

怎能利用 统计撒谎

〔美〕达瑞尔·哈夫 著
欧文·盖斯图

沈恩杰 马世宽译
马 安 吴世农校



中国统计出版社

怎能利用统计撒谎

[美] 达瑞尔·哈夫 著
欧文·盖斯 图
沈恩杰 马世宽 译
马 安 吴世农 校

中国统计出版社

How to Lie with



By

DARRELL HUFF

Pictures by IRVING GEIS

W · W · NORTON & COMPANY

New York · London

本书根据 1982 年版本译出。

怎能利用统计撒谎

ZENNENG LIYONG TONGJI SAHUANG

602 8/17
[美] 达瑞尔·哈夫 著
欧文·盖斯 图

沈恩杰 马世宽 译
马 安 吴世农 校

*

中国统计出版社出版

新华书店北京发行所发行

北京印刷一厂印刷

*

787×1092毫米 32开本 4印张 7.5万字

1989年11月第1版 1989年11月北京第1次印刷

印数: 1— 5 000

ISBN 7-5037-0304-0/C·159

定价, 1.90元

谎话有三种：一般的谎话、糟糕透顶的谎话和统计。

——迪斯雷利

统计思维，如同读写能力一样，总有一天会成为讲求效率的公民所必需的本领。

——H·G·威尔斯

使我们陷入困境的，倒不是我们不知道的东西，而是我们所知道的不正确的东西。

——阿特默斯·沃德

我有一个重大的题目[统计]要写，但是深感我的写作能力不足以将它表述得明白易懂而不损害它的准确性和透彻性。

——弗朗西斯·高尔顿爵士

译者前言

《怎样利用统计撒谎》是美国统计学家达瑞尔·哈夫的名著。该书迄今已重印 34 次，畅销不衰。许多统计学教科书将该书列为重要参考文献，或者引用该书的警句和事例。按全书主旨和我国情况，中译本改称《怎能利用统计撒谎》。

在经济发达的美国，统计数据是人们传递信息、了解社会、制定决策的基本依据。但也有些人用统计骗人。作者对误用统计、滥用统计、利用统计骗人的现象给予了无情的揭露和鞭挞。本书取材广泛，例子众多，内容生动，插图幽默，妙语叠出，耐人寻味。本书在美国出版后曾遭到某些非议，然而作者的真实用意在于提醒人们正确地进行统计、利用统计，而非贬低统计。这精神贯串全书，但在前几章中，作者使用了寓庄于谐的“文学技巧”，在最后一章，则正面地、集中地阐述了如何辨伪求真，从统计数字中获得真实的信息。这种认真、严肃、求实的科学态度，着实令人钦佩。

随着我国社会主义商品经济的发展，统计作为认识社会和制定决策的有力工具，日益受到人们的重视。但是在一些报刊上也不乏误用统计方法和统计数据的例子。因此我们选译此书，供统计工作者和广大社会公众参考借鉴。

限于译者水平，译文难免有不妥之处，请读者指正。

译者

1989 年 10 月

鸣 谢

使这本书辛辣有趣的那些荒诞、诡诈的事例，其所以收集得比较广泛，是因为得到了许多方面的帮助。我通过美国统计学会发出呼吁之后，许多统计学家（请相信我，他们对统计的误用象任何人一样深恶痛绝）把他们搜集的一些事例提供给我。我猜想，他们一定不愿意让我在这里提到他们的姓名。另外，我还在一些书中找到一些有价值的事例。这些书主要是：*Martin A. Brumbaugh* 和 *Lester S. Kellogg* 的《商业统计学》，*Hadley Cantril* 的《民意测定》，*Willard Cope Brinton* 的《统计图》，*Frederick E. Croxton* 和 *Dudley I. Cowden* 的《实用商业统计学》，*George Simpson* 和 *Fritz Kafka* 的《统计学基础》，*Helen M. Walker* 的《基本统计方法》。

