



“十二五”普通高等教育本科国家级规划教材

第十一届全国统计科学研究优秀成果奖（统计教学类）二等奖



北京高等教育精品教材

BEIJING GAOJING JIAOYU JINGPIN JIAOCAI

教育部经济管理类核心课程教材

Statistics

统计学

——思想、方法与应用（第二版）

▶▶ 袁 卫 刘 超 编著



“十二五”普通高等教育本科国家级规划教材

第十一届全国统计科学研究优秀成果奖（统计教学类）二等奖



北京高等教育精品教材
BEIJING GAODENG JIAOYU JINGPIN JIAOCAI

教育部经济管理类核心课程教材

Statistics

统计学

——思想、方法与应用（第二版）

► 袁 卫 刘 超 编著

中国人民大学出版社
• 北京 •

图书在版编目 (CIP) 数据

统计学: 思想、方法与应用/袁卫等编著. —2 版. —北京: 中国人民大学出版社, 2016. 9
教育部经济管理类核心课程教材
ISBN 978-7-300-20851-0

I. ①统… II. ①袁… III. ①统计学-高等学校-教材 IV. ①C8

中国版本图书馆 CIP 数据核字 (2015) 第 034038 号

“十二五”普通高等教育本科国家级规划教材
第十一届全国统计科学研究优秀成果奖 (统计教学类) 二等奖
北京高等教育精品教材
教育部经济管理类核心课程教材
统计学——思想、方法与应用 (第二版)
袁 卫 刘 超 编著
Tongjixue: Sixiang, Fangfa yu Yingyong

出版发行	中国人民大学出版社	邮政编码	100080
社 址	北京中关村大街 31 号		
电 话	010-62511242 (总编室)	010-62511770 (质管部)	
	010-82501766 (邮购部)	010-62514148 (门市部)	
	010-62515195 (发行公司)	010-62515275 (盗版举报)	
网 址	http://www.crup.com.cn		
	http://www.ttrnet.com (人大教研网)		
经 销	新华书店	版 次	2011 年 9 月第 1 版
印 刷	北京东方圣雅印刷有限公司		2016 年 9 月第 2 版
规 格	185 mm×260 mm 16 开本	印 次	2016 年 9 月第 1 次印刷
印 张	21 插页 1	定 价	39.00 元
字 数	455 000		

版权所有 侵权必究

印装差错 负责调换

第二版前言

《统计学——思想、方法与应用》自 2011 年 9 月出版以来，受到了读者的欢迎，也获得了一些重要的奖励，包括 2012 年获得第十一届全国统计科学研究优秀成果奖（统计教学类）二等奖，2013 年获得“北京高等教育精品教材”称号，并于 2014 年入选“十二五”普通高等教育本科国家级规划教材。

大数据时代已经来临。正如电影《美国队长 2》中的台词所说：21 世纪就是大数据书。现在，大数据的概念已经深入人心。各行各业、不同人群都知道了大数据，知道了统计方法。考虑到这个历史大背景以及广大师生的迫切要求与建议，我们在总结第一版教材的基础上进行了修订，其变化主要体现在以下几个方面：

第一，与第一版相比，本书对全书进行了较大幅度的修订，包括数据的更新、内容的增删等。尤其是在书中增加了很多公式，但仍将烦琐的计算交给统计软件完成，便于读者结合数据理解相关统计方法，突出各种统计方法的介绍和使用。

第二，考虑到教材的广泛使用，在写法上力求简洁清晰，使得内容更加紧凑；更新了许多例子，更多地使用与老百姓息息相关的例子（如淘宝店、PM2.5 等），以期读者更好地理解、掌握统计方法的使用。

第三，对各章的练习题做了适当增删。我们特别感谢广大读者和授课教师，是他们的鼓励和建议帮助我们不断完善教材。感谢中国人民大学出版社的相关编辑，没有他们的支持、鼓励和辛勤劳动，新版教材也不可能这样快与读者见面。

