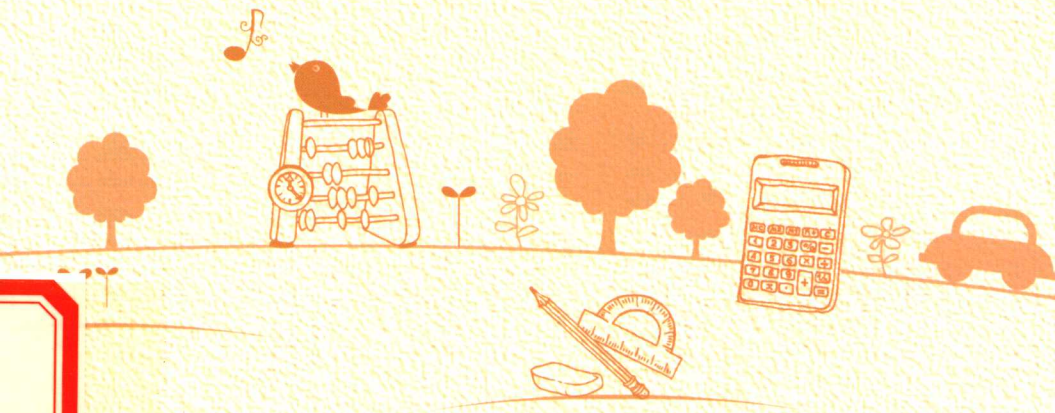
项 昭　分册主编
赵 慧　编　著

快乐阅读书屋

数据“炼金术”——统计

happy reading 数学知识类



贵州出版集团
贵州人民出版社

数据“炼金术”

——统　计

赵　慧　编著

贵州出版集团
贵州人民出版社

图书在版编目(CIP)数据

数据“炼金术”——统计 / 赵慧编著. —贵阳：
贵州人民出版社，2013.9
ISBN 978-7-221-11362-7

Ⅰ.①数… Ⅱ.①赵… Ⅲ.①统计学 Ⅳ.①C8

中国版本图书馆 CIP 数据核字(2013)第 201392 号

数据“炼金术”——统计

赵 慧 编著

出版发行	贵州出版集团 贵州人民出版社
地 址	贵阳市中华北路 289 号
责任编辑	徐 一
封面设计	熊 锋
印 刷	贵阳经纬印刷厂
规 格	850mm×1168mm 1/16
字 数	100 千字
印 张	8.5
版 次	2014 年 7 月第 1 版
印 次	2014 年 7 月第 1 次印刷

书 号：ISBN 978-7-221-11362-7 定 价：17.00 元

出版说明

兴趣是最好的老师，知识的学习更是如此。如果学习者缺乏兴趣，阅读就将是一个枯燥无味的过程，轻松快乐的学习也就无从谈起。基于这样的事实，本着“兴趣阅读、快乐学习”的理念，我们经过深入调研，与国内的众多专家学者及一线教师全力合作，为所有希望将学习变得轻松愉快的朋友奉献上“快乐阅读”书屋。

“快乐阅读”书屋，以知识的轻松学习为核心，强调阅读的趣味性。它力求将各种枯燥无味的知识以轻松快乐的方式呈现，让读者朋友便于理解接受。它的各种努力，只有一个目标，即力图将知识学习过程轻松化、趣味化。读者朋友在阅读过程中，既能保持心情愉快，又能学有所得。在轻松愉快的氛围中学习，让知识学习成为读者朋友的兴趣，本身就是提高学习效率最有效的途径。

“快乐阅读”书屋首批图书分为“语文知识”、“作文知识”、“数学知识”、“文学导步”、“文学欣赏”、“语言文化”、“个人修养”七大板块，各个板块之下又有细分。英语、生物、化学等相关的知识板块将会在以后陆续推出。针对不同学科知识的特点，本书屋以不同的方式来达到轻松快乐的目的。要么是以故事的形式，在故事的展开之中融入相关知识；要么是理清该知识点的背景，追根溯源，让读者朋友知其然，更知其所以然，让理解更为轻松。总而言之，就是以最恰当的方式呈现相关的知识。