目 录

鸣 谢

引 言	(1)
1. 有偏性的样本	(5)
2. 选择得当的平均数	(19)
3. 没有透露的小数字	(28)
4. 无事空忙	(43)
5. 惊人的曲线	(49)
6. 平面图	(55)
7. 牵强附会的数字	(63)
8. 死灰复燃的伪因果论	(74)
9. 怎样变统计戏法	(85)
10. 如何识别统计数字的真实性和有用性	(104)



警察追捕卖淫少女

引 言

“这里的犯罪事件真多！”我的岳父从衣阿华来加利福尼亚不久，便这么说。的确，从他读的那份报纸上看，情况就是如此。那家报纸从不漏登当地的任何一个犯罪事件，并因对衣阿华州的一个凶杀案所作报导比该案件发生地主要日报所作报导更为详尽而闻名。

我岳父的结论是一种非正式的统计。它根据的是一个样本，一个带有明显偏差的样本。正象许多复杂的统计量一样，它犯了根据不足的错误。因为它假定，报导犯罪事件的报纸版面的多少是测度犯罪率大小的一个指标。

几年前，十几个调查人员各自独立地提出了关于一种抗组织胺药片的数字。每个调查报告都表明，这种药片对感冒的治愈率达到一个相当高的百分比。接着是大事宣扬，至少在广告中是这样，于是该药片的生产一片兴旺。这是由于人们长期存在这样的愿望，也是由于人们奇怪地拒绝越过统计去留心一下人们长久以来所已熟知的事实。正如亨利·G·费尔森——一位谈吐幽默，却决非医学权威的人很早就曾指出的那样，适当的治疗可以在7天之内使感冒痊愈；但若听其自然，感冒会延续一个星期。

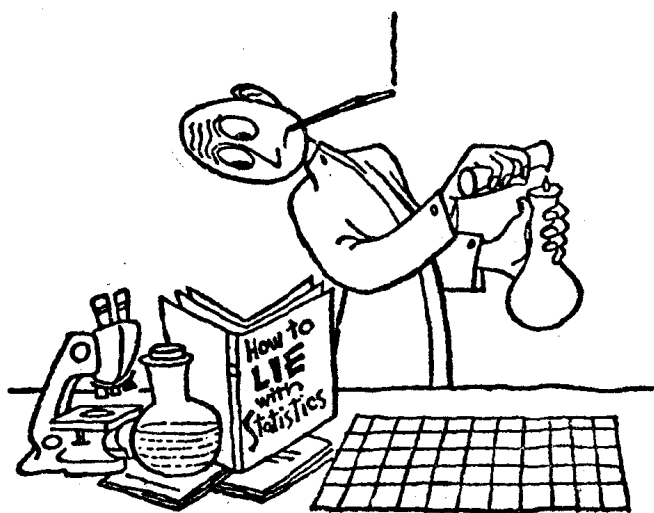
人们读到的和听到的东西有许多往往也是这样。不论是平均数、相对数、趋势线、图示，并不总是完全符合实际的。实际数量比眼睛看到的会更多或者要少好多。

