

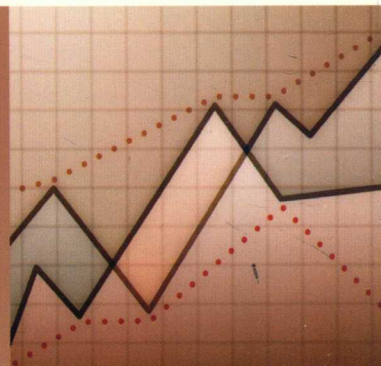
丛书主编：赵焕光



文化数学欣赏丛书 - 4

真理相遇统计

赵焕光 章勤琼 王 迪 著



科学出版社

丛书主编：赵焕光

文化数学欣赏丛书 -4

真理相遇统计

赵焕光 章勤琼 王 迪 著

科学出版社

北 京

内 容 简 介

本书从统计学科的特色、人文欣赏的视野着手,运用通俗生动的语言、精彩有趣的故事、丰富典型的案例,介绍统计文化的常识及统计在现实世界中的广泛应用.主要内容包括为何学统计、统计应用概说、统计陷阱概说、统计学与相关学科的关系、统计历史人物故事精选、统计数据概说、统计数据的收集、统计数据的组织、统计数据的概括、统计指数概说、时间序列概说、随机抽样与抽样分布、参数估计、假设检验、 χ^2 检验与方差分析及其相关分析与回归分析.

本书可作为高等院校所有专业的本(专)科生、硕士生、中学优生、中学数学教师,具有一定数学与统计基础的高校教师以及各行各业的行政管理人士的数学与统计文化修养提高读本,也可作为高等院校本(专)科各个专业的选修课教材或教学参考书.

图书在版编目(CIP)数据

真理相遇统计/赵焕光等著. —北京:科学出版社, 2015.3
(文化数学欣赏丛书)

ISBN 978-7-03-043506-4

I. ①真… II. ①赵… III. ①统计学-普及读物 IV. ①C8-49

中国版本图书馆 CIP 数据核字(2015)第 039300 号

责任编辑:王丽平/责任校对:张怡君
责任印制:张 倩/封面设计:王 浩

科学出版社 出版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

三河市骏杰印刷有限公司印刷

科学出版社发行 各地新华书店经销

2015年3月第 一 版 开本:720×1000 1/16

2015年3月第一次印刷 印张:17 1/4

字数:350 000

定价:69.00 元

(如有印装质量问题,我社负责调换)

序

统计支配世界

统计是什么？简言之，统计是一门关于数据、特别是大数据的科学，统计也是如何从数据中获取信息的艺术。

不论你是否知道，也不论你喜欢与否，统计或者说统计现象，已然从出生到死亡支配着你的一生。众所周知，创造一个新生命只需要一个精子与卵子，但却需要数以百万计的精子去博取一次成功的机会。世界卫生组织研究表明，每毫升中最少含 2000 万的精子数是成功受孕必须具备的条件。自然地，这里还有其他的统计因素。因此，生命的创造是一个统计现象，确切地说，归功于统计或大数据的威力。同时，你也必须清楚地认识到形成生命的过程中，任何人几乎有同等的机会成为男性或女性。

在你接下来的人生旅途中，当你准备进入高校学习时，当然你能就读怎样的大学也受统计结果以及你生活在何处的影响，因为入大学不同批次的分数线并不仅仅是由你的努力程度与天资所决定的。此外，在你的生命里，你会遇到谁、会与谁结婚也充满了随机性。到了你的退休时间，保险的支出费用和所获福利也都是由统计确定的。

没错，统计的确相当重要，许多现象都可以由它来解释。不过，统计只有非常短暂的历史，尽管它与数学密切相关，但它更像艺术。统计需要跟数据打交道，但关于数据却包含多个不同内容。譬如，其中一个重要内容是如何有效地收集数据；而一旦数据收集完毕，如何整理数据并清楚地展示出来也非常关键；此外，统计最重要的作用还体现在如何由数据推断结论。

在信息时代，获取数据也许并不十分困难，重要的是如何获取好的数据。整理和解释数据也颇具难度。人们需要深刻的洞察力在众多的变数和选择之中作出正确的决定并完成正确的事情。当然，人们需要数学（或概率）理论来实施这些步骤，但概率论与统计学不同。下面一个例子可以解释两者的差异：假设一个容器里装有一些红球与黑球，如果已经知道容器里面两种球的组成情况，然后问随机取出一个球

是红球或黑球的可能性是多少,这是概率.相反地,如果你已经取出了一些球,而且已经知道取出球的颜色,现在想要知道容器里面红球跟黑球的组成,这是统计,或者更确切地说是推断统计——利用样本数据中的模式得到对该样本代表的总体的推断.