由于时间仓促，书中错误在所难免，敬请读者指正。

袁 卫 刘 超

第一版前言

为了满足当代学生和广大读者对统计方法在信息社会中日益增长而又尚未满足的需求，我们专门为各领域应用统计和数据分析编写了这本教材。

随着统计知识的普及以及经济、社会的发展，特别是在信息化社会的今天，理、工、农、医、经济、管理以及人文社会科学对统计数据分析的需求在持续增加，需要处理大量的数据信息，统计是有力的分析工具和决策方法。

大多数统计应用不是由统计学家实施的，而是实际工作者完成的，教师首先应该了解学生的想法和需要，才能达到教学相长。但是很多统计教师在对实际工作者讲授统计时，总是习惯性地希望教学对象也运用自己学习统计时的思维方式，这是一个误区！

目前，我国统计教学过程中存在两个困难或误区：（1）国内的很多教材过于重视数学公式和推导，使得读者轻视对统计思维的培养，导致面对实际数据时感到恐惧而束手无策；（2）现有很多教材虽然也强调应用，但是忽略统计方法的软件使用，虽然可以通过软件操作手册掌握软件的使用，但这增加了读者学习的负担，而且这些使用手册过于强调操作过程，忽略对软件输出的统计结果的解释，导致读者不能正确解读统计分析结果，面对一大堆输出而束手无策。

本书作为一本统计学入门教程，体现了以下准则：

- 读者通过分析实际数据，解决实际问题，得到最大的启发。
- 读者应该有机会去使用统计软件分析真实的数据。
- 相对于概率论的基础知识的学习，我们更应该重视概率意识的建立。
- 强调统计的应用，而非统计的理论知识。

本书的最大特色就是应用性强，不仅介绍各种统计方法而且注重在实际中的应用。以大量实际数据运用为基础，为读者深入领会统计思想与方法提供了最有效的工具，编排新颖，讲解清晰。每一种统计方法的介绍都有许多例子和案例，并在每章配有大量的例题和练习。这些统计信息不仅能使读者产生兴趣，还可以让读者从

中学习统计方法及其应用。

本书的另一特色就是可读性强,语言轻松活泼,通俗易懂。本书借助于界面越来越方便友好的统计软件,避免了烦琐的数学推导和统计公式,采用深入浅出、循序渐进的语言系统介绍统计学的知识,帮助读者从中得到有用的信息,并且能够很容易地利用 SPSS 和 Excel 完成各种统计分析的运算。其间穿插了大量丰富、鲜活、生动的扩展阅读,使得统计技术不再枯燥,方便了读者的理解和学习。

本书的特色具体体现在以下几个方面:

1. 案例引子。所有的新知识都通过实际数据的例子来阐述。各章安排了“案例引子”,通过实际生活中的各种实例引导读者进行学习。既可以激发读者学习的兴趣,又可以将复杂的事物简单化、具体化,帮助读者更好地理解和学习统计方法与知识,让读者逐步掌握解决实际问题的能力。

2. 学习目标。在每章的“学习目标”部分,指明教学时应该明确的相关问题。帮助读者系统、全面、深入地理解和掌握基本概念、理论和方法。

3. 想一想。每章均安排了“想一想”的小习题,其中的概念和问题涉及的方式都与示例相似,通过这部分练习,读者可以实时测验自己对知识的理解,巩固所学知识。如果你肯花上一点儿时间把书中讨论的话题和你自己的实际经验结合起来,将会更清楚地认识到,你对统计理解了多少,哪里还存在问题,该如何把课堂内的体验和课堂外的生活联系起来,等等。这种行为将引起你对统计更持久的学习。

4. 专栏或扩展阅读。每章均安排了“专栏”或“扩展阅读”,给出一些与统计知识有关的历史故事或注意事项,帮助读者了解所学知识的重要性和实践性。

5. 无公式化教学。计算机和教学的结合,尤其是界面越来越方便友好的统计软件的使用,已经使旧的学习方法——特别是记忆并运用统计公式——不再适用于大部分读者。本书在每章的讲解中尽量不使用统计公式,而是把统计公式放在每章末的附录里。