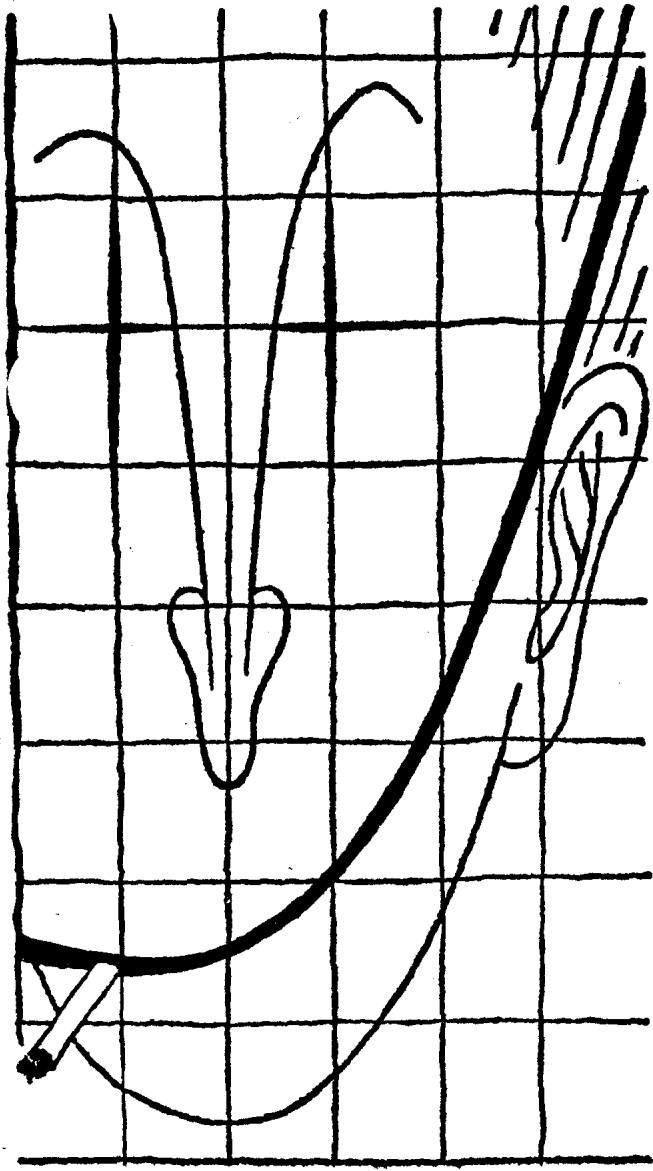
统计的秘密语言，在注重事实的社会里是这样地吸引人，但是往往被使用得成了耸人听闻的、虚夸的、迷惑人的和过分简单化的东西。在报导社会经济发展趋势、商情、民意测验、普查的大量资料时，统计方法和统计术语是必需的。但是，如果没有态度诚实而又深刻理解统计语言的作者和了解统计语言含义的读者，结果就会造成语义上的荒诞不经。

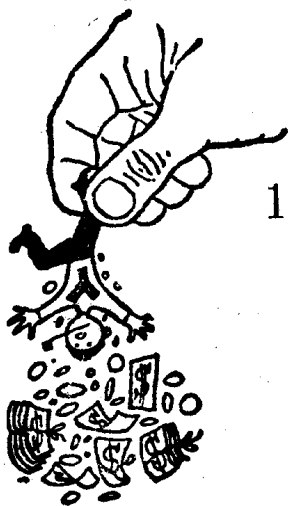
科普读物中有张画像，主人公是一位在光线幽暗的实验室中加班干活而又不拿加班工资的穿白大褂的人，而被滥用的统计则突出于画面。正如一小撮药粉，一小瓶染色剂一样，统计正在把许多重要事实搞得似是而非。一份装帧精致的统计资料，胜过希特勒的“弥天大谎”。它使人深入歧途，却不负任何责任。

本书是一本怎样用统计数字骗人的入门书。它看起来也很象一本骗子手册。一个已经洗手不干的窃贼写了一本回忆

录，它在怎样放轻脚步去撬锁这个方面达到了研究生必修课程的水平。我或许可以模仿他的样子，替本书说句公道话：骗者们早已知道这些骗人的手段，诚实的人们，你们也必须学会它们，以便自卫。







1. 有偏性的样本

《时代》杂志有一次评论纽约《太阳》报的某些报道时提到，“1924 级的耶鲁大学毕业生一般年收入 25 111 美元。”

啊，运气真好！

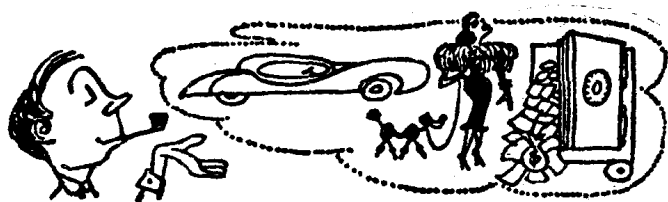
但是，等一下，这个给人印象很深的数字意味着什么呢？是否象它表面上所说的那样，足以证明如果你把你的孩子送到那里去学习，你年老时就不必工作了，他年老时也不必工作了？