希望这套“快乐阅读”书屋能陪伴每一位读者朋友度过美好的阅读时光。

编　者

2014 年 5 月

亲爱的读者朋友，你有没有发现：我们生活的世界中到处都能发现数据的身影。例如，班里每一位同学的身高、体重，每一次考试的成绩，就是一串串数据。又如，产品的合格率、商品的销售量、电视台的收视率等等都要用数据表示。那么，生活中的这些数据背后有没有隐藏着什么秘密呢？

细心的读者朋友，你有没有留意：老师总喜欢计算每一次考试的平均成绩？电视媒体上总说某一年度城市居民的平均收入、平均消费？这个“平均”又有怎样的含义呢？

睿智的读者朋友，你有没有注意：在班级里，大多数同学的身高、体重都相差无几，只有少数同学的身高、体重格外突出。这又说明什么呢？

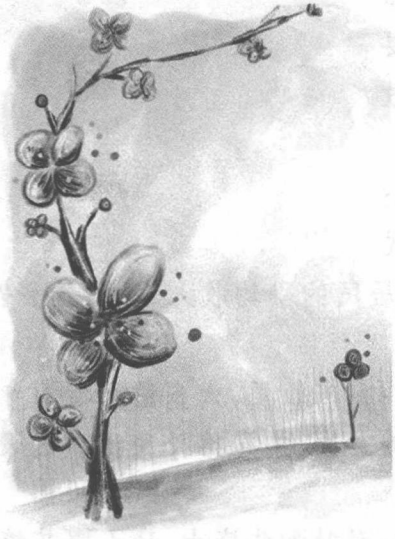
聪明的读者朋友，你有没有想过：面对如此庞大、让人眼花缭乱的数据世界，有没有简单明了、好学好用的方法来处理这些数据，揭开它们隐藏的秘密，从中提炼出有用的信息呢？

我们知道，要分析、处理、揭示数据的规律，就要从数据的收集、整理做起。那么，你是否知道如何科学地收集数据呢？或许有人会说：这些数据是可以通过调查获得的，那么，又该怎样调

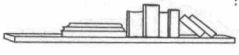
查呢？

难道为了调查饮料的容量是否达标，就要把所有的饮料都打开吗？要检验节能灯的使用寿命，就要把所有产品逐一尝试吗？……现实生活中，谁也不会这么做！我们往往会选择整体中的一部分作为调查对象，可这些部分的信息就能够代表整体吗？……

为了解答这些问题，就让我们跟随本书的两个小导游，聪明的小蓬蓬和可爱的小依依，通过收集到的瓶装饮料容量的数据，在一问一答中，走进统计的世界，将这些问题一一揭晓吧。



目　录



第一章

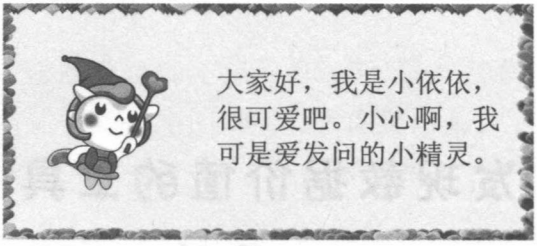
发现数据价值的工具
——统计

大家好，我们是数据世界的小精灵——小蓬蓬和小依依。在这个科技发达的世界里，随时有各种各样的数据在不断穿梭，有些数据很直观，而有些却不那么明显。大千世界丰富多样的色彩、悦耳动听的声音、绚烂多姿的语言等等，所有这些文字、图像和声音其实都是蒙着面纱的数据，因为它们都可以被人们转化，并借助适当的工具作为数据传播、处理和存储。所以说，我们生活在一个数字化的时代，时刻都在与数据打交道。如果你细心一点，就可以听到这些数据在耳边吟唱。接下来，请你和我们一起仔细聆听这些数据演奏的美妙乐曲吧！

