



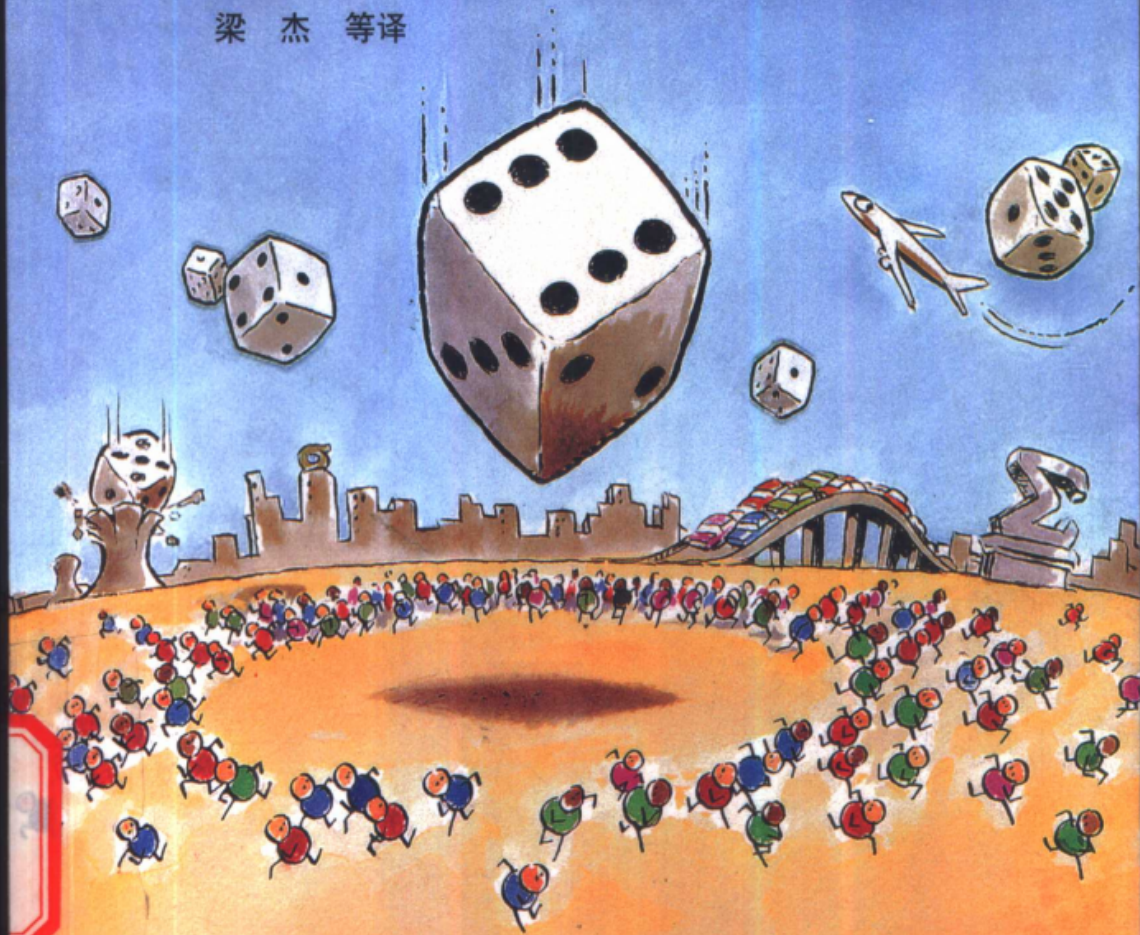
小恐龙科普漫画丛书

漫画统计学入门

THE CARTOON GUIDE TO STATISTICS

〔美〕拉里·戈尼克 沃尔科特·史密斯 著

梁杰 等译



辽宁教育出版社

你知道在点菜的时候需要用统计学的知识吗？

在风雨交加的夜晚能打到出租车的可能性有多大？

你买过彩票吗？

产品的不合格率如何计算？

《漫画统计学入门》一书不仅有你想要的答案，而且它还涵盖了现代统计学的所有精髓：数据的汇总、整理；随机变量；伯努利实验；中心极限定理；假设检验；估计置信区间；林林总总，所有这一切都在书中用简洁、明了的文字和妙趣横生的插图加以解释。