一眼看去这个数字有两点很可怀疑：一是准确得令人惊奇；二是大得难以相信。

任何一群职业不同的人，其平均收入很难搞得准确到个位数字。你绝不可能把你自己的去年的收入搞得那么准确，除

非你的收入完全来自薪金。25 000 美元的收入往往不会全是薪金，有这种收入的阶层可能有较多的投资收入。

而且，这个可爱的平均数无疑是从耶鲁毕业生自报所挣收入的总额计算而得的。尽管他们1924年在纽哈芬*有优异成绩，我们也难以相信，四分之一世纪之后，他们的“答卷”还那么认真，他们自报的数字那么可靠。某些人，当别人问到他们的收入时，往往由于虚荣心或者过于乐观而夸大。



另一些人则尽力缩小，尤其是在填写所得税申报表的时候；心跳，踌躇，生怕与其他文件中的记录不一致。谁知道税务人员会看到什么材料呢！这两种倾向——虚夸和隐瞒，可能互相抵销；但事实上是不大可能抵销的。某一种倾向可能比另一种倾向大得多，但我们不知道哪一种倾向大。

我们是在说明一个常识告诉我们很难符合实际的数字。现在让我们指出这个最大的错误的来源，即产生某些人的平均收入为25 111美元这个数字的可能来源，这些人的实际平均收入很可能只达到此数的一半。

这个来源，就是抽样方法。人们平日接触到的有关各种

* 纽哈芬——康乃狄格州南部的一个海港城镇，耶鲁大学所在地。

——译者注

问题的统计资料，大都是用这种方法得到的。它的基本原理是很简单的，尽管在实际应用中它被加以各种修饰而出现了各种各样的方法，其中有些实在不值得推崇。假如你有一桶豆子，其中一些是红的，另一些是白的。要想得知每一种颜色的豆子各有多少的确切数目，那只有一种方法，就是一颗一颗地数。但是，你可以用一种更容易的办法：抓一把豆子数一下，就可以近似地得出红色豆子占多少的数字，以此代表它们在整桶中所占的比重。如果你的样本足够大，而且抽选适当，就许多目的而言，它就能很好地代表整体。否则，它可能远不如明智的猜测来得准确，除了给人一种象煞具有科学准确性的虚假印象之外，实际上毫无可取之处。可悲的事实是，我们所读到的，或者我们认为已经知道的东西，有许多是从这种有偏误的样本、或者太小的样本，甚至是既有偏，又太小的样本得出的结论。

关于耶鲁大学毕业生的报导来自一个样本。我们可以肯定这一点，因为理智告诉我们，没有一个人能掌握 1924 级毕业生中现在活着的全部成员的情况。可以肯定地说，经过 25 年后，有些人已经不知去向。



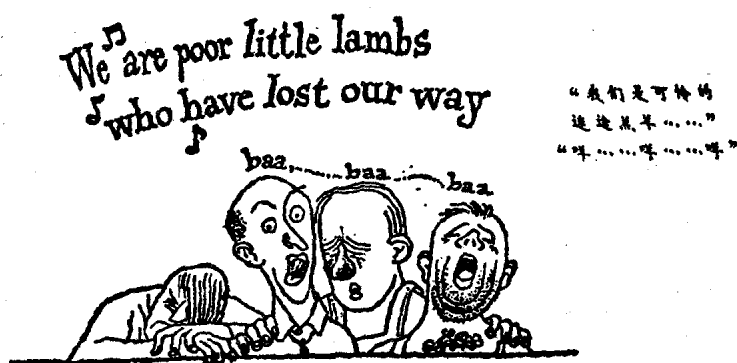
还有些人，虽然知道他们的地址，但他们之中有许多人可能对调查表、特别是有关私事的调查表不肯作出答复。某些邮寄的调查表，能够收回 5% 至 10 % 就算不坏。耶鲁大学毕业生的回信率可能高些，但绝对不会是百分之百。