6. 历史上的统计学家简介(人物小传)。本书简要介绍了一些著名的统计学家以及他们的成就。通过对他们的了解,扩展读者的视野,使读者了解统计学发展历史上重要的人物以及学科发展历史事件,读者会更加全面地了解统计这个学科。

7. 真实的应用案例。每章后面都有应用案例,学生需要分析案例中的数据去解决所提出的问题,方便学生利用本章所学知识解决实际问题。

8. 挑战性的案例分析。每两到三章统计知识结束之后安排一个挑战性的案例分析,读者需要用前面章节中提出的办法去分析这些案例,复习学过的知识。使得读者借此拓展自己的思维,挑战自己运用统计的能力。

9. 用统计软件分析数据。书中所有的统计分析方法,都运用了使用最为广泛的两种统计软件(Excel 和 SPSS)的分析输出结果来进行讲解。读者在这个高科技的商业社会中肯定要接触这些统计软件输出结果。

10. 统计软件指南。在每一章的最后都给出了“软件的使用”,介绍如何使用 SPSS 和 Excel 操作该章学习到的统计分析方法。读者可以从中得到有用的信息,



并且能够很容易地利用 SPSS 和 Excel 完成各种统计分析的运算。

本书可作为理、工、农、医、经济、管理、人文社会科学专业以及其他领域的统计学教材，也可以供从事商务活动 and 经济分析等实际工作的各类人员参考。读者可以根据时间和需要，有选择地学习有关内容。本书所采用的数据和课件可以在中国人民大学出版社工商管理分社网站（www.rdjg.com.cn）下载。

由于本书是作者第一次编写通俗易懂的统计教材，疏漏和错误在所难免，恳请读者批评指正，以便再版时修改完善。希望本书能引起读者对统计的兴趣，不再惧怕统计，并帮助读者读懂数据。

袁 卫 刘 超

2011 年 8 月

目 录

第 1 章 统计学导论	1
1.1 什么是统计学	3
1.2 如何运用统计学	5
1.3 统计学的基本概念	6
1.4 数据的收集	10
1.5 数据阵/数据文件	18
1.6 统计学和数学、统计软件的关系	19
附录：软件的使用	20
练习	28
挑战性的案例分析	29
 第 2 章 分类数据的描述分析	31
2.1 数据分析的三个原则	33
2.2 频数表	33
2.3 统计图	34
2.4 列联表	44
附录：软件的使用	52
练习	54
应用案例	56
挑战性的案例分析	57
 第 3 章 定量数据的描述方法	59
3.1 展示数据的分布	60
3.2 定量数据的其他图形表示	68
3.3 分布的集中位置	74



3.4 分布的离散程度	85
3.5 分布形态的度量	89
3.6 标准得分	91
3.7 利用描述方法扭曲事实	92
附录：软件的使用	94
公式	95
练习	96
统计图应用的经典案例	100
应用案例	102
挑战性的案例分析	104
第4章 概率	107
4.1 什么是概率	108
4.2 怎样求概率	109
4.3 概率的计算	111
4.4 离散随机变量的分布	113
4.5 连续随机变量的分布	119
4.6 使用概率来检验假设	123
附录：软件的使用	125
公式	129
练习	131
应用案例	133
第5章 区间估计	135
5.1 样本统计量和总体参数	136
5.2 点估计	138
5.3 抽样分布与中心极限定理	140
5.4 区间估计：给结论留一些余地	149
5.5 合适样本量的确定	163
附录：软件的使用	165
公式	166
练习	170
应用案例	172
挑战性的案例分析	173
第6章 假设检验	177
6.1 假设检验的基本思想	178
6.2 正态总体均值的假设检验	183