获得隐藏在大数据背后的关键模式是统计的重要内容,但这不是统计的全部.如何从大总体中取得真正随机的样本,如何从小总体中取得可靠的有用的数据,这些都需要原创性的新想法.例如,假设一所大学发现有学生在一次重要的考试中作弊,并试图找出到底有多少学生作弊.这所大学可以怎么做?如果他们随便问一个学生他是否作弊,有很大可能他们并不能得到直接的答案.事实上,即使他们去问再多的学生也得不到更多真实有用的信息.但是,他们可以利用统计的思想和方法设计一些恰当的问题去得到想要的结果.诸如此类的许多问题都能在由赵焕光教授等著的有趣且非同寻常的有关统计的著作《真理相遇统计》中得到答案.

西方许多人都听说过盖洛普民意测验.而在美国,每当总统选举临近时,人们对这个名词更是耳熟能详.但是它是如何进行的?一个女政治家和她的女婿对这个测验曾经产生了怎样的影响?该书给出了详细解释.

该书中出现的另一个重要人物是著名数学家哈代 (Godfrey Hardy, 1877~1947).哈代在数学上有着非凡的成就和荣耀,但他却度过了一个并不幸福的晚年.在他的名著《一个数学家的自白》中提到:不论他在数学中做了什么,对于日常生活都是无用且无害的^①.但事实并非如此,该书就陈述了哈代聪明地将概率应用于色盲疾病的研究.

这里我想提及一位当代伟大的纯数学家对统计的认识的转变. David Mumford (1937~) 的工作领域是代数几何,获得了菲尔兹奖、沃尔夫奖以及众多其他荣誉.在由美国数学学会出版的著作《数学:前沿与观点》中,他有一篇题为《统计时代的黎明》的文章,其中提到:“作为一名随机方法的再生信徒,我在这里必须坦承说明这种转向的实情.上周, Dave Wright 提醒我, 70 年代在我热衷于代数几何的日子里,曾给过一个研究生这样的忠告:‘天哪,不要把你的时间浪费在统计上——那不过像烹饪手册一样毫无意义!’现在我收回这句话!”^② 他进而解释道:“两千多年以来,亚里士多德 (Aristotle, 公元前 384~ 前 322) 的逻辑一直统治着西方的知识分子的思维方式.一切精确的理论、科学的方法,甚至连思考本身的过程的模型,都必须在原理上遵循逻辑的规范.然而,从其并不光明的发端——设计赌博策略

① Hardy G H. A Mathematician's Apology. With a foreword by C. P. Snow. Reprint of the 1967 edition. Cambridge: Cambridge University Press, 1992: 153.

② Arnold V, Atiyah M, Lax P, Mazur B. Mathematics: frontiers and perspectives. American Mathematical Society, Providence, RI, 2000: xii+459. (在庆祝 2000 年为数学年时,此书作为国际数学联合会的代表出版).

以及中世纪伦敦清点尸体——开始, 概率理论以及统计推断已经成为科学模型的更好的基础了. 尤其是其思维的过程, 以及可以将其作为理论数学的核心成分甚至数学本身的基础. 从我们的观点来看, 可以预计这种突变将在接下来的这个世纪最终影响整个数学.” 他的新建议对每一位数学学生应该都有价值, 而且统计的影响将不仅仅局限于科学领域.^①

随着计算机的普及及其强大的计算能力, 获取和储存数据变得越发简单, 而统计也随之走进许多学科的前沿领域. 如今市场上已有众多关于统计的书籍, 每年的新书更是层出不穷. 该书作者花费了大量的时间和精力去系统搜集相关而有趣的例子. 你将会为统计结论和理论背后隐藏着如此众多的精彩故事感到意外, 而又惊奇于生活中众多令你困惑却又十分重要的事情都能够通过统计来解决或解释. 人们一直认为统计是干枯冰冷的, 但该书给你展示了它真实且有趣的另一面. 同时, 该书又不像一些统计的通俗读本, 它还以一种可以接受且扎实的方式帮助你理解一些基本概念与理论.