拉里·戈尼克是哈珀柯林斯公司已出版的四本漫画入门的作者与合作者，沃尔科特·史密斯是积极从事科研与咨询的知名统计学家，他们两人同时又是坦普尔大学的教授。

“戈尼克别树一帜。”——《探索》杂志



ISBN 7-5382-6216-4



9 787538 262162 >

ISBN 7-5382-6216-4/C · 226

定 价：15.00元

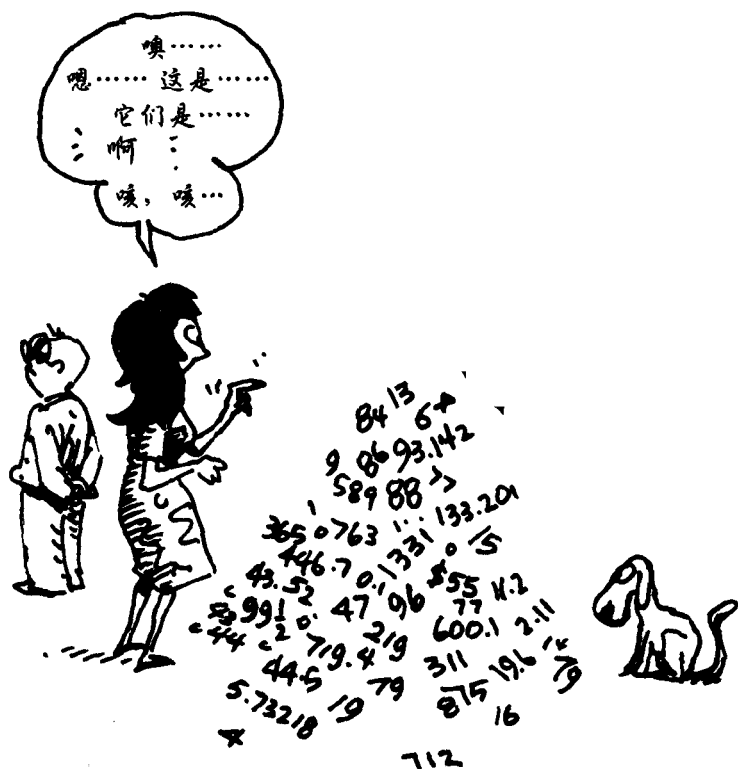
小恐龙科普漫画丛书

漫画统计学入门

THE CARTOON GUIDE TO STATISTICS

〔美〕拉里·戈尼克 沃尔科特·史密斯 著

梁杰 宋长来 李凤杰 孙奕驰 译



辽宁教育出版社

版权合同登记: 图字-06-2001-79号

图书在版编目(CIP)数据

漫画统计学入门/[美]戈尼克, 史密斯著; 梁杰等译. - 沈阳: 辽宁教育出版社, 2002.1

书名原文: The Cartoon Guide To Statistics

ISBN 7-5382-6216-4

I. 漫… II. ①戈… ②史… ③梁… III. 统计学-通俗读物 IV. C8-49

中国版本图书馆 CIP 数据核字(2001)第 091831 号

Cartoon Guide to Statistics

Copyright © 1993 by Larry Gonick and Woolcott Smith

Chinese (Simplified Characters only) Trade Paperback copyright © 2001 by Liaoning Education Press (Liaoning Jiao Yu Chu Ban She)

Published by arrangement with HarperCollins Publishers, Inc. (USA)
through Arts & Licensing International, Inc. USA

ALL RIGHTS RESERVED

辽宁教育出版社出版、发行

(沈阳市和平区十一纬路 25 号 邮政编码 110003)

沈阳六〇六所刷厂印刷

开本: 890 毫米 × 1240 毫米 1/32 字数: 60 千字 印张: 7 1/8

印数: 1—5,000 册

2002 年 1 月第 1 版

2002 年 1 月第 1 次印刷

责任编辑: 唐日松 吴晓红

责任校对: 王 玲

封面设计: 杜 江

版式设计: 赵怡轩

定 价: 15.00 元

目 录

第一章	什么是统计学	1
第二章	数据的描述	7
第三章	概率	27
第四章	随机变量	53
第五章	两种随机变量的分布	73
第六章	抽样	89
第七章	置信区间	111
第八章	假设检验	137
第九章	两个总体的比较	157
第十章	实验设计	181
第十一章	回归分析	187
第十二章	总结	211
参考书目		220
作者简介		223