6.3 总体比例的假设检验	192
6.4 何时拒绝零假设	193
附录: 软件的使用	195
公式	198
练习	199
应用案例	201
挑战性的案例分析	202
第7章 方差分析	204
7.1 单因素方差分析	206
7.2 双因素方差分析	215
附录: 软件的使用	222
公式	224
练习	225
应用案例	228
第8章 相关与回归分析	231
8.1 两个变量间的相关分析	233
8.2 一元线性回归分析	236
8.3 多元回归分析	245
8.4 虚拟变量	251
8.5 模型选择和逐步回归	253
8.6 多重共线性	255
附录: 软件的使用	260
公式	262
练习	263
应用案例	268
挑战性的案例分析	269
第9章 时间序列分析	272
9.1 时间序列的成分	274
9.2 趋势分析	277
9.3 指数简介	285
附录: 软件的使用	289
公式	290
练习	290
应用案例	292



第 10 章 国民经济统计初步	294
10.1 概述	295
10.2 国内生产总值 (GDP)	296
10.3 国民收入指标	305
10.4 价格指数和总体价格水平	307
10.5 失业率	309
10.6 国际收支	311
练习	313
应用案例	314
统计术语	316
附录 1 标准正态分布表	322
附录 2 t 分布表	324
参考文献	325

21 世纪就是本数据书!

电影《美国队长 2》中有句台词：21 世纪就是本数据书。调查显示，中国互联网平均 1 分钟时间里，腾讯收入约 8.3 万元；百度搜索约 347 万次；迅雷下载约 1.1 万 GB；UC 浏览器打开约 460 万个网页；新浪微博有约 9.5 万条微博发送；约 14.8 万人访问天猫；360 手机卫士拦截约 13.6 万条垃圾短信；中国移动净赚约 24.6 万元……此外，2013 年我国产生的数据总量超过 0.8ZB，即 8 亿 TB，相当于 1 200 万个中国国家图书馆藏书量。而 2013 年世界上所存储的数据如果印刷成书，则可以覆盖整个美国 52 次。

所有这些行为都由海量的数据来呈现。这些数据未必都准确，但信息很清晰。这里既有无法改变的，也有可以改变的；既有真实的生活，也有背后的贫富差距。

如今，大数据被越来越广泛地应用。最近中国的雾霾问题备受世界瞩目。利用大数据解决雾霾问题是一种比较可行的思路。2014 年 2 月，美国航空航天局公布了 VIIRS（美国航空航天局的 NPP 卫星搭载的可见光红外成像辐射套件）设备拍摄的亚洲上空的雾霾画面。从中可以清晰地看到中国华北一带的上空是厚厚的灰色雾霾层。有研究机构称，可以根据现有监测站提供的空气质量数据以及城市里的其他多种数据来源（包括气象情况、交通流量、人员流动趋向、路网结构、人口集中点等），运用数据挖掘和机器学习技术，对大数据加以充分利用，并在监测信息和对应结果之间建立一个隐式映射，从而实时推断出包含细颗粒物信息的城市空气质量数据。据悉，中国准备在京津冀、长三角和珠三角地区建立雾霾应急监控对策系统，这个系统依托于“天河一号”计算机，可以对采集到的海量空气质量数据进行分析，以对雾霾作出全面的分析及准确的预报。

我们如何看待统计数据，其实就是如何看待自己，也是如何看待人类。对数据进行科学的分析，不仅对我们自己会有影响，对整个人类都会产生影响，这就是统计科学的力量。



... 学习目标

- (1) 用一句话说明统计学。
- (2) 理解我们为什么要学习统计学。
- (3) 懂得如何运用统计学。
- (4) 知道统计学的基本概念。
- (5) 懂得什么是描述统计学和推断统计学。
- (6) 了解数据如何收集以及数据收集中存在的问题。
- (7) 知道常用的统计软件。

在当今的时代，当你浏览报纸、杂志或者网页时，会发现大量的信息。如何判断各种报道里根据数据所得出的结论是否有道理？样本是否具有代表性？样本是如何抽取的？很多人已经意识到，作为一个信息接收者或者使用者，在日常生活中，对统计学基本概念的了解以及懂得如何用统计去解决问题是非常重要的。正如书名所示，本书可以帮助你理解统计学，熟悉统计语言，并知道如何评估统计结果。如果你想研究统计学，本书将成为你开始漫长却又乐趣无穷的“统计”之旅的一个合适的起点。

