

HZ BOOKS
华章经管

风靡德语世界数十年
· 原书第7版 ·

统计数据 的真相

首先要掌握事实，然后你可以随意歪曲它们。
——马克·吐温

(德) 瓦尔特·克莱默 著 隋学礼 译
(Walter Krämer)



机械工业出版社
China Machine Press

· 原书第7版 ·

统计数据 的真相

(德) 瓦尔特·克莱默 著 隋学礼 译
(Walter Krämer)



机械工业出版社
China Machine Press

Walter Krämer. So lügt man mit Statistik.

Copyright © 1991, 1997 by Campus Verlag GmbH, Frankfurt/Main.

Simplified Chinese translation Copyright © 2009 by China Machine Press.

All rights reserved. This translation published under license.

本书中文简体字版由Campus Verlag GmbH, Frankfurt通过Andrew Nurnberg Associates International Ltd.授权机械工业出版社在中华人民共和国境内独家出版发行。未经出版社预先书面许可,不得以任何方式复制或抄袭本书的任何部分。

版权所有,侵权必究

本书法律顾问 北京市展达律师事务所

本书版权登记号:图字:01-2008-2446

图书在版编目(CIP)数据

统计数据的真相 / (德) 克莱默著; 隋学礼译. —北京: 机械工业出版社, 2008.12

书名原文: So lügt man mit Statistik

ISBN 978-7-111-25511-6

I. 统… II. ① 克… ② 隋… III. 统计数据—研究 IV. 0212

中国版本图书馆CIP数据核字(2008)第170193号

机械工业出版社(北京市西城区百万庄大街22号 邮政编码 100037)

责任编辑: 宁 珊 版式设计: 刘永青

北京瑞德印刷有限公司印刷·新华书店北京发行所发行

2008年12月第1版第1次印刷

165mm×205mm·10印张

标准书号: ISBN 978-7-111-25511-6

定价: 32.00元

凡购本书,如有缺页、倒页、脱页,由本社发行部调换

本社购书热线:(010) 68326294

投稿热线:(010) 88379007



人们一直宣称，世界会由数字来统治。但我所知道的是，数字会告诉我们，它被利用得是好还是坏。

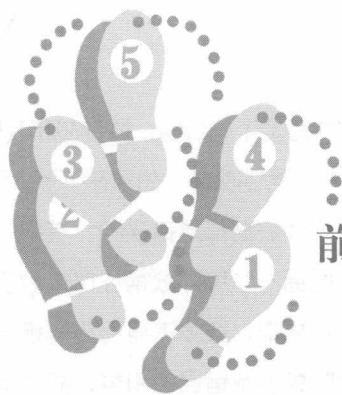
——《歌德对话录》，爱克尔曼

每个人都知道：可以人为操纵数字，可以用数字来蒙人，甚至可以用数字来撒谎。部门主管美化其季度业绩，联邦政府过于乐观地估计我们的国家（德国）能够顺利实施欧元计划。瓦尔特·克莱默对此做了详细的考察与研究，并告诉我们究竟哪里出现了问题。在通过图形处理数据的过程中，他向我们揭露了各种阴谋诡计，揭露了统计学中准确度的假象、先人为主的抽样检验、骗人的虚假趋势，揭开了人工合成的最高级，以及受操纵的平均值的真面目，并且特别留意到基本的人为操纵形式中的数据伪造者。四种基础的计算方法和一个确定的怀疑对于所有类型的数据商人来说已经足够了。

IV

对于所有希望用统计学解决问题的人来说，无论他们是想解决职业方面的问题，还是想处理个人事务，这本了不起的、能够使人产生兴趣的并给人以启迪的书都是非常专业和有益的。

瓦尔特·克莱默出生于1948年，是德国多特蒙德大学经济和社会统计学教授，著述丰富。他最重要的著作是1996年与格茨·特伦克勒（Götz Trenkler）合著的《日常误区辞典》（*Lexikon der populären Irrtümer*）。此外，2000年，他还出版了《时尚话题（德语版）》（*Modern Talking auf Deutsch*）。



