露的小小数据 // 033

第四章 无事瞎忙 // 051

第五章 惊人的图形 // 061

第六章 一维图形 // 071

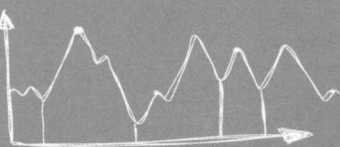
第七章 看似相关的数据 // 081

第八章 因果颠倒 // 097

第九章 如何操纵统计 // 113

第十章 如何反驳统计数据 // 137

致 谢 // 161



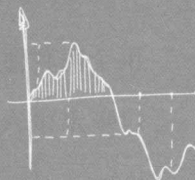
How to Lie with



S tatistics

第一章

带有偏差的样本



在评论《纽约太阳报》(Sun)有关报道时,《时代》(Time)杂志曾特别指出:“耶鲁大学1924届毕业生平均年收入高达25 111美元。”

哇哦,他们实在太厉害了!

但稍等一下。这一令人惊叹的数据究竟意味着什么?它是否像表面看起来的那样,足以证明如果你将自己的孩子送进耶鲁大学,你就可以安享晚年,你的孩子也可衣食无忧了?

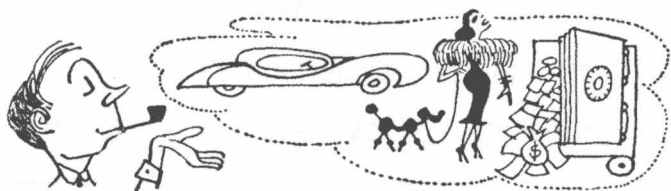
在充满怀疑地乍一看之后,我们发现这个数据有两个可疑之处:首先,它精确得让人吃惊;其次,它给人带来难以



置信的好处。

把那些四处分散的人的平均年薪搞清楚并精确到个位数，这个可能性实在很小。就算你对自己去年的收入要做到一清二楚都不太可能，除非这些收入全部源自薪水。而 25 111 美元的年收入通常并非全部来自薪水，这部分耶鲁毕业生很可能还有零零散散的各种投资。

毫无疑问，这一令人羡慕的平均数是根据耶鲁毕业生自己报出的数据计算而来的。即便 1924 届毕业生在位于纽黑文市的耶鲁大学念书时谨遵其信用制度，我们也不能因此肯定多年之后，他们依然恪守该信用制度，他们口中的这些数据依然真实可靠。当被问及收入时，他们中的有些人难免会因虚荣或自负而有所虚报；有些人恐怕会尽量瞒报收入，特别在涉及所得税的问题上，他们也许会因担心与其他表单上的数据互相矛盾而犹豫不决，那些税务官们会不会看出什么端倪来呢？这两种倾向——虚报和瞒报——也许会互相抵消，但事实上这不可能。其中一种倾向也许会占上风，但究竟哪种占上风我们不得而知。



根据常识，这一数据绝对不真实，那么就让我们来分析一下。我们先来看看导致最大误差的原因，就是这个原因炮制出了某些人“平均年收入高达 25 111 美元”的神话，然而事实上，他们的真正年收入也许仅仅接近这个数字的一半而已。

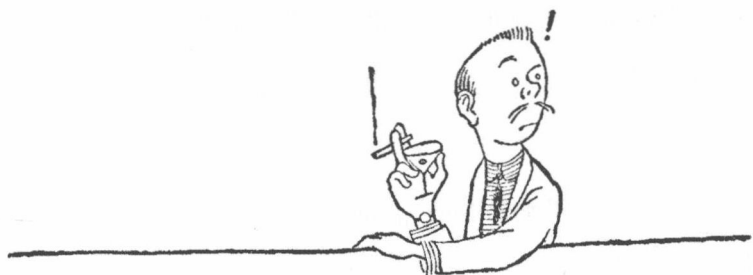
这就是抽样过程。在处理各类问题时，抽样过程是你所接触到的统计学主要内容的核心。其基础原理非常简单，尽管在实践中其精细到已衍生出了各种分支原理，而其中一些并不可信。如果你有一桶豆子，有些是红色，有些是白色，现在只有一个办法能让你准确得知每种颜色的豆子各有多少：一颗一颗地数。不过，还有个方法更加简便可行：随便抓一把豆子，然后数出红色豆子有多少颗，设想红色豆子在这一把豆子中所占的比例和桶里红色豆子在所有豆子中所占比例相同，由此便可大致知道桶里红色豆子究竟有多少颗。如果样本规模够大，且选择得当，那么它在大多数情况下完全可以代表整体。否则，这种方法还不如高明的猜测来得精确，除了营造出一种所谓的科学精确性的虚假氛围，这种方法毫无可取之处。可悲的是，各种各样的结论就是从这些要么带有偏差，要么过于微小或二者兼而有之的样本中获得的，而我们却并不知道自己所读到的这些结论或自以为清楚的这些结论来自这样的样本。

有关耶鲁毕业生的收入报告就来自这样一个样本。我们之所以如此肯定是因为理智告诉我们，要联络到所有尚在人世的1924届毕业生根本不可能。多年之后，许多人的地址已经无从得知。



而且，就算知道了这些人的详细地址，他们也未必会答复调查问卷，尤其是这种关系到个人隐私的问题。就一些邮寄问卷而言，有5%或10%的回复率就已然很高了。耶鲁大学毕业生的收入调查问卷的回复率也许更高一些，但绝不可能高到百分之百。

因此，我们得出结论——该收入数据来自这样一个样本，即样本对象由既有明确地址又回复了问卷的毕业生成员组成。该样本具有代表性吗？也就是说，我们能否假定该样本群体成员的收入与那些联系不到或没有回复问卷的成员的收入处于相同水平？



在耶鲁毕业生名单中，谁是那些“地址不详”的迷途羔羊呢？是华尔街富商、公司董事、制造公司和公共事业部门的高管等这些高收入者吗？不，要想获知这些富人的地址并不难。即使有些巨商富贾早已忘了和校友办公室保持联系，我们仍可以在《美国名人录》(Who's Who in America)以及其他通信录中找到他们的地址。不妨这样猜测一下，那些失去联系的人在获得耶鲁大学学士学位之后的多年间，没有实现当初的抱负、功成名就。他们沦落成了普通的小职员、机械工、流浪汉、失业的酒鬼、勉强糊口的落魄作家和艺术家……在25 111美元的收入单上，这些人占据了一半甚至更多。他们不常参加同学聚会，因为他们连路费都没有。

哪些人将问卷丢进了废纸篓？我们对此不能确定，但至少能猜到这些人的工资还没有高到能拿出来炫耀的份儿上。他们有点像这样的人：第一次领薪水时，发现工资单上附了张小纸条，要求对自己的薪水保密，不要打探同事之间的薪