

汉 · 译 · 管 · 理 · 学 · 世 · 界 · 名 · 著 · 丛 · 书



Mastering Statistics

Tim Hannagan

统 计 学

[英] 堤姆·汉拿根 著

陈宋生 朱 丽 译



经济管理出版社
ECONOMY & MANAGEMENT PUBLISHING HOUSE

汉 · 译 · 管 · 理 · 学 · 世 · 界 · 名 · 著 · 丛 · 书

Mastering Statistics

Tim Hannagan

统 计 学

[英] 堤姆·汉拿根 著

陈宋生 朱 丽 译



经济管理出版社
ECONOMY & MANAGEMENT PUBLISHING HOUSE

北京市版权局著作权合同登记：图字：01—2005—2899号

Mastering Statistics by Tim Hannagan.

Copyright © T.j. Hannagan, 1982, 1986, 1997.

This Translation of Mastering Statistics, First edition is published by arrangement with Palgrave Limited.

Chinese translation(simplified characters) copyright © 2007 by Economy & Management Publishing House.

All rights reserved.

图书在版编目 (CIP) 数据

统计学/(英)汉拿根著;陈宋生,朱丽译. —北京:经济管理出版社, 2011.4

ISBN 978-7-5096-1359-7

I. ①统… II. ①汉… ②陈… ③朱… III. ①统计学
IV. ①C8

中国版本图书馆CIP数据核字(2011)第061127号

出版发行：经济管理出版社

北京市海淀区北蜂窝8号中雅大厦11层

电话：(010)51915602 邮编：100038

印刷：三河市海波印务有限公司

经销：新华书店

责任编辑：申桂萍

技术编辑：晓 成

责任校对：全志云

787mm × 1092mm/16

15印张 309千字

2011年8月第2版

2011年8月第1次印刷

定价：30.00元

书号：ISBN 978-7-5096-1359-7

· 版权所有 翻印必究 ·

凡购本社图书，如有印装错误，由本社读者服务部

负责调换。联系地址：北京阜外月坛北小街2号

电话：(010)68022974 邮编：100836

致 谢

本书作者和出版商感谢英女王殿下之文书局管理人员允许本书使用《社会趋势》的数据与信息。



本书介绍了大量的统计基础知识，书中涉及有限的一些简单的代数和基本的数学技巧知识，适合各类读者学习。

本书旨在全面深入地介绍统计资料、统计概念、统计方法和统计应用，所以非常适合作为高级数量分析或数值分析的基础课程。

书中着重强调了各种日常商业环境中涉及的统计。本书主要适合于下列读者：

- 参加考试、培训课程的学生；
- 希望对事实和数字有更深了解的普通读者；
- 需要了解、使用统计信息，行业中需要使用统计方法的商务人士。

本书内容主要覆盖英国普通中等教育考试（GCSE）的大纲要求，适用于国家职业资格考试（GNVQ）和其他同类考试。第三版在第二版的基础上，根据此期间的技术发展进行了章节更新。每章的开始增加了“学习目标”，内容表述更加通俗易懂。本书中增加了统计考试知识及关于归纳统计的讨论，更深入介绍了“数据收集方法”。书的最后还介绍了一些统计应用领域，继续学习的方向以及统计对信息技术的影响，更新了公式表、统计的主要术语和参考资料表。“如何通过统计考试”一节专门给参加考试的考生提供一些建议和帮助。

本书可以作为社会学科、公共管理学科和其他很多学位资格考试的统计用书。书中提供了诸如 GNVQ、NVO 及其他管理课程、职业课程需要的数字分析基础知识。

每章的最后都备有习题，以供巩固章节知识。作业适合个人习作，也可以由小组完成。尽管不做任何习题也可以读懂书中的内容，但在允许的情况下尽可能完成习题。习题不仅为考试提供练习的机会，也有助于理解特定的方法和概念在实际中如何应用。如果需要更多的习题，可以参考本书的辅助教材《学会统计学》（提姆·汉拿根，麦克米兰出版公司，第二版，1996），其中有大量习题及答案。

如果读者通过习读本书，了解了统计信息的主要来源，理解了数字概念，掌握了基本的统计方法，能够使用各种统计技术解决商业问题或日常生活问题，即达到了本书的目的。

提姆·汉拿根

目 录

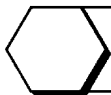
第一章 统计信息	1
第一节 统计和统计学的含义	1
第二节 统计信息	2
第三节 数字数据	2
第四节 统计的误用	3
第五节 统计的使用	4
第六节 第一手数据和第二手数据	5
第七节 第二手统计结果	7
第八节 计算工具	8
习题	9
第二章 信息的准确性	11
第一节 近似	11
第二节 容忍度	12
第三节 误差	13
第四节 四舍五入	14
第五节 绝对误差	17
第六节 相对误差	17
第七节 偏倚误差	18
第八节 无偏误差 (修正误差)	18
第九节 涉及近似值和误差的计算	19
第十节 表面上的准确性	22
习题	23
第三章 统计信息的收集	25
第一节 调查	25
第二节 调查方法	27
第三节 观察调查	28