前言

首先要掌握事实，然后你可以随意歪曲它们。

——马克·吐温

每天早上我们都会翻阅报纸，在读到报纸最后一页时，我们已经看到了比歌德和席勒生前看到的信息多得多的统计数据。失业者按照其来源不同进行统计，数据或多或少；癌症风险和臭氧黑洞不断增加；1/3的人生活在最低生活水平线以下；一个联邦德国公民一年内吃掉了5.8升食用冰；音乐家可以比其他人更长寿；农村的空气有益于健康，或者农村的空气不利于健康；全球到2020年将会有100亿人口；道琼斯指数又涨了30点；网球运动员B在与一个左手握拍、带着眼镜的选手比赛时从未输过，该选手比B年轻；与男性相比，超过25%的女性经常错误地拐入单行道；恐龙灭绝于6000万年前；吃猪肉会导致心脏病；外国人更容易成为罪犯；绿党的选民经常阳痿（这绝对不是开玩笑——这是一个汉堡的性学研究者在德国联邦议会上所做的说明）；人们在纽约睡觉时，

VI

最安全的地方是中央公园；还有30年我们就要与下一次的核灾害分别了，目前存在着气候灾害的威胁；如果在下一个星期天投票，某党X不会进入联邦议会。

许多统计数据都是错误的。其中一些统计数据是在人们有意识地操纵下形成的，而另一些统计数据仅仅是源于抽样时没有运用适当的方法所致。在一些统计数据中，数字本身已经是错误的；而在另一些统计数据中，人们是在用正确的数字做错误的引导，把“苹果”和“梨”重新混到一起，提出带有暗示性的问题，继续以不在意的方式推导趋势，笨拙地计算比率、份额或者平均值，任意篡改概率或者扭曲抽样检验的真实含义，上述种种做法对于许多人来说会产生一个印象：谎言和统计学就像一对不能拆分的连体婴。自从本杰明·迪斯累里（Benjamin Disraelis）谈及“世界上有三种谎言：谎言、该死的谎言，还有统计数据”之后，相关的俏皮话、讽刺语就不断地落到可怜的统计学家头上。

这种批评、讽刺虽然“微不足道”，但在某种意义上却是真实的。之所以说“微不足道”，是因为人们不仅可以使用而且能够滥用每一种统计工具。统计数据在这里肯定不是孤立无援的，而是具有许多相互关联的因素。之所以说“真实”，是因为我们所有人看到的世界都是以我们喜欢的方式看到的世界，而不是世界的客观本来面目。我们使用统计数据的目的仅仅是像“一个喝醉的酒鬼使用电线杆：首先是为了寻找能够支撑我们立场的基石，很少是为了进一步说明客观事实”（安德鲁·朗）。

作为一个统计学家，本书总结了我24年以来职业生涯的相关经验。在此我绝不是凭借着专业数据收集家的身份自以为是，从而觉得高人一等，事实是，在各种情况下，我都避免出现这种情况。提供真实的实际数据不是可能不可能的东西，而是愿意不愿意的事情，这是许多数据吹嘘者明显的相通之处。

VII

媒体有一个君子协定 (Gentlemen's Agreement), 在做一件好事时, 真实并不那么重要 (即手段要为目的服务)。如此一来, 《德国医生》杂志 (*Deutsche Ärzteblatt*) 凭借着很容易就能申请到的研究经费对艾滋病统计数据中的一项错误进行辩护, 这个错误就是: 通过所谓的“积累”, 当前的病人数量要多于实际存在的数量。“如果积累有益于这个效果的话,” 我们读到的可是白纸黑字写得清清楚楚的东西, “那么, 我们就应该让这种情形再保持一段时间。”

为了服务于主观或者客观方面好的目的, 各个理论派别对“许可说谎” (License to Lie) 这种说法都持相同的赞成意见; 从左派党团到右派党团, 正如从激进人士到保守人士, 所有人都是这么做的。如果在某种情况下数据可靠性增加, 那么那些数据美化者会根据这种可靠性误以为自己掌握着绝对真实的数据。谁能够确知世界将在20年后毁灭呢? 若不是发生了上述这种情况或是发生了其他事情, 如果这个人仅仅通过忠于事实和客观实际等常规性思考方式来进行补救工作, 他就会明显感到补救工作具有难度。“如果人们看到的某件事情越波折起伏, 形成的原因越复杂, 那么人们对此就越有兴趣”, 这就是一个美国气候研究者对其相当大胆的、根据逻辑推理所做的天气预报 (当然是错的) 道歉时所说的笑话。作为科学家, 有时必须“要引起一些小的混乱, 这样做的目的主要是为了让大家注意到自己”, 来自德国的臭氧研究专家也持相同的意见。