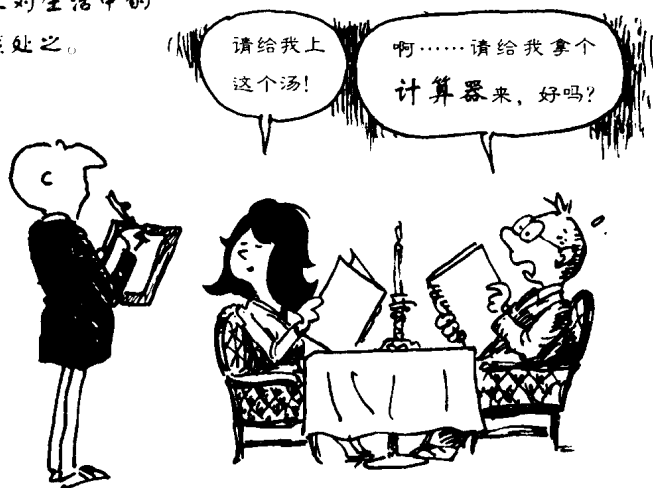
第一章

什么是统计学

在混日子的过程中，我们曾经根据不足的信息做出过选择……



我们多数人对生活中的
不确定性都泰然处之。

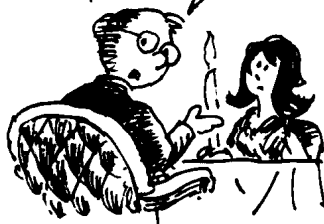


统计学的独特之处在于, 它能对不确定性进行量化, 使其精确。为此, 统计学家需要对不确定程度非常准确地进行明确的陈述。

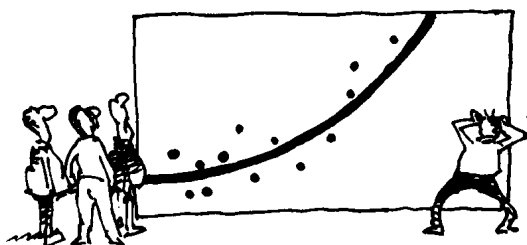


统计学不仅反映在点汤一类的小事上，它还涉及到生死攸关的大事……

嗨——以往在价格不优惠的晚上，你到这儿喝过这种汤吗？

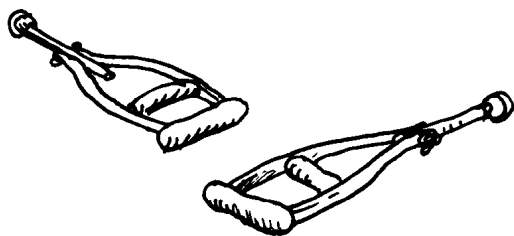


例如，1986年“挑战者”号宇宙飞船爆炸了，7名宇航员丧生。原因是他们甚至没有对低温操作的数据进行简单分析，就决定在29℃气温时发射了。



噢……曲线图的那部分！

另一个更有力的例子是接种小儿麻痹疫苗。1954年，对大约40万个孩子进行接种疫苗试验，严格控制，以消除偏差。试验结果的详细分析充分证实了疫苗的功效，当今小儿麻痹几乎绝迹。



为了取得他们的数学高招，统计学家求助于相互联系的三个学科的知识。

数据分析

收集、整理、汇总数据；

概率

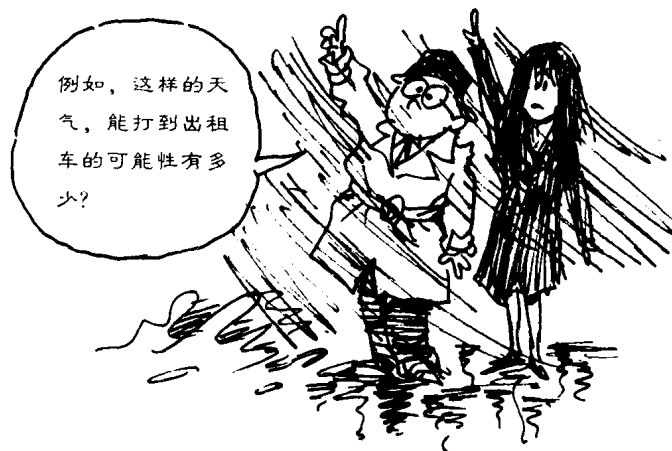
赌场内外的机会法则；

统计推断

运用概率知识从特定的数据中得出统计推论的科学。



在这本书中，我们将注意到，所有三方面的内容都应用到现代世界的各个领域，其中统计学起关键作用。



在第二章，我们将看到简单的数据集，是一组大学生的体重报告。

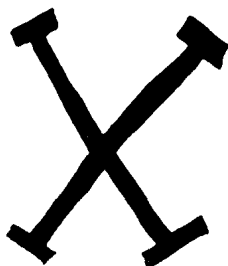


在第三章，我们在概率的发源地——赌场学习概率法则。

大约在 17 世纪
(误差为 $\pm 3\%$)