为帮助你漫游统计乐园，本书的每一章都是以一些和本章内容有关的实际问题开始的。我们希望这些问题成为你对每道统计大餐的开胃菜。下面就以一些问题来开始第1章。

(1) 根据国际货币基金组织（IMF）2014年4月8日公布的数据，2013年按照国际汇率计算显示中国的国内生产总值（GDP）为9.1814万亿美元，GDP增速为7.7%左右。尽管2013年中国总量GDP约为日本的两倍，中国人均GDP与日本相比从1比10上升到了1比6，在世界排名第86位左右，但是仍然不到世界平均水平的一半。与此同时，我国2013年全国居民收入基尼系数为0.473。这样的结果对我们意味着什么呢？毫无疑问，我们需要知道一些统计知识。

(2) 在金融危机发生后，各国政府对于国家经济的现状和对未来经济走势的预测都十分关注，每个月都要编制消费价格指数（CPI）等各类指数以评价通货膨胀情况。有关商品销售额、新开工的住宅、货币流通以及工业生产的信息仅仅是构成预测基础的成百上千类信息的一小部分。该如何解读这些数据呢？知道一些统计知识有助于我们知道这些数据是如何与我们的生活息息相关的。

(3) 作为一名潜在的汽车购买者和一名有责任心的市民，你乐意为了保护地球生态环境做出应有的贡献。根据最新的研究结果，消费者的行为对自然资源到底有怎样的影响呢？你应该购买柴油发动机的汽车，还是购买电动车，或者干脆骑自行车呢？报纸、杂志或网络上的统计结果对于你的决定非常关键，这些结果会影响你该有怎样的消费观。

(4) 假设你是市场部的新任经理，有一项广告活动的统计结果摆到了你面前，声称某个结果是“统计显著”的。你如何解释这份报告而又不暴露你对该术语的无

知呢? 赶快学点儿统计, 这对你和你的事业都非常有用。

(5) 大数据应用层出不穷。谷歌利用用户搜索记录判断出美国流感疫情的现状, 比疾控中心快一至两周; 对冲基金通过剖析社交网络推特的数据信息来预测股市的表现; Netflix 应用大数据制作热播剧……你会用大数据在未来的生活中产生什么样的发明创造呢?

(6) 近年来受欢迎的新职业, 是统计学家、软件工程师、图形设计师和作家的结合, 目的是从大数据中得到新发现。谷歌首席经济学家哈尔·瓦里安 (Hal Varian) 说: “数据非常之多而且具有战略重要性, 但是真正缺少的是从数据中提取价值的能力。这也就是为什么统计学家、数据库管理者和掌握机器学习理论的人是真正了不起的人。”

经典科幻小说《星际战争》和《时间机器》的作者威尔斯曾说: “终有一天, 统计思考的能力将会和读写的能力一样重要。”威尔斯 100 多年前的预言, 在今天变成了现实。

★人物小传

赫伯特·乔治·威尔斯

出生于英国的作家赫伯特·乔治·威尔斯 (Herbert George Wells, 1866—1946) 在 1895 年出版了他的第一部小说《时间机器》。该书是英国文学中的讽刺文学, 提醒人们发展是不可阻挡的。除了是著名的科幻小说家以外, 威尔斯还是记者、社会学家、历史学家和心理学家, 是一位硕果累累的作家。他对统计思考能力的言论, 只是他对生活众多精辟的言论之一。下面是威尔斯一些有名的言论:



“广告就是合法的撒谎。”

“粗糙的分类和错误的一概而论都是有组织的生活的祸端。”

“今日的危机只是明天的玩笑。”

“愚蠢的人做研究者, 聪明的人利用研究者。”

“衡量成功唯一真正的标准是, 我们可能会做什么和我们实际已经做了什么的比较, 以及我们已经取得什么和我们自己为别人做出什么。”

1.1 什么是统计学

在 300 多年前, 统计的英文单词 statistics 首次被人们应用, 指政府部门记录人们出生和死亡信息的工作。时至今日, 统计仍然是世界上各个层次的政府机构的支