不仅仅是人类的拯救者想让事情表现出不同于它们实际本来的样子, 绿色和平组织的示威游行者在分享着这种尝试, 他们经常在全体人员参加的大会上扮演着一个不可能取得任何结果的集团首脑角色, 或是追逐猎物的猎人。他们研究不忠诚的婚姻配偶的所做所为, 其行为就像一个招揽更多人进

VIII

行疗养的疗养院主任，也像赞成或者反对速度限制的全德汽车俱乐部(ADAC)的主席。自从有了数据，也就有了数据的人为操纵。

对于这些或多或少无耻的人为操纵行为，我在后面的内容中不会给出它们之间的差异，而是对这些统计学意义上所存在的数据伪造者采取一视同仁的态度。当然我也不会认为，一个使人产生错觉的洗涤剂广告应该像由人为操纵而产生癌症风险的行为一样，要受到人们的谴责。从一个较高的层次来看，它们之间肯定存在着差异。我想在这里主要是方法上的差异，而不是原因或者目的上的差异。

如果仅仅是为了弄清楚这些不同的方法之间的差异，那么人们不需要专门学习统计学。对此人们可以完全放心，谁还能依稀记得中学或者大学里那些相关课程的内容呢？对于识别日常生活中大大小小的数据骗子，这些知识已经足够了，人们根本不必来阅读和理解本书。最后，如果一个人为了倾听并辨别另一个人的歌唱得是否正确，他也没必要一定是意大利著名男高音卡鲁索(Enrico Caruso)才行。

人们已经发现许多人为操纵统计学行为的基础形式，并进行了详细的分析。对于我来说，灵感的一个永恒源泉就是达雷尔·赫夫(Darrel Huff)与他那本永不过时的经典著作《统计陷阱》“*How to Lie with Statistics*, New York, 1954”，其德语译本于1956年在瑞士苏黎世出版(德语名字：*Wie lügt man mit Statistik*)。自从《统计陷阱》出版以来，许多统计学家与利用数据骗人的无耻之徒一样，从中学到了不少有用的知识，以至于出版了许多在书名方面相似的书籍，当然本书在内容方面与它们并不重合。托马斯·拉克斯、埃娃·玛莉亚·尼斯华巴(Thomas Lachs, Eva Maria Nesvadba)于1986年在瑞士维也纳出版了一本统计学方面的书：《统计学——谎言或者真实》，这本

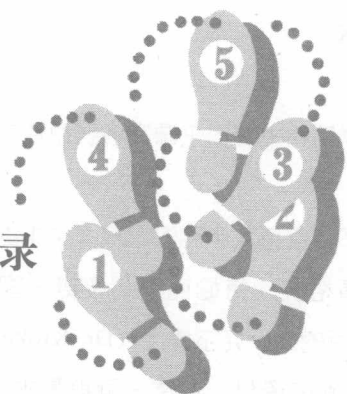
IX

德语书也具有相同的思想。其他的文献与相关资料来源我会在每章的章末尽可能详细地列出。

我要感谢的人很多，特别是我在多特蒙德大学和汉诺威大学中的那些卓越的学生代表。另外也包括我的同事格茨·特伦克勒和彼得·富劳默尔（Peter Pflaumer）。科隆德国经济研究所的克罗克尔博士（Dr. Kroker）非常友好地给我提供了有关星期六工作方面的资料。海德·阿斯豪夫（Heide Aßhoff）和艾尔克·格布哈特（Elke Gebhardt）给予我在画图 and 手稿编辑方面的支持。在核实资料来源和引文方面，克里曼斯·蒂尔克（Clemens Tilke）和奥拉弗·科恩（Olaf Korn）与我一起开展了有意义的合作。拉尔夫·容德（Ralf Runde）审读了手稿，以保证我在各种文体方面不致出问题。Tilmann Deutler、Monika Eberstaller、Michael Eberstaller、Hermann Göppl、Doris Krämer、Sonja Michels、Axel Scheffner、Johannes Schornstein、Claudia Schütze、Michael Schmidt等提出了很多不同的想法与建议。最后还要感谢坎普斯出版社（Campus）的贝内迪克特·布卡德（Benedikt Burkard），他为最终的编辑工作付出了很多心血，并对手稿做了多处修改，使本书具有了更高的可读性。在此我感谢所有人。我还要特别说明的是，如果您在阅读本书时发现了遗留下来的错误，特别是争论激烈的价值判断和成见，那么，所有问题都文责自负。