我们都知道基本的统计概念, 如平均与平均数. 但还有哪些其他的基本概念与方法呢? 我们经常讨论两样看起来类似的事情是如何联系的, 但两者之间具体的相关关系如何? 该如何量化并测量它们? 如何理解两者之间的关系? 所有这些疑惑都在本书中有所提及并给出了解释.

该书不是一本标准的统计教科书, 也不是一本有关统计的通俗读本, 它更像是两者的结合. 该书通过故事与生活中的实际事例解释了许多有趣的统计理论与结果. 它让你感受并目睹到统计并不仅仅是处理枯燥的数据, 而是向你展示了这些数字是如何产生, 该如何理解.

毛泽东曾说过: “枪杆子里面出政权.” 也有许多人说笔杆子比枪杆子更厉害, 因为文字与书籍能持续更长时间且更具影响力. 的确, 许多国王和王朝已经消失在历史的尘埃之中, 也早已从人们的记忆中淡去, 但是圣经中的文字却已流传了数千年并将永存下去. 当然如此, 圣经是宗教真理, 而真理永存. 正如该书作者在前言中所解释的, 统计就是一种在各种复杂情形下接近真理的有效工具. 在当下这个信息时代以及接下来的新时代中, 如何理解大数据并正确而有效地运用它们将是成功之匙. 不论谁如果做到了这一点, 就可以支配世界!

统计也可能是危险的并有可能被滥用. 许多谎言和虚假信息在统计的伪装下迅速地传播, 而识破它们的最有效途径正是正确地理解和运用统计知识. 获取对统计的正确理解并加以应用是世界上每个公民的权利与责任. 该书可以看成是一个良

^① 事实上在我国的经典著作《易经》中, 就已经可能运用概率来设计“筮占”, 这一点并不为很多人所知. 详见 Diaconis P, Graham R. *Magical Mathematics: The Mathematical Ideas That Animate Great Magic Tricks*. Princeton: Princeton University Press, 2011. 如果书中这一描述属实, 那么这可能是人类历史上对概率理论的最早应用.

好的起点。

该书作者在前言中表达了一个美好愿望，希望有读者能从书中获得对统计的美好感觉并爱上统计。我确信这一愿望能够成真。

季理真

密歇根大学，美国密歇根，48109

前言

印度著名统计学家 C.R. 劳 (C.R. Rao, 1920~) 在其名著《统计与真理》中说:“人类一切努力的最终目的是寻求真理,而在严格意义下的真理是不可能得到的,替代的是要寻求可接受的知识.严格地讲,知识不是真理,但它应最好地运用真理.”他还说:“统计学是探求真理必不可少的工具.”拙著取名“真理相遇统计”就是从这里获得启迪的.

如何探求真理?从古至今全世界都没有找到明确的答案.古希腊哲学家赞诺芬 (Xenophanes, 约公元前 434~前 355) 有诗云:

真理未知亦难知,

上帝人间布迷离,

恰好诸事我所提,

偶尔逢机出奇迹,

永恒真理非彼知,

茫茫谜网尽猜疑.