柱。全球统计已成为许多国际组织，像跨国公司、联合国和一些关注人口密度、生态灾难和疾病流行的组织等重点关注的对象。2010年6月3日，第64届联合国大会第90次会议通过决议，确定当年10月20日为首个“世界统计日”。2011年初，我国国务院学位委员会在新的研究生专业目录中将统计学上升为一级学科，为统计学科和统计教育的发展提供了更广阔的平台。同时，也凸显出统计对科学研究、经济增长和社会发展的重要性。

“statistics”这个单词通常有两个含义：当它以单数名词出现时，表示一门学科的名称“统计学”；当它以复数名词出现时，表示“统计数据”或“统计资料”，此时，统计可以认为是某种形式的数字，例如北京的月平均降水量、上海的周平均气温、中国持有的外汇储备余额等。我们面对着所谓的“信息爆炸”，数字化的信息无处不在。现代社会似乎对统计有一种永不满足的需求。正是基于这种需求，统计学家收集越来越多的统计数字。统计数字又叫数据。

“statistics”的一个简单意思就是数值数据。进入信息时代后，人们趋向把所有存储在计算机上的信息，无论是数字还是音乐、视频都统称为数据。2010年，美国总统科学技术委员会在给总统和国会的报告中提到：数据正在呈指数级增长，联邦政府的每个机构和部门，都需要制定一个应对“大数据”的战略。现在大数据战略被认为是世界下一个创新、竞争和生产力提高的前沿，是争夺全世界的下一个前沿。在本书中，我们感兴趣的是如何收集数据和处理数据。然后，我们希望通过统计的帮助把数据中的信息变成实际的知识。

简而言之，统计学是研究收集数据、整理数据、分析数据以及由数据分析得出结论的方法，简称为“数据的科学”。统计学有时候也指一门学科，例如你正在通过本教材学习的统计学课程。在统计学中，统计学家探索、开创各种收集数据和分析数据的方法以得出结论。他们从数学公式中为统计设计出新的应用，并通过实践来检验理论模型。虽然统计学没有自己的基于实验或观察的经验研究对象，它不像化学、社会学或心理学这类学科研究已经定义得很好的某种现象，但是，统计学为经济学家、社会学家、化学家、心理学家以及其他人士提供了一套研究对象的方法。

英文“statistics”的起源和中文“统计”一词的来源

英文中的“statistics”大约在18世纪中叶由德国学者阿亨瓦尔（Gottfried Achenwall）所创造，是由状态 status 和德文的政治算术联合推导得出的，首次由辛克莱（John Sinclair）所使用，1797年出现在《大英百科全书》中。

统计一词的本意是国势学，即关于一个国家基本情况的调查。这些情况可以是描述性的事实，也可以是数据，其中最重

要的一部分，就是人口情况。18世纪，又有人把这种基于实情的条目放到了表格中，按行和列组织起来，这样更容易对比，因此促进了数据的使用。

基本上和德国同期，英国人也开始探索统计在政治和社会生活中的作用。1662年，约翰·格朗特（John Graunt）调查了伦敦的人口死亡情况，分析了生男孩和生女孩的比例，发展了现在保险公司所用的那种类型的

死亡率表。随后出版了他第一本也是唯一一本手稿《对死亡的自然和政治观察》，该书被后世公认为是统计学的开山之作。

中文的“统计”一词来源于日本。在日本明治维新时期，日本向西方国家学习，成立了国家统计局。1872年进行了第一次

全国人口普查，1882年出版了第一本统计年鉴。1903年，中国学者翻译了日本人横山雅南所著的《统计讲义录》一书，把“统计”一词引入中国。

资料来源：徐子沛：《数据之巅》，北京，中信出版社，2014。

★人物小传

瓦尔德

有人把统计学定义为数据处理的一门艺术，瓦尔德(A. Wald, 1902—1950)的事例就是明证。瓦尔德是第二次世界大战时期的统计学家，他发明的一些统计方法在战时被视为军事机密。当有人向瓦尔德咨询飞机上什么部位的钢板应该加强时，他开始研究从战役中返航的军机上受敌军创伤的弹孔位置。他画了飞机的轮廓，并且标示出弹孔的位置。资料累积一段时间后，几乎把机身各部位都填满了。于是瓦尔德提议，把剩下少数几个没有弹孔的部位补强……因为这些部位被击中的飞机都没有返航。这是一个简单但近乎完美的实例，简单的统计方法一旦融入了统计学家的智慧，便显得生动而唯美！