瓦尔特·克莱默

目 录



关于本书

前言

第 1 章

准确度的假象/1

第 2 章

错觉的误导/14

第 3 章

小心曲线/23

第 4 章

自夸的百分数/36

第 5 章

受操纵的平均值/46

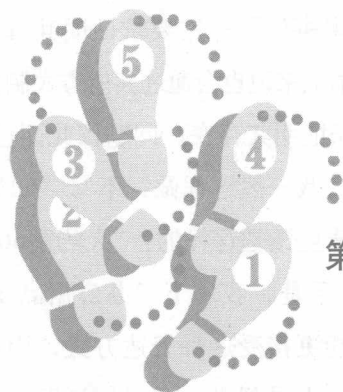
第 6 章

骗人的虚假趋势/58

第 7 章

人工合成的最高级/71

- 第  章 先入为主的抽样检验/80
- 第  章 美化后的图标/95
- 第  章 投之以木瓜，报之以琼琚/104
- 第  章 统计失业抑或实际失业/120
- 第  章 贫穷和富裕：所有都是平等的/128
- 第  章 所有关于图的计谋/139
- 第  章 相关关系PK因果关系/147
- 第  章 飞机噪音产生艾滋病/159
- 第  章 国民生产总值神话/169
- 第  章 还有哪些数字可以相信/178
- 译者后记/184



第 1 章

准确度的假象

缺乏数学修养不仅是因为对数学的一无所知，而且也是因为出于对数字计算的过度依赖。

——卡尔·弗里德里希·高斯

不久以前，我的纳税申报单摆在了我的面前。这种情形对于大家来说早已经是司空见惯。当初为了购买各种各样的东西而发生了这样和那样的大量开支，可时至今日，发票哪去了呢？可能已经丢失了或者不知道被自己搁到了什么地方。如果人们仅仅是因此而必须多缴税费的话，那一定是非常糟糕的事情。

但是，且慢！财政局不也是接受了那些虽未附上发票但已经发生

了的费用吗？按照联邦税务法院的判决标准，可能会出现下面的情形：如果“小概率事件”一定要求证明其真实性，那么产生这种真实性的可能性从根本上来说也会通过其他方式制造出来，即使是通过最简单的方式——收据来证明其存在。问题仅此而已。

事实上，我一般这样做：不写“60马克”，因为这样写会让人感觉不可信，事情应该是这样的：“大约是60马克，但准确的数字我已经记不清楚了”，于是，我写下“办公用品支出：58.24马克”。一般来说，专业人员可能更接受这种表达方式，而不是“大约60马克”之类的说法；如果专业人员想进一步了解证明文件（如收据）的真实性，那么这种风险（编造数据的风险）在此也不会太大。换言之，我是在编造数据而已。

1909年4月6日，美国人罗伯特·埃迪温·皮尔里（Robert E. Peary）到达地球的北极，从而成为世界上第一个到达北极的人。至少皮尔里本人是相信这一点的。说得确切一些，他希望全世界的人都相信他所取得的这个结果，所以，他确定了他到达北极的位置是北纬89度57分11秒，距离北极点大约5公里——这些数据对于皮尔里的实际目标来说已经是一个巨大的成功了，并且准确地说，这种效果也是他精心策划出来的。

实际上，皮尔里根本不可能把他自己的位置精确到30米的距离（因为11秒不会发生任何事情）。即使在今天，借助于卫星定位这种现代化的辅助手段，也不可能取得如此微小的、精确的结果。所以，这件事情在那个时候根本就是不可能的。就连皮尔里的朋友也承认，在

最理想的情况下，皮尔里通过其有限的工具也只能准确地标注到6分（纬度）或10公里，其他指标干脆就是他自己杜撰的。

我曾经在某个媒体读到下面的信息：世界上有8 523 012人（请注意，这里的人数精确到12人！）把匈牙利语作为母语。从常识上看，这个统计结果肯定是错误的，除非是上帝这个全能的造物主创造了这么多的人！大概是一个嗜好数字的“恶人”发明了全部的指标，但不包括像皮尔里一类的第一批到达北极的人。