第四节 访问调查	30
第五节 问卷调查	32
第六节 为什么抽样	34
第七节 抽样的优点	34
第八节 抽样的目的	36
第九节 抽样的基础	36
第十节 抽样误差	37
第十一节 样本量	38
第十二节 抽样设计	38
第十三节 抽样中的偏倚	39
第十四节 抽样方法	40
习题	49
第四章 数据的列示 (1)	51
第一节 列示的目的	51
第二节 表格	53
第三节 数据分组	54
第四节 频数分布	56
第五节 数据报告	58
第六节 直方图	59
第七节 频数多边形图	62
第八节 频数曲线	64
习题	65
第五章 数据的列示 (2)	67
第一节 条形图	67
第二节 饼图	70
第三节 比较型饼图	71
第四节 象形图	73
第五节 比较型象形图	74
第六节 统计地图	74
第七节 分层图	75
第八节 坐标图	75
第九节 半对数曲线	79
第十节 比例曲线	80
第十一节 甘特图	81
第十二节 收支平衡图	82

第十三节 Z形图	83
第十四节 洛伦兹曲线	84
第十五节 列示和理解	86
习题	89
第六章 数据的分布特征：平均数	91
第一节 平均数的作用	91
第二节 算术平均数	93
第三节 中位数	98
第四节 四分位数	101
第五节 众数	103
第六节 几何平均数	105
第七节 调和平均数	106
习题	106
第七章 数据的分布特征：离散程度	109
第一节 离散程度	109
第二节 极差	111
第三节 四分位差	112
第四节 标准差	115
第五节 方差	122
第六节 离散系数	122
第七节 结论	123
习题	123
第八章 统计决策	125
第一节 估计	125
第二节 概率	125
第三节 正态曲线	131
第四节 概率与正态曲线	134
第五节 标准误差	135
第六节 显著性检验	136
第七节 置信区间	138
第八节 统计质量控制	139
第九节 结论	140
习题	141

第九章 统计量的比较：指数和重要的统计数据	143
第一节 指数	143
第二节 加权指数	145
第三节 环比指数	147
第四节 指数构建中的一些问题	148
第五节 零售价格指数	149
第六节 重要的统计数据	150
习题	151
第十章 统计量的比较：相关性	153
第一节 什么是相关性	153
第二节 散点图	155
第三节 交叉列联表	157
第四节 简单相关系数	157
第五节 等级相关系数	159
第六节 伪相关	160
第七节 回归分析	161
习题	162
第十一章 趋势和预测	163
第一节 趋势和预测	163
第二节 时间序列	164
第三节 移动平均	165
第四节 季度性波动	166
第五节 不规则及残差波动	170
第六节 线性趋势	171
第七节 结论	176
习题	176
第十二章 掌握更多的统计知识	177
第一节 更多的统计知识	177
第二节 统计证明	177
第三节 实验	180
第四节 更多的推断统计量	183
第五节 更多的数据收集方法	186
第六节 调查表和评级	187

第七节 量表	189
第八节 金融统计	194
第九节 信息技术	198
第十节 未来统计	200
习题	200
附录 1 推荐读物	201
附录 2 面积表	203
附录 3 数学基础知识	205
附录 4 如何通过统计考试	221



第一章 统计信息

目标：

1. 掌握统计和统计学的含义
2. 分析统计信息和数据的本质
3. 掌握正确使用统计的方法

第一节 统计和统计学的含义

统计学是一门收集、整理、总结、描述和分析数据，并在分析的基础上做出正确结论和合理决策方法的科学。

数据 (Data) 是一个英文复数名词，单数形式很少使用，这表明数据不是指单个的数字，而是由多个数据构成的数据集。随着社会经济情况日趋复杂，数据越来越受到我们的重视，数据分析已经成为科学判断的重要基础。统计学是一套自成体系的数据收集和分析方法。