我是小蓬蓬

我是小依依

当然，在这趟数据海洋的旅程中，大家可要注意了，小依依可是喜欢不断地提出各种各样的问题，挑战大家的啊！

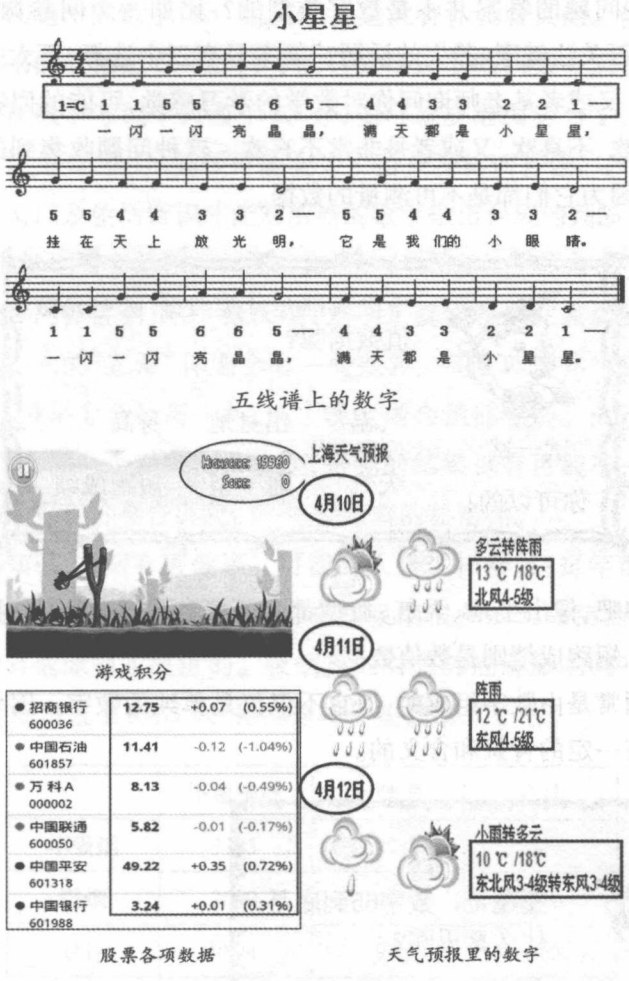


你的生活中是否也出现过这样的场景呢？

早晨6点30分起床，洗漱完毕后7点出门上学。路上买了份3.5元的200ml盒装牛奶和一块2元的面包做早餐，乘坐公交车还花了1元钱。7点40分到学校开始一天的学习。数学老师教会大家如何计算期中考试各科平均分，音乐老师教会大家识别《小星星》的五线谱，信息老师教会大家设计图片的大小和位置。放学回家后，再兴高采烈地在每个小格子里涂上数字对应的油彩，完成数字油画——毕加索的向日葵。然后开心的玩上30分钟的小游戏——愤怒的小鸟，看看自己的积分，乐此不疲。妈妈在厨房里烹饪从市场里买来的18元一斤的鲈鱼、2元一斤的白菜、4.5元一斤的辣椒，心里还计算着明天的生活费该如何支出。而爸爸则看着股市信息，分析着股票从17.8元跌到17.3元的对策，一筹莫展。爷爷奶奶看着天气预报，说明天的温度在13摄氏度至24摄氏度之间，嘱咐着明天该穿什么，是否要带雨具。

这就是我们平凡普通的一天，而每一天我们都会得到许多数据。

钟表上的数字



上面图中框出来的可都是最直观的数据。再如，产品的合格率、农作物的产量、商品的销售量、电视台的收视率等也都是我们在生活里触手可及的各种各样的直观数据。

有一些问题的答案是明明白白的数据。当有人问起你的年龄时，你会回答 15 岁；当有人问起你的体重时，你会回答 49 公斤；当有人问起你的考试成绩时，你会回答 85 分。这些可测量的数据被称为数值数据。