尽管如此，我们还是相信他们所给出的各种指标。当我们在陌生的地方问一个人：“从这里到最近的邮局有多远？”其中一人回答：“嗯，向前走3公里左右的样子，然后左拐。”而另一个人则回答：“直行2.4公里，然后右拐。”那么，我们到底相信谁？在需要抉择的十字路口，我们是向左走还是向右走？

我相信，绝大多数人会向右走。我们之所以向右走是出于一个相同的原因，这就是为什么我们相信一个探险北极的研究者，他能够知道自己的精确位置，直到秒的程度。或者我们相信一个语言学者，他告诉我们每一个人，世界上哪些人在说这种或者那种语言。常言道“人靠衣装，佛靠金装”，同样也可以说数字装扮指标。如果指标拥有的数字越多，那么我们就越相信这些指标。

1.1 玛士撒拉^①有多大岁数

与一个整数相比，非整数更容易得到人们的肯定。通过这种非逻

① 玛士撒拉（Methusalem），《圣经》中的人物。——译者注

辑的思考方式，一个虚假的非整数数字便得到了人们先入为主的信任。而这一切产生自日常人们所说的“整数（约数）通常都是错的”，或是对现实的不正确反映（并非是真的映像），也就是说，一个准确无误的数字总是不真实的。例如，在我已往的生活中，电话账单还从未出现过精确到100马克的现象。我不能准确地回忆起与自己的妻子或朋友吃饭时花了多少钱，究竟是否恰好200马克或300马克，类似这些整齐的数字在实际生活中从未出现过。当然也有一些例外现象。

NETZGEREI SACK			
PFINZTALSTR. 13			
	A00	U04	5628
	03,95	e12403	816
K	DM/Kg	DM	
FLEISCHWURST			
	24,7% FETTGEHALT		
0,202	18,80	3,80	
TH. ROTWURST			
	14% FETTGEHALT		
0,182	23,80	4,33	
ZIGEUNER-KASSELER			
0,104	31,80	3,31	
MORTADELLA BOLOGNA			
0,091	22,90	2,08	
HAUSMACHER LEBERW.			
0,187	17,80	3,33	
PFEFFERBRATEN			
0,096	32,80	3,15	
6 **		20.00	
N.181			
ES BEDIENTE SIE:			
FRAU LÖFFLER			

这样的账单非常罕见：一个大的整数数字来自众多小的、非整数数字。

因此，我们就会本能地得出一个相反的结论，即每一个非真实的数字必须是无可指摘的（即经过加工处理过了的，从形式上看没有问

题，也指真实的数字都不是整数)。当然这种认识是错误的。超级市场中商品的价格就属于非整数，其价格几乎都是非整数（在德国等发达国家几乎无一例外）。例如，超级市场在销售果酱时，不是标价3马克，而是标价2.97马克。这种情形与下面的事情是一个道理，载重汽车方面的专家在法庭上提供证据，指出某辆汽车的制动距离是63.59米。或者就像一个钓鱼者，他向我们描述最近一次钓鱼的成绩。他从来不会吹嘘自己钓了两打金枪鱼，他会说23条，并且还会提及有一条金枪鱼从自己的鱼钩上逃脱。

这种在数字上所使用的花招就像数字本身一样古老。《圣经》是这样描述的：亚当活了930岁，他的儿子塞特活了912岁，而有名的玛士撒拉甚至活了969岁。天才的作者说不定会这样认为：所有这些男人都是非同寻常的长寿。然而，这个作者已经完全清楚地看到，一个认真严肃的专业书籍作者是不应该这样写的，原因是人们能够完全相信专业作者所做出的结论，即使这个结论是没有经过深思熟虑的研究而草率做出的判断。与此相反，969岁的玛士撒拉这一表述，对于每一位读者来说已经清楚地表明：该书作者对细节部分的资料也是运用自如。

因此，歌利亚（Goliath，被戴维杀死的巨人）也不是“非常大”或者“巨大”，而是精确到六肘零一虎口（肘的算法有几种，最短的一种是0.5米为1肘，6肘便等于3米。六肘零一虎口，即3米多）。还有那顶神圣的帐篷，以色列人在他们行军过程中就是携带着这顶帐篷通过沙漠的，它不仅具有令人吃惊的长度和宽度，而且在尺寸上还精确到