人们已经赋予“统计”一词多种含义，不同的场合“统计”可以有不同的含义。它可以指：

- 事实数据，如一个小镇某一时点的常住人口数；
- 收集和分析数据事实的方法，以及关于这些方法的研究，即统计学。

我们也可以说，数字本身并不代表任何客观事实。只有当数字被分析和解释之后，才能对研究和决策产生影响。

第二节 统计信息

现代社会中信息无处不在，每一个人都在收集、分析和使用信息，这些信息很多以数字或统计数据的形式存在。人们每天通过谈话、看电视、使用电脑、上网、收听广播、看报纸、广告、通知和说明书等能够得到大量的信息。可用的信息充斥着整个生活，人们需要选择接受哪些信息，拒绝哪些信息，从大量事实中筛选出对自己有兴趣的或对自己有用的信息。无论在商业、工业还是日常生活中，统计信息都是非常必要的，我们需要掌握收集这些信息的渠道和方法。

作为消费者，我们在决定买什么之前都要比较不同商品的价格和质量；作为应聘者，我们要比较不同公司的工作环境和待遇；作为老板，公司希望尽量降低成本，增加利润。

统计最主要的用途就是对我们要做的这些选择提供信息和帮助。不利用统计方法提供的信息做决策判断很可能就会造成决策失误，就好像医生不测量病人的体温、脉搏和血压，不检查病人的身体情况就对病症做结论，这个诊断肯定会受到病人的质疑。统计就是描述事物的现状，刻画过去，并预测未来，这些都能为我们的判断提供科学的基础。

第三节 数字数据

数字数据的处理简单方便，因此被广泛使用。但是，实际中又有一种尽可能使用文字、图表信息，避免数字的趋势。非常有趣的是，我们在普及教育，扫除文盲的同时对身边的数盲却无动于衷。只有在体育运动中，我们不会避开数字，数字和表格是可以广泛接受的。人们对理解棒球的击球率和足球彩票毫无困难，也许这说明人们都有能力去理解复杂的数字信息，缺乏的只是足够的动力和热情。

许多事实只有用数字才能清楚地说明。比如金融信息，足球比赛中的排名表都惯用数字来传递大量信息（表 1.1）。如果你能进行加减乘除运算，看得懂表 1.1，那掌握统计学应该没有什么大问题（数学基础知识见附录 3）。

显示数字数据

表格是简洁明了地显示统计信息的方法之一。表 1.1 中描述的信息可能需要密密麻麻的几页纸才能写完，如果要分析这些信息，还要再写几页纸。比如在表 1.1 中，每个队伍获胜的场次和排名之间关系密切，但不是完全对应的关系。比如阿森纳队取得了 46 分，赢了一半的场次，谢菲尔德星期三队也取得了 46 分，但仅赢了 46% 的场次。谢菲尔德星期三队平了 37% 场次的比赛，而阿森纳队仅平了 14%，平场数目平衡了它们的得分。谢菲尔德星期三队比阿森纳队排名靠前是因为球差（进球—失球）更大，净胜球数更多。

表 1.1 联赛积分榜

队名	比赛				进球		积分
	场次	胜	平	负	进球	失球	
埃弗顿	27	17	5	5	60	31	56
托特纳姆	27	16	6	5	53	27	54
曼联	28	14	7	7	51	32	49
利物浦	28	13	9	6	40	23	48
谢菲尔德星期三	27	12	10	5	42	26	46
阿森纳	28	14	4	10	48	37	46
温布尔登	28	13	7	8	38	33	46
利兹	27	14	3	10	41	35	45
切尔西	28	10	10	8	42	33	40
阿斯顿维拉	27	10	7	10	38	41	37
布莱克本	27	10	7	10	33	37	37
纽卡斯尔	29	9	9	11	43	53	36

第四节 统计的误用

用统计来支撑一个事实或结论，可以使结论更科学，或者对统计证据的收集和分
析使得事实更真实、更准确。但是，我们很可能常常误用统计学。

你可以从数据中分析出任何结果，有的有意义，有的无意义。(Carlyle)

世界上有三种谎言：谎言，邪恶的谎言和统计数字。(Disraeli)

做小说家不如做统计学家，统计学家能够展示更多、更丰富的想象力。(Mel
Calmen)

使用统计就好像一个醉汉在夜路上靠着灯柱——是用来依靠而不是用来照亮。

(Andrew Lang)

上面的这些引言都是对的，又都是错的。使用统计是一个解释的过程，而解释是一个基于知识进行判断的过程。例如什么是通货膨胀，在这个问题上至少对数字和计算结果的正确理解和计算方法和结果的准确性一样重要。如果根本不知道通货膨胀为何物，用准确的方法计算出一个准确的通货膨胀率就毫无意义。同时，对准确的统计信息也要有选择地使用。比如，我们想说明现在的通货膨胀率比较低，我们可以根据当月物价相对于上月物价的增长率 0.5%，说通货膨胀率是 0.5%。如果我们想说现在的通货膨胀比较高，我们可以根据 10 年来物价的平均增长率 18%，说通货膨胀率是 18%。