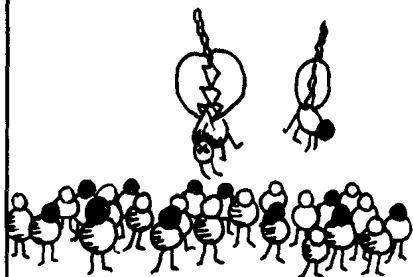
第四章和第五章介绍如何运用随机变量的概念，通过概率模型，描述客观世界。



哇！
一个符号！



在第六章，介绍统计学家必需的方法，在总体中选取样本。



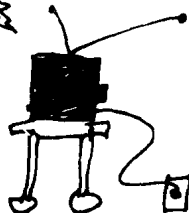
在第七章及以后的章节，我们描述如何在大千世界中进行统计推断，诸如投票时的民意测验、产品质量控制、医学检验、环境监控、种族歧视以及法律等。

总之，统计学
无处不在。

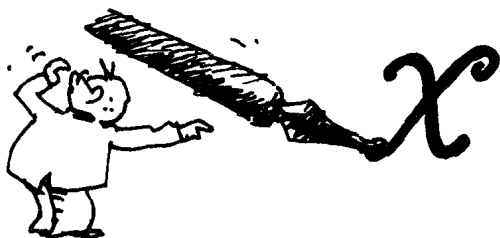


最后，在讨论统计学时，确实很难回避另一件事：在当今世界，对统计学普遍存在着怀疑。日常生活中几乎没有客观真实的统计分析，所以大家都说：“统计学与谎言同在。”我们应如何看待统计学？

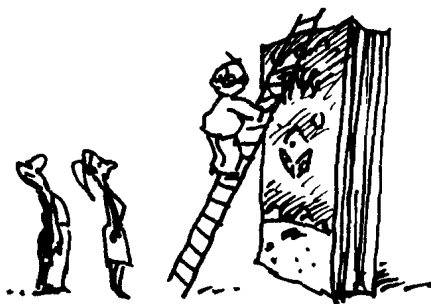
3/4 的医生忠告：不要相信任何以“3/4 的医生……”开头的报道。



我们私下认为，对这门功课涉猎一点不是坏事……这也是我们写这本书的初衷！



接下来，我们将尽量生动、直观地阐述统计学的基本内容。你只需要有一点耐心，一些推理能力和对代数的一定程度的忍耐——或许这就是对学一门课程的基本要求吧！！



第二章

数据的描述



在问题难倒我们之前，我们先把问题搞定，怎么样？

这里有一些真实数据：作为班级实验的一部分，92名宾州大学的学生报告了自己的体重，结果如下：



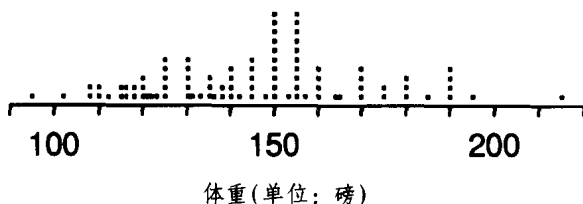
男生

140 145 160 190 155 165 150 190 195 138 160 155 153 145 170 175 175 170 180 135
170 157 130 185 190 155 170 155 215 150 145 155 155 150 155 150 180 160 135 160
130 155 150 148 155 150 140 180 190 145 150 164 140 142 136 123 155

女生

140 120 130 138 121 125 116 145 150 112 125 130 120 130 131 120 118 125 135 125
118 122 115 102 115 150 110 116 108 95 125 133 110 150 108

我们若归正信，直接绘制圆点图：一个点代表一名学生，与其自报的体重对应。



在这儿你会发现一个问题：体重150至155磅的学生很多。学生们倾向于以5磅的增值报告自己的体重。在此类情形的现实生活中，这样的舍零取整能使数据的一般模式模糊……但现在，也只能围绕这些数据工作。