什么是真理?学术界一直争论不休.法国著名哲学家、数学家笛卡儿 (Rene Descartes, 1596~1650) 曾经这样表述:“当我们不具备决定什么是真理的力量时,我们应遵从什么是最可能的,这是千真万确的真理.”很清楚,真理隶属于哲学的范畴,通常百姓很少去思考关涉真理的大问题.显然,真理奠基在真相的基础之上,从本质上说,它们之间没有根本的区别.然而与真理相比,真相却与我们的日常生活接近许多.例如,对于揭露发生在 2014 年的 MH370 失联真相以及 MH17 坠毁真相,许多人都会很有兴趣去关注.当然,真相有时候会被各种各样错综复杂的信息所掩盖,欲在混乱不堪的无序现实世界中把真相揭示出来,光凭相互争雄的哲学思辨方式是难以实现的.我国著名概率统计学家严加安院士说:“无序隐有序,统计解迷离.”现代统计通过对客观事物 (现象) 的观察、收集并筛选相关的信息 (数据),去除那些扭曲事实的并澄清混淆关系,使得让事物的真相原形毕露成为可能. C.R. 劳说,“在终极的分析中,一切知识都是历史;在抽象的意义下,一切科学都是数学;在理性的基础上,所有的判断都是统计学”.由此可见,统计无疑成为帮助人们披露真相不可或缺的重要工具.

20 世纪初,被称之“科幻小说之父”的作家、思想家威尔斯 (Herbert George

Wells, 1866~1946) 曾经预言“在未来社会统计学思维将像阅读能力一样成为社会人必不可少的能力”。的确, 统计思维作为一种认知世界的思维方式, 它与哲学、数学一样具有同等重要的基础性地位。其实, 统计更重要的意义应该体现在统计文化与统计观念层面。美国哈佛大学教授哈克英 (William E. Hocking) 说: “那些默默无闻的统计学家已经改变了我们的世界, 不是发明了新的事实或技术, 而是改变了我们推理和试验的方法, 以及我们对这个世界的观念的形成方式。”从他的话中我们可以看出, 统计观念对现代文明的贡献是非常巨大的。特别地, 现代文明从追寻绝对真理到探求相对真理的华丽转身, 正是从现代统计观念那里获得灵感源泉的。

目前, 国际统计教育界关于统计教育目标比较流行的提法是: 尊重统计文化, 形成统计观念, 提升统计素养。我们认为, 所谓统计素养可理解为养成尊重事实、用数据说话的习惯, 养成能够自觉地从统计角度思考现实问题的习惯。关于统计观念, 有不少统计学者作过深入探讨。英国的 Peter Holmes 先生在 1980 年提出统计观念具体表现为下述五个方面, 值得我们借鉴:

- (1) 数据的收集;
- (2) 数据的记录与表示;
- (3) 数据的提炼;
- (4) 数据与概率的关系;
- (5) 对数据进行解释并作出推断。

20 世纪末有一首流行歌曲中唱到: “……借我一双慧眼, 让我把这纷扰看得清清楚楚, 明明白白, 真真切切……”实际上, 形成统计观念、提升统计文化素养, 就是让统计借给我们一双慧眼, 让统计帮助我们打开瞭望世界的窗口、带领我们走进生活更加广阔的天地、指引我们走向人生更加美好的未来。

全书分三章。第 1 章标题为“统计启蒙”, 主要内容包括为何学统计、统计应用概说、统计陷阱概说、统计学与相关学科的关系及统计历史人物故事精选。这一章的主要亮点是通过讲故事的方式阐明所探讨的主题, 运用通俗易懂的生活语言行文, 尽可能吸引读者的阅读兴趣。第 2 章标题为“统计描述”, 主要内容包括统计数据概说、统计数据的收集、统计数据的组织、统计数据的概括、统计指数概说及时间序列概说。这一章的主要亮点是把老百姓普遍关心的经济热点问题与描述统计的基本常识有机地融合在一起, 让读者感受到统计描述并不是那么枯燥无味的。第 3 章的标题是“统计推断”, 主要内容包括随机抽样与抽样分布、参数估计、假设检验、 χ^2 检验与方差分析, 以及相关分析与回归分析。这一章的主要亮点是充分运用实际生活中有趣的典型案例, 以深入浅出的表述方式介绍统计推断的常识及其内容丰富的实际应用, 揭开统计推断的神秘面纱, 让读者感受到统计推断并不那么深奥而且魅力无限。

本书第一作者长期从事大学数学分析类课程的教学与研究工作, 并非统计学

专业教育人士. 可以说, 仅凭个人对统计文化的兴趣与爱好, 认真学习与欣赏了统计文化最基础的内容. 然后, 再将自己平时陆续积累的少许学习体会整理成小册子, 权作文化数学欣赏系列丛书之一正式出版. 我们写作本书的目的很明确, 就是借此机会努力做一点力所能及的统计文化传播工作, 同时为从事数学教育工作的朋友提供一些拓宽数学视野的阅读材料 (跳出数学看数学). 如果有哪一位读者朋友, 通过阅读本书, 不再觉得统计理论离我们的日常生活很遥远, 那么我们会感到很欣慰; 如果有哪一位读者朋友能从本书中找到爱上统计的感觉, 那么我们会更加心满意足了!