不过还有很多的数据并不直观，它们属于不同类型的数据。你有没有

留意过有些问题的答案并不是数字类型的？比如一个问卷调查的问题："你平时是否关注健康、养生的话题？"答案只有三个选择：不太关注、一般、非常关注。又或者某老师询问你对数学的学习感觉，可能的回答有非常喜欢、有点喜欢、不喜欢，又或者是非常不喜欢。这种问题收集到的数据称为分类数据，因为它们都是不可测量的数据。

试一试

你可以的！

下面哪些是分类数据，哪些是数值数据呢？

气温　　出身地　　身高

天气　　血　型　　短跑成绩

很简单吧，像出生地、天气、血型都是不可测量的，属于分类数据。而气温、身高、短跑成绩则是数值数据。

数据通常是由数字组成的，但它不仅仅是单纯的数字。因为任何一个数字都是有一定的背景和含义的。

蓬蓬哥，数字65到底是什么意思呢？

小依依

让我想一想，想一想啊！

小依依问题中的数字“65”就其本身是没有什么含义的。但是如果我们得知一本书的价格是65元时，可能会觉得这本书稍微有点贵。如果某位同学在一次数学考试中只考了65分，他可就要继续努力了。又或许某位姐姐的体重是65公斤，她很可能会为了苗条身材而苦恼。所以，我们需要配合上下文以及生活常识才能对所给的数字做出合理的判断。

然而，数据也会有差异。大家都知道人的体温一般都在37摄氏度，可如果你测量到的体温高于37摄氏度，是不是就代表你在发烧呢？也许不是。因为每个人的“正常”体温会有一些差异。甚至就连你自己的体温也会有变化，一般早上会稍高一些，到了晚上则会稍低一些。由于数据总是有差异，甚至对同一个对象测量多次，得到的结果也有可能不一样。这些出现差异的数据也许会告诉我们一些隐藏着的秘密呢。

大家都知道：吸烟有害健康。可能有人会这样辩解：每年都会有一些从不吸烟的人因为患肺癌而病逝，可有些吸烟很多的人却活到八九十岁，最后是因为其他原因才离世的。吸烟真的对患肺癌有影响吗？某肿瘤研究所“随机”地调查了9965人，得到这样一些数据。

	不患肺癌	患肺癌	总计
不吸烟	7775	42	7817
吸烟	2099	49	2148
总计	9874	91	9965

通过这张数据表格，可以粗略地估计出：在不吸烟的调查对象中，有0.54%的人患肺癌；在吸烟的调查对象中，有2.28%的人患肺癌。直观上可以得到的结论是：吸烟和不吸烟的人患肺癌的可能性是有差异的，吸烟更容易引发肺癌。聪明的统计学家会这样撰写总结报告：不吸烟可以将肺癌的死亡率减少17%～34%，我们有95%的信心确保真正的比例会落在这个范围内。细心的你有没有注意过诸如“95%的信心”以及“有统计上的显

著意义”等常用字眼呢？

喜欢打网络游戏的学生一定就比不打网络游戏的学生表现得差吗？最好不要太快下结论。如果喜欢打网络游戏的某位同学学习基础好，学习方法好，专注听讲，认真完成作业。而有位不喜欢打网络游戏的同学听课却从不认真，而且学习效率低。那么很明显，前者应该表现得比后者要好吧。这说明有很多因素是有关联的。这些信息中的关联，如学生的学习基础、学习能力与考试成绩的关系是强还是弱，又该如何表达呢？