我们希望你学习本书之后，能够通过统计学的帮助欣赏到数据是如何转化为比数字本身更为复杂、更为有用的知识的。

1.2 如何运用统计学

统计在许多学科中得到广泛应用，可以毫不夸张地说，我们已经被统计分析结果包围。比如说，学术研究杂志就依赖于统计结果。在许多学科中，一篇文章是否能够发表在主要杂志上，在很大程度上取决于该文章是否能正确地使用统计方法。

除了学术领域，统计在其他领域也广泛使用。统计在工业中得到大量使用，尤其用于研究新产品、质量控制和市场开发中；市场调查也需要统计。在当今时代，很难想象不需要市场调查来获知销售者对于企业产品或服务的看法。产品或服务的改进依赖于由统计得到的消费者的反馈信息；同样，没有一份报纸、杂志不刊登以统计为基础的文章。我们在报纸、杂志、网络等媒介上可以读到诸如在2010年上海世博会期间参观的人数，为汶川和舟曲等地震灾区捐款的金额，大学生就业率等



统计数据；甚至在电视节目中，主持人的出现和电视剧的播放都依赖于统计，因为只有排名靠前受观众喜爱的主持人和节目才能在竞争中生存。



想一想 1.1

请从你的实际经验中想出一些以统计分析为决策依据的例子，并思考出现这种统计结果的原因。

学习统计可以帮助我们理解什么可能出差错。作为信息接收者或者使用者，如果我们希望能完全理解统计的广泛应用，就应知道为得到所读到的结果而需要的原则和方法。对统计的了解可以帮助我们评价这些结果，并使我们随时保持警觉，因为从开始叙述问题到完成最终报告的过程中随时可能出现一些错误。

《读者文摘》的总统选举调查

这个故事有关一次著名的失败的统计调查，它一直是一个统计逸事。在1936年美国总统选举前，一份名为《读者文摘》(Literary Digest)的颇受人尊重的杂志进行了一次民意调查。调查的焦点当然是谁将成为下一届总统——是挑战者，堪萨斯州州长阿尔夫·兰登(Alf Landon)，还是时任总统富兰克林·德拉诺·罗斯福(Franklin Delano Roosevelt)。为了解选民意向，民意调查专家根据电话簿和车辆登记簿上的名单给一大批人发了简单的调查表(电话和汽车在1936年并不像现在这样普及，但是这些名单比较容易得到)。尽管发出的调查表大约有1000万份，但收回的比例并不高。在收回的调查表中，兰登非常受欢迎。于是，该杂志预测兰登将赢得选举。

如果你有一些统计知识，你会对这个声称兰登将赢得选举的预测结果怀有疑问。正如你所怀疑的，在经济大萧条时期调查拥有电话和汽车的人，并不能很好地反映全体选民的观点。此外，只有少数的调查表被收回，这一点也是值得怀疑的。事实表明，最终是罗斯福而不是兰登赢得了这次选举。由此可见，该杂志的调查结果有多么错误。当前大多数应用统计不会像这个例子错得那样离谱，但即便在今天，我们也很容易发现统计被误用的情况，尤其在需要考虑选择正确的样本时。

需要提及的一点是，当时盖洛普(Gallup)也进行了调查，它的调查结果是罗斯福获胜。在这次总统选举之后，盖洛普成了一家比较成功的调查公司。

1.3 统计学的基本概念

1.3.1 随机性和规律性

当不能预测一件事情的结果时，这件事就和随机性联系起来了。例如，掷硬币