由此我们得知，这些收入数字来自一个样本，该样本由全班成员中那些知道其住址，又肯提供答案的人组成。这个样本有代表性吗？也就是说可以设想这一组人的收入与无法送达调查表或者不肯提供答案的那一组人的收入是相等的吗？



在耶鲁大学毕业生通讯录上，注明住址不详的走失了的羔羊是哪些人？是华尔街老板、公司的头头、制造业或公用事业的经理这些赚大钱的人吗？不是。这些富人的地址不难找到。这个年级中那些万事亨通的人，尽管他们忘了与校友会联系，但是可以通过《美国名人录》或其他参考资料找到他们的地址。那些被漏掉的人，可能是在获得耶鲁大学文学学士学位后 25 年来不能实现其闪光誓言的一些人。他们是一些店员、技工、流浪汉、失业的酒鬼、苟延余生的作家和艺术家……这些人的收入，要把六、七个人的加在一起，才能达到 25 111 美元的数目。这些人仅仅是由于付不起旅费，就不

能那么频繁地在同班校友联欢会上登记。



把调查表随手扔进附近废纸篓的又是谁？我们不可能确切知道。但可以猜想他们之中有许多是工资还不足以炫耀于他人之前的人。他们有点象第一次领工资的人，工资袋里附着一个便条，要求对自己的工资保密，不在办公室里谈论。他们会对老板说：“您不必担心，我同您一样为此感到惭愧。”

现在可以非常清楚地看出，这个样本缺少两组最能使平均数降低的人。25 111 美元这个数字，开始得到了说明。如果它可以说是一个真实的数字，那也只能代表 1924 级中那些知道地址而又愿意站起来报告他们挣多少钱的人。即使情况如此，也要假定这些先生说的都是实话。

不要轻易地做这种假定。根据一种叫作“市场研究”的抽样调查的经验，几乎不能做这种假定。有一次，为了研究杂志读者的情况，进行逐户调查。其中的一个关键问题是：你的家庭在读什么杂志？当调查结果列成表并经过分析，得出的

结论是：许多人喜欢《哈帕斯 (Harper's)》而不喜欢《真实故事 (True story)》*。但是，出版商提供的大约同一时期的数字清楚地表明，《真实故事》发行量已经超过一百万份，而《哈帕斯》只有几十万份。这项调查的设计者们，也曾讨论是不是调整范围搞得不对头？但是，不。调整范围包括了全国各种类型的居民区。合理的结论只能是，许多被调查者对问题没有做出真实的回答。几乎所有已拆封的调查表都有点势利。

后来发现，想了解人们究竟读些什么，问是没有用的。最好的办法是到他们家去收购旧杂志。那么，看看你能收到什么吧。你可能会收到许多《耶鲁评论》和《爱情罗曼史》。当然，这种没有把握的手段，仍然不会告诉你人们爱读什么；而只能告诉你，他们有过什么。

与此相似，下次你若在报刊中读到普通美国人（最近有关普通美国人的报导很多，但大部分是不大可能的）每天刷牙1.02次（这个数字是我刚才编造出来的，但它象其他人编造的一样巧妙），你不妨先自问一下，人们是怎样了解到这种情况的？一位在许多广告中看到不刷牙是违反社会公德的妇女，自然不会对一个陌生人说她经常刷牙。这个数字对于只想知道人们对刷牙的看法的人来说，或者有用。但是，它不能反映人们的门牙接触牙刷的次数。

我们知道，一条河的水位不会高于它的源头。若是在隐蔽的什么地方设一个泵站，情况可能不同。与此相似，一个抽样调查的结果，不会比它所根据的样本更好。当资料经过

* 《Harper's》是一种知识分子阅读的较高雅的读物，《True Story》则是一种读者较广泛的通俗读物。——译者注