其实，数据收集是任何一门科学的基础，但是只有数据是不够的。任何仔细收集起来的信息，在发现它的价值之前，都会对这些数据提出一系列的问题。比如：

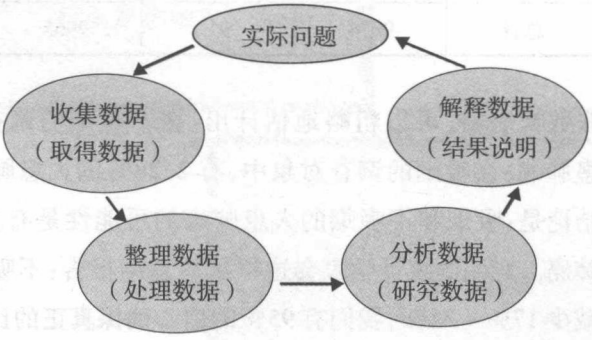
☆ 数据中的最大值是多少？数据中的最小值是多少？

☆ 所有数据的平均值是多少，又意味着什么？

☆ 这些数据是如何围绕平均值分布的？

☆ 各种不同类的数据之间又有怎样的关系？

所有的这些问题都很重要，因为每个问题都可以帮助研究者更多地了解数据代表了什么。倘若人们要揭开数据背后的秘密，就需要数据分析的帮手——统计的鼎力帮助。统计学通过图表和计算工具的运用，从数据中找信息、找线索，从而发现这些数据所表示的意义，并且做出结论，进而为人们下一步的工作给出参考意见和建议。



统计研究的过程

在统计学研究的过程中，需要用到很多数学知识。但统计学和数学的差别在于：统计学不是一门纯粹演绎的学科，它既是艺术，也是科学，既涉及个人判断，也涉及仔细的逻辑推导。

在日常生活中，人们对“统计”术语常常有不同的用法。例如，企业将每年“统计”的销量和利润，作为常规工作来看待；股民将“统计”的成交额和股票指数，作为买卖股票的指导信息来应用。那么究竟该如何理解“统计”呢？所谓统计，就是人们认识客观世界总体数量变动关系和变动规律的活动的总称，是人们认识客观世界的一种有力工具。

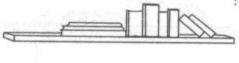
从一般意义而言，统计学是描述一系列可用于描述、整理和解释资料或数据的统计工具和技术。也有人将统计分为两类：一是描述统计，二是推论统计。“描述统计”是指对采集的数据进行登记、审核、整理、归类，并在此基础上进一步计算出各种能反映研究对象的综合指标，以图表的形式表示经过归纳分析而得到的各种有用的统计信息。描述统计常用于整理、描述所收集数据的特征。而“推断统计”是指在采集的数据进行描述的基础上，利用一定的方法（如参数估计与假设检验方法）去估计或检验研究对象的数量特征。推断统计通常利用较小群体的数据来推论较大群体的特征，是数据收集和汇总后的下一步。

假如表1－1是班级某次数学期中考试成绩，而班里只有15位同学，从这些成绩里你能看出些什么信息呢？

表1－1　数学期中考试成绩

学号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
成绩	72	85	55	49	66	84	82	78	95	91	83	80	37	100	83

还好只有15个成绩数据，还是比较容易的找出：班级的最高成绩为100，最低成绩为37。然后，老师会说这次考试平均分为76.33。15位同学中仅有5位同学的数学成绩低于平均分，多数成绩还是与平均分比较接近的。通过这些简单的分析还是可以描述出这些数据的特征的，这就是一个

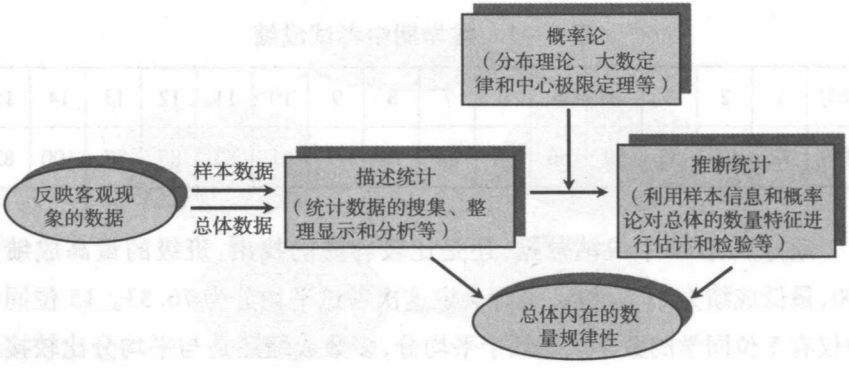


简单描述统计的过程。

可再假设一下，这15位同学的数学期中成绩仅仅是全年级300位学生中的一个很小的部分，那么如何通过少量的数据去估计全部学生的数学考试成绩呢？考试成绩不理想是不是与数学练习不足有关呢？这就是收集描述数据之后的下一步工作——推断统计了。