本书写作的思路与框架是由我与章勤琼博士、王迪博士共同讨论形成的, 初稿由我提供, 章勤琼与王迪参与修改及定稿工作. 本书在写作过程中参阅了大量文献, 为此向被本书引用的参考文献的作者表示特别的感谢. 吾妻钱亦青在书稿打印及文献查阅方面付出巨大的努力; 本书在修改与出版的过程中得到温州大学数学学院高利新教授、洪振杰教授、王玮明教授、方均斌教授、应裕林副教授、黄忠裕副教授、程国正副教授、张笑钦副教授、黎祥君副教授、蔡风景副教授等诸多同仁的大力帮助; 本书出版得到浙江省重点学科“应用数学”、温州大学重点学科“数学”、温州大学优势专业“数学与应用数学”、温州大学教务处、温州大学科学技术处、温州大学人文社会科学处相关建设项目资助, 在此一并表示感谢! 美籍华人数学家季理真教授为本书写了精彩的序, 使本书增色不少, 特此表示诚挚的感谢!

限于作者水平, 不足之处在所难免, 恳请读者批评指正.

赵焕光
2013 年 10 月

1.3 统计陷阱概说	25
1.3.1 李普杰悖论式陷阱	25
1.3.2 平均数陷阱	26
1.3.3 百分数陷阱	29
1.3.4 信息不充分陷阱	30
1.3.5 样本有偏陷阱	31
1.3.6 伪造数据式陷阱	32
1.3.7 统计滥用式陷阱	33
1.4 统计学与相关学科的关系	35
1.4.1 统计学与数学的关系	35
1.4.2 统计学与相关实质性学科的关系	36
1.5 统计历史人物故事精选	37
1.5.1 格兰特和哈雷的故事	37
1.5.2 高斯与凯特勒的故事	38
1.5.3 高尔顿与卡尔·皮尔逊的故事	37
1.5.4 戈塞特与费希尔的故事	31

目 录

第 1 章	统计启蒙	1
1.1	为何学统计	1
1.1.1	从两个典型案例看统计	1
1.1.2	“统计”的含义解释	7
1.2	统计应用概说	9
1.2.1	统计在日常生活中的应用	10
1.2.2	统计在法律中的应用	11
1.2.3	统计在军事上的应用	14
1.2.4	统计在考古学中的应用	15
1.2.5	统计在遗传学中的应用	17
1.2.6	统计在流行病学中的应用	19
1.2.7	统计在经济管理中的应用	21
1.3	统计陷阱概说	25
1.3.1	辛普森悖论式陷阱	26
1.3.2	平均数陷阱	28
1.3.3	百分数陷阱	29
1.3.4	信息不充分陷阱	30
1.3.5	样本有偏陷阱	31
1.3.6	伪造数据式陷阱	32
1.3.7	统计滥用式陷阱	33
1.4	统计学与相关学科的关系	35
1.4.1	统计学与数学的关系	35
1.4.2	统计学与相关实质性学科的关系	36
1.5	统计历史人物故事精选	37
1.5.1	格兰特和哈雷的故事	37
1.5.2	高斯与凯特勒的故事	43
1.5.3	高尔顿与卡尔·皮尔逊的故事	47
1.5.4	戈塞特与费希尔的故事	51