所以，也许更准确的说法是：“你不能根据统计证明任何问题。”数字是用来支持你的理论、看法，甚至偏见的。在上面的例子中，我们首先考虑的是说明的目的，想要说明通货膨胀率高还是低，然后我们才决定用哪种统计量来支持这个目的。

理想情况下，统计数据的收集和分析应该是尽可能客观公正和准确的。但实际操作中，统计常常被用来渲染、夸张、混淆或过分简化事实的，所以了解数据收集和分析的方法对你所做出的判断至关重要。

第五节 统计的使用

统计数据和统计分析结果最基本的功能是能够扫除人们在讨论中可能产生的争议，帮助我们做一个明智的决策。统计为争论和决定提供了一个事实的基础，从而能驱散传闻和谣言，调和不同的人因为从个人角度或个人情况出发而持有的不同观点。没有统计的事实，我们很可能片面地根据一些理论和看法就对传闻和谣言做出判断。因此，统计往往能被看成是一个维护稳定的力量。

祖母定律

祖母定律基于这样的理论：我的祖母（或祖父）每天抽 50 支烟（或者喝一瓶威士忌，从来不运动，吃一整只马等），而她（他）的寿命长达 105 岁；所以抽烟（喝酒、不运动、暴食）对人没有什么害处。在某种意义上，统计数据悖于“祖母定律”。统计数据并没有证据说每个大量吸烟的人的寿命都很短，但统计确实证明了大量吸烟的人，与吸烟量小或不吸烟的人相比，寿命缩短的可能性更大。

这就是统计信息与其他信息最大的不同点，它进行了系统的数据采集工作。所以统计能够提供系统科学的方法来概括复杂的经济和社会问题。

通常我们把统计分为两种：

● 描述统计，包括数据的图表描述和数据的分布特征测度（如计算百分比、平均数、离散程度和相关程度等），从而凸显数据的特征，把数据分类汇总到我们可以简便处理的程度。

● 推断统计，即从样本的某些特征出发推断总体的某些特征的方法。这些方法主要基于概率论。

事实上，很多知识在一定程度上都是一种概率知识。世间没有绝对确定的陈述，如果确定，也需要有一定的限制条件。比如，“我比我爸爸出生晚”、“黑猫是黑色的”和“一加一等于二”，这些都是赘述或虚伪的定义。即使我们承认它们是正确的，它们也只能提供有限的信息。大部分信息都来源于样本数据，而不是一个对全部总体进行的庞大的调查，正是概率论推进了统计的发展，从描述统计向推断统计发展（见第三章和第八章）。

统计工作的重心已经从对过去的回顾转移到对现状的分析和对未来的预测和运作上。几乎所有的商业机构和政府部门都在日常工作中大量使用统计。随着电脑技术的发展，近年来统计学在数据的收集、处理和存储方面都取得了突飞猛进的发展。英国作家赫伯特·乔治·威尔斯（H.G.Wells）曾经说：“统计思想终有一天将会与读书写字的能力一样。”也许这一天真的来了。

相当多的人在使用电脑和电子计算器，除了工作和学习的需要，日常生活中购物、纳税，甚至看体育新闻，他们也要使用这些工具。这些电子设备的使用，避免了处理数据时大量烦琐重复的计算工作。但是，当这些数据处理的过程变得自动化时，我们有必要去弄明白整个过程和方法。数字和事实不能照单全收，我们需要搞清楚数据是怎样收集的、怎么处理的，然后我们才能掂量这些结果的分量，看出它们的缺陷。数学基础知识见附录 A3。

第六节 第一手数据和第二手数据

统计数据有两种来源：一种是直接来源；另一种是间接来源，它们分别被称为第一手数据和第二手数据。

第一手数据指为当前的数据使用者专门收集的数据。收集工作包括调查设计、抽样方法设计等（见第三章）。数据经过收集、处理和公布之后就变成了第二手数据。

第二手数据是指他人收集的数据，不是当前使用者或为当前使用者的目的而收集的数据，它们是第二手的。对于大多数使用者来说，受财力、物力、人力的限制，亲自调查试验往往是不可能的。我们只能在总量中采集一些样本数据，而且每个人可能

都做过这个工作，但这种收集数据的方法也不一定是系统的、科学的。消费者决定购买物品前要对比几家商店的价格；公司收集产出、收入等数据来反映公司的经营状况。但是，对于更深入的研究可能有其他的要求，比如要获得大量的统计资料，要求系统的数据收集方法，这种情况下只能使用他人调查和试验的结果了。