又如，民意测验中某一位候选人是否能够当选？全国婴儿的性别比例如何？某种电子产品的使用寿命有多长？等等，都是需要用到推断统计的方法来解决的。

描述统计学和推断统计学的划分，一方面反映了统计方法发展的前后两个阶段，同时也反映了应用统计方法探索客观事物数量规律性的不同过程。显然，描述统计和推断统计是统计方法的两个组成部分。描述统计是整个统计学的基础，推断统计则是现代统计学的主要内容。由于在对现实问题的研究中，所获得的数据主要是样本数据，因此，推断统计在现代统计学中的地位和作用越来越重要，已成为统计学的核心内容。当然，这并不等于说描述统计不重要，如果没有描述统计收集可靠的统计数据并提供有效的样本信息，即使再科学的统计推断方法也难以得出切合实际的结论。从描述统计学发展到推断统计学，既反映了统计学发展的巨大成就，也是统计学发展成熟的重要标志。



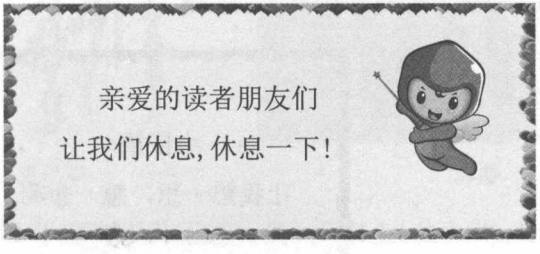
描述统计与推断统计的关系

换句话说，统计学是帮助我们认识理解周围世界的工具。这是通过我们收集到的数据来实现的，而且接着还可以让我们做出特定的推断，也就是怎样将那些数据的特征应用到新的情况当中。描述统计和推论统计可以一起发挥作用，使用哪一种、何时使用取决于你想要回答的问题。

随着社会、经济和科学技术的发展，统计的范畴已覆盖了社会生活的一切领域，并发展成为有着许多分支学科的科学。如生物统计、工程统计、心理统计、教育统计等等。运用统计学的知识和方法，可以计算保险政策下的保费，可以用来阐述经济政策，可以为交易股票和债券方面作决定，甚至还可以用来识别罪犯。很难想像一个科学机构、媒体、大公司或政府部门，不收集、分析和使用统计学的。它是一门必不可少的科学，是畅游在数据海洋里的精灵。甚至有统计学家这样说到："统计方法的应用是这样普遍，在我们的生活和习惯中，统计的影响是这样巨大，以致统计的重要性无论怎样强调也不过分。"

近几十年间，随着计算机技术不断发展，使统计数据的搜集、处理、分析、存贮、传递、印制等过程日益现代化，提高了统计工作的效能，无形中也促使统计科学和统计工作发生了革命性的变化。

统计的发展及其未来，已经被赋予了划时代的意义。也难怪有人说：未来将是统计的时代。



第二章

统计思想的发展历程

一日，聪明的小蓬蓬准备款待朋友，还请来可爱的小依依帮忙。两个人一大早赶到超市，购买了各种各样好吃、好喝、好玩的东西。忙碌了一天后，两个人终于可以喝着可口的饮料稍事休息一下，聊聊天了。此时，刚好一栏电视节目正在介绍密码的发展历程。小依依皱了皱眉头，心想：生活是数据的海洋，统计是畅游在数据海洋里的精灵，那精灵也该有自己的成长历程吧？

蓬蓬哥，任何一门学科都有自己的成长经历，那么“统计”又是如何形成、发展的呢？

小依依

让我想一想，想一想啊！

统计作为一种社会实践活动已有悠久的历史，当然有必要从历史角度