第 2 章 统计描述	54
2.1 统计数据概说	54
2.1.1 统计数据简述	54
2.1.2 统计规律简述	55
2.1.3 大数据简述	56
2.1.4 数据挖掘简述	57
2.2 统计数据的收集	59
2.2.1 统计调查概说	59
2.2.2 统计调查问卷设计	62
2.2.3 盖洛普民意测验	65
2.2.4 随机对照实验	67
2.3 统计数据的组织	72
2.3.1 统计数据分组	72
2.3.2 分布数列概说	76
2.3.3 统计图简介	80
2.4 统计数据的概括	87
2.4.1 总量概括与常用经济总量指标	87
2.4.2 相对比较与常用相对经济指标	92
2.4.3 数据集中趋势的概括	101
2.4.4 数据离散程度的概括	113
2.5 统计指数概说	116
2.5.1 统计指数概述	117
2.5.2 综合指数与平均指数	119
2.5.3 居民消费价格指数	125
2.5.4 股票价格指数 (SPI)	130
2.6 时间序列概说	133
2.6.1 时间序列概述	133
2.6.2 时间序列的水平指标	135
2.6.3 时间序列的速度指标	141
2.6.4 时间序列指标运用	145
2.6.5 经济增长率与人口自然增长率	146
第 3 章 统计推断	151
3.1 随机抽样与抽样分布	151
3.1.1 统计学中的常用概念	152
3.1.2 随机抽样概述	154

3.1.3	随机抽样误差	161
3.1.4	抽样推断的数学基础	165
3.1.5	抽样分布概说	176
3.2	参数估计	180
3.2.1	参数估计典型案例	180
3.2.2	参数的点估计	185
3.2.3	参数的区间估计	192
3.2.4	样本容量的确定	197
3.3	假设检验	199
3.3.1	假设检验概述	199
3.3.2	总体均值的检验	205
3.3.3	总体比例的假设检验	212
3.3.4	假设检验与区间估计的关系	215
3.3.5	p -值检验	217
3.4	χ^2 检验与方差分析	219
3.4.1	χ^2 检验	219
3.4.2	方差分析	227
3.4.3	方差估计	232
3.5	相关分析与回归分析	233
3.5.1	相关分析概述	234
3.5.2	相关系数的测定与检验	236
3.5.3	统计相关误用陷阱	240
3.5.4	回归分析概说	241
3.5.5	回归方程的统计检验	247
3.5.6	回归方程的应用	253
3.5.7	简单非线性回归模型	256
	参考文献	260

第 1 章

统计启蒙

本章作为全书的开篇语, 伴随统计启蒙的梦想, 试图回答下述问题.

- (1) 为什么学习统计?
- (2) 统计有哪些具体应用?
- (3) 统计应用有哪些常见陷阱?
- (4) 统计与数学有什么关系?
- (5) 统计与相关学科有什么关系?
- (6) 有哪些伟大统计学家贡献出重要的统计思想方法?



1.1 为何学统计

学习统计的理由有千百条, 每个人都有各自的独特答案. 学生学习统计的首要理由是要取得好的课程成绩, 教师学习统计的理由是不断提高自己的教学水平, 统计研究人员学习统计的理由是为了努力提高自己的研究水平, 管理人员学习统计的理由是更好地提高管理服务水平, 我们作为统计文化普及的一员, 学习统计的理由是力求写出通俗易懂的统计科普文化著作, 为启蒙普通百姓的统计意识更好地尽自己的微薄之力…….

1.1.1 从两个典型案例看统计

例 1.1.1(敏感问题调查) 大家知道, 统计调查中调查者往往要通过问卷或口头问答的方式从被调查者处获取某些数据, 这里有些问题的数据是易于得到的. 例如, 人们的年龄、性别、职业、出生地等, 一般情况下可以期望回答是如实准确的. 但对某些敏感的社会问题则不然, 例如:

问一个成年人: 您吸毒吗? 您是同性恋者吗?

问一个商人: 您曾偷税漏税吗?

问一个学生: 您考试时作过弊吗?