因为第二手数据是经过一次传递的，我们要尽可能地掌握它的详细信息，就好像在买一辆二手车之前要尽可能掌握它的过去一样。了解数据的收集方法和处理过程比较重要，这可以帮助我们评价数据的可靠性和它的真正含义。表格可能会产生误导：比如一张表里显示，公司今年一个元件的产量与去年相比增长了 100%。这个结论看起来相当可喜，但如果去年仅生产了 5 个元件，今年生产了 10 个，100% 的意义就不大了。如果去年的产量是 5000 个，今年增长到 10000 个，100% 增长率的形容似乎才比较适当。

使用第二手数据关注以下几点比较重要：

- 数据是怎样收集的；
- 数据经过怎样的处理；
- 数据的准确性如何；
- 数据是何时公布的；
- 数据的度量口径是什么；
- 怎样解释数据，尤其是当使用的目的与原数据的使用目的不同时。

通常使用第二手数据都要在自己想要的资料和有限的资料来源之间权衡。但是，使用第二手数据的利大于弊，其主要优点有：

(1) 第二手数据的取得成本低。很多政府出版物包含着丰富的数据信息，而且售价非常便宜，图书馆也保存着大量的政府、企业和其他组织的数据。

(2) 大量第二手数据很容易获得——尽管有时候它们并不是我们最想要得到的数据。

(3) 各类学科和专业收集有各种形式、各种类别的数据。

(4) 第二手数据能够提供一个时期（几年或几十年）的数据，我们可以利用这些积累的历史数据来寻找趋势。

第二手数据记录对以下部门有重要的价值：

- 政府——用来帮助做出决策或制定将来的政策；
- 商业和工业领域——在诸如营销和销售工作中，用来了解大的社会环境和经济情况，同时掌握对手的信息；
- 研究机构——用来提供社会、经济和工业的信息。

公布的统计数据有的来源于生产管理过程的记录，有的来源于专门的询问或调

查。失业情况统计就是前者的一个很好的例子。这些数据来源于失业人员登记处，失业人员为了领取失业补助金都需要到管理部门进行登记，这样统计数据在管理的过程中就产生了，它们并不为其他的目的而收集。我们从这些登记资料中就能收集到我们需要的数据。

失业情况统计用来估计当前的失业人数，而不是准确的失业人员数的记录，因为很多没有工作或正在寻找新工作的人并没有到失业人员登记处登记。没有登记的失业人员包括不能领取失业养老金的人，如家庭主妇，还有自认为一定能找到工作而不愿去登记的人。同时，失业情况统计没有提供未充分就业情况的信息，还有很多人在工作，但没有全职的职位。所以，最好的情况下官方的失业情况统计提供的也仅仅是一个低估的社会失业人员数字。

来自民意测验表明，人口普查和家庭消费调查的统计数据是调查和询问的好例子，它们是专门设计来获取特殊信息的。英国全国范围的人口普查每十年举行一次，政府通过人口普查来掌握家庭部门的大量信息。这些信息能用来帮助政府进行公共管理决策，但请注意这并不是管理过程中产生的数据。

与管理过程的副产品数据相比，调查和询问更容易被用做第二手数据。这是因为调查是为专门的目的、为提供特殊的信息而收集数据，更便于我们根据自己的使用目的去引用。同时，关于调查和普查如何进行的资料也更详细、更容易获得。

第七节 第二手统计结果

根据第二手数据进行整理、分析得到的统计结果称为第二手统计结果。根据政府公布的数据计算得到的数据，比如失业率，就是第二手统计结果。计算失业率所依据的公开统计表就是第二手数据。

使用第二手统计结果和使用第二手数据一样，需要注意数据的含义、计算口径和计算方法等。因为第二手数据容易得到并广泛使用，第二手统计结果也随处可见。大量的第二手数据可以从政府统计资料、私人出版物和互联网络中轻易获得。在公司经营运作的过程中，为及时掌握公司的财务、人事等状况，公司会积累大量数据，同时，在公司向政府主管部门上报各项报表的过程中，这些数据也会作为政府数据来源的一部分。政府根据公司上报的数据，编著不同行业、不同地区的统计资料，这些又为公司与同行业、同地区的平均水平进行对比提供了基础。

工商业企业使用统计数据来简洁地描述公司的经营活动，通过实施抽样调查来分析企业的经营活动特征，预测未来的经营状况，这些都为企业的经营决策提供了一个科学的基础。