问一个职员：您觉得您的直接上司称职吗？

显而易见，很难指望被调查者会坦率如实地回答这些问题，这时直接调查得到的资料将是极其不可靠的，通过这种资料作出的判断将导致错误的结论。然而鉴于此类问题的重要性，我们必须获得如实回答的数据。很幸运，利用简单概率论的知识，统计学家已经帮助人们找到了此类问题的处理方式。下面通过一个具体问题的调查说明其原理。

校方想了解该校学生在某次考试中作弊的情况，通常做法是选定一个或几个有代表性的班级进行问卷调查。问卷设计包含两个问题：

第一个问题是希望得到真实回答的敏感问题，此处就是“您在考试中作过弊吗？”；

第二个问题是一个普通问题，任何人都会回避给出真实答案，例如，“您的学号是偶数吗？”或者“您喜欢读武侠小说吗？”。

明确告诉被调查者，每人只需回答其中一个问题，至于具体为哪一个，则由他们自己抛掷一枚硬币来决定，当硬币上的国徽向上时回答第一个问题，否则回答第二个问题；不论回答哪一个问题，都只选择一个字的答案，即“是”或“否”，而且不用注明回答的是第一个问题还是第二个问题。这样的问卷设计使被调查者完全打消了顾虑，因为即使对一份回答“是”的问卷也无法证实回答者承认在考试中作弊，这一答案完全可以针对第二个问题的。但是这样的问卷对调查目的而言已经足够了，从得到的数据中我们已经可以推断作弊学生的比例，其方法如下。

假设被调查对象总数为 N ，回答“是”的人数为 N_1 。又假设考试中作弊学生的比例为 λ ，也就是当一个被调查者选定回答第一个问题时，回答“是”的概率是 λ 。将被调查者选定回答第二个问题时回答“是”的概率记为 β 。请注意可以认为 β 是已知的，对此仍以上面的例子来说明。如果问卷中的第二个问题是“您的学号是偶数吗？”，显然，当被调查的学生人数足够多时，回答“是”的概率，即 β ，应该很接近 $1/2$ ；如果第二个问题是“您喜欢读武侠小说吗？”，事情会复杂一些，但这是一个普通非敏感问题，可以通过另一次独立调查来解决。

实际上，统计学家已经设计出了很有效的问卷方式，将所需要的独立调查合并在上述敏感问题调查之中，对此不再详述，但无论如何，都可以认为 β 是一个已知数。由上面所规定的回答第一个问题还是第二个问题的选择方式，可以知道每个问题各有一半学生作答，由此上述各个量之间有以下关系：

$$\lambda \frac{N}{2} + \beta \frac{N}{2} = N_1.$$

从中不难得到

$$\lambda = \frac{2N_1}{N} - \beta.$$

上述方法可用于各种各样的敏感问题调查，这里作为引例，仅提供一个简单介

绍, 针对更一般的情况, 统计学家已经设计了多种更为完善的方案. 关于这一方面, 想了解更多的读者可到更深入的统计学著作中寻找答案.

例 1.1.2(色盲遗传问题) 色盲虽然不是什么严重疾病, 但却也是一种生理缺陷. 大约在 20 世纪初, 有人发现色盲是可以遗传的. 于是人们提出了一个令人担忧的问题:

色盲既然能遗传给下一代, 那么将来会不会有一天使全世界的人都成为色盲呢?

如果真是这样, 那么这个世界真是太可怕了! 要解决这个问题, 首先要弄清楚色盲是怎么回事. 为此, 先得弄明白人们为什么能看到颜色, 这就要研究视网膜的复杂构造及其性质, 还得了解不同的光波所能引起的光化学反应等. 因为眼睛是人体很复杂的器官, 只从解剖学的角度来考虑, 就已经十分困难了, 何况还与遗传因素有关. 当时, 人们还不了解遗传基因的结构, 根本没法了解色盲与遗传基因方面的关系. 因此, 从生理学上来讲, 当时这是一个无法解决的难题. 生理学家请英国大数学家哈代 (G. H. Hardy, 1877~1947) 帮助解决这个难题. 哈代出手不凡, 他以概率统计的观点, 仅用初等代数知识, 便非常巧妙且彻底地解决了这个难题.

哈代首先从大量临床统计资料中了解到以下三种情况:

- (1) 色盲中男性远多于女性;
- (2) 色盲父亲与正常母亲不会有色盲孩子;
- (3) 正常父亲和色盲母亲的儿子是色盲, 女儿则不是.

据此, 哈代判断色盲的遗传与性别有关. 当时的生理学已经搞清楚男女性别的差异, 与遗传基因中的性染色体有关. 每个人的体内有 23 对染色体, 一半来自父亲, 一半来自母亲; 女性性染色体是 XX. 在遗传给下一代时, 母亲赋予 XX, 给予子女的总是 X, 父亲赋予 XY, 随机地选择一 X 或者 Y 给子女, 其比例是 21:22. 若为前者, 则是女性; 若为后者, 则是男性. 所以男、女出生的比例是 22:21(注: 这里实际上已经回答了统计资料中为什么说男性比例略高于女性的问题).

既然色盲与性别有关, 所以色盲者一定是性染色体出了问题. 那么究竟是 X 出了问题, 还是 Y 出了问题呢? 一定是 X, 而且这个异常染色体会世代遗传下去. 为什么能肯定病态染色体是 X 呢? 这可用反证法证明.

假如病态染色体是 Y, 女性就不会有色盲, 因为女性性染色体中没有 Y. 但是, 女性有色盲存在, 只是比男性色盲少而已.

很自然, 人们必须弄明白为什么男性色盲比女性多. 这是因为女性有两个 X, 如果其中有一个异常、一个正常, 仍然可以维持正常视力. 这种女性, 不妨称其为“次正常”. 这样, 男性分为两类: 正常和色盲; 女性分为三类: 正常、次正常和色盲.

在基本生理常识分析的基础上, 哈代运用非常简单的概率统计方法估计出下一代人中的色盲比例. 他首先根据概率统计中的随机原则作了如下假设:

- (1) 在两类男子和三类女子之间, 夫妇配对的机会是随机的;

(2) 异常染色体 (记作 $\bar{X}$), 在所有染色体 X 中所占比例为 p , 在男、女染色体中保持不变;

(3) 父、母和子女中男女出生比例假设为 1:1.

在上述假设的基础上, 不难推知, 若男性中正常和色盲两类分别以 F 和 S 表示; 女性中正常、次正常和色盲三类分别以 Z, C 和 K 表示, 则 F, S 在男性中所占比例分别为 $q, p(q = 1 - p)$; Z, C, K 在女性中的比例分别为 $q^2, 2qp, p^2$. 易见男、女配对有 6 种夫妇类型. 各种配对类型的夫妇所生子女为色盲的比例可通过下述列表的方式计算:

第一类 (F, Z) 配对, 即丈夫、妻子均为正常. 发生这种类型的概率为 q^3 , 子女中不会有色盲, 如表 1.1.1 所示.

表 1.1.1 正常丈夫与正常妻子

男 \ 女	X	Y
	X	X
X	(X, X) 正常女儿	(X, Y) 正常儿子
X	(X, X) 正常女儿	(X, Y) 正常儿子

第二类 (F, C) 配对, 即丈夫正常, 妻子次正常. 发生这种类型的概率为 $2pq^2$, 其子女的四种情况中有一种是色盲 (表 1.1.2), 即这类夫妇的子女中有 $1/4$ 是色盲, 在下一代人口中所占的比例是 $2pq^2 \times \frac{1}{4} = \frac{1}{2}pq^2$.

表 1.1.2 正常丈夫与次正常妻子

男 \ 女	X	Y
	X	X
$\bar{X}$	($\bar{X}$, X) 次色盲女儿	($\bar{X}$, Y) 色盲儿子
X	(X, X) 正常女儿	(X, Y) 正常儿子

第三类 (F, K) 配对, 即丈夫正常, 妻子色盲. 发生这种类型的概率为 p^2q , 其子女的四种情况中有两种是色盲 (表 1.1.3), 即这类夫妇的子女中有 $1/2$ 是色盲, 在下一代人口中所占比例是 $p^2q/2$.

表 1.1.3 正常丈夫与色盲妻子

男 \ 女	X	Y
	X	X
$\bar{X}$	($\bar{X}$, X) 次色盲女儿	($\bar{X}$, Y) 色盲儿子
$\bar{X}$	($\bar{X}$, X) 次色盲女儿	($\bar{X}$, Y) 色盲儿子

第四类 (S, Z) 配对, 即丈夫色盲, 妻子正常. 发生这种类型的概率为 pq^2 , 其子女不会有色盲 (表 1.1.4).