我们能利用频数分布表汇总那些数据。将体重数列分成若干组，并计算每组组限。频数是分布在每组的学生人数，频率是各组学生的比率，即频数除以学生的总数。

组距	组中值	频数	频率
97.5-102.4	95	2	0.022
102.5-117.4	110	9	0.098
117.5-132.4	125	19	0.206
132.5-147.4	140	17	0.185
147.5-162.4	155	27	0.293
162.5-177.4	170	8	0.087
177.5-192.4	185	8	0.087
192.5-207.4	200	1	0.011
207.5-222.4	215	1	0.011

合计 92 1.000

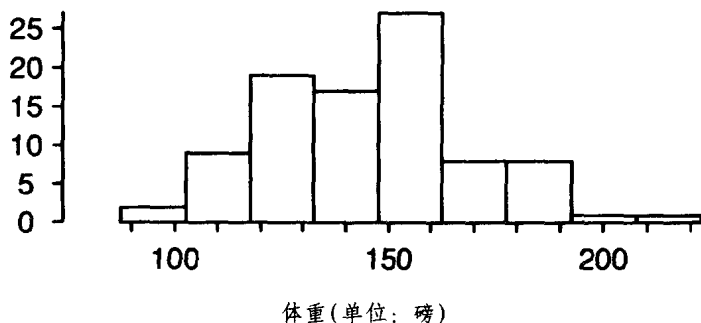
注意：每组组限都避开了讨厌的5磅的倍数，这样可抵消学生的报告偏差。

分组原则：

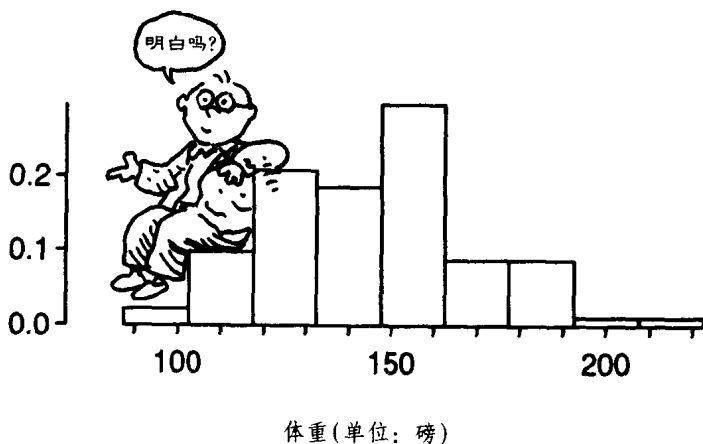
- 1) 采用等距分组，组中值选取合适的整数数值。
- 2) 对于较小的数据集，组数要小一些。
- 3) 对于较大的数据集，组数可多一些。



频数分布表显示每组“离”每个值有多少个数据圆点，我们也能用图形来表示该信息。这种条形图叫直方图。每个条形都覆盖以组中值为中心的区间。条形的高度表示每组中数据圆点的数量。



我们也能绘制频率的直方图，对照那个体重频数分布表标出频率，除了纵坐标外，频率直方图与频数直方图很相似。



统计学家约翰·托奇设计一种叫树图的灵巧方法，既汇总数据又保持特殊数据点。



对于体重数据，树干是一纵列体重值，由以 10 为单位的体重数据构成（即去掉了个位上的数字）

9
10
11
12
13
14
15
16
17
18
19
20
21



现在，把体重值的个位数字加在对应的行中。

树干：树叶

9 :
10 :
11 : 628
12 : 0155005
13 : 080015
14 : 05
15 : 0
16 :
17 :
18 :
19 :
20 :
21 :

意味着有 116 磅, 112 磅, 118 磅, 120 磅等体重值。



像这样填写

9 : 5
10 : 288
11 : 628855060
12 : 01553005525
13 : 8500850600153
14 : 05505580502
15 : 50537055055050500500
16 : 050004
17 : 055000
18 : 0500
19 : 00500
20 :
21 : 5

最后，将“叶子”排序。

9 : 5
10 : 288
11 : 002556688
12 : 0001235555
13 : 0000013555688
14 : 00002555558
15 : 0000000000355555555557
16 : 000045
17 : 000055
18 : 0005
19 : 00005
20 :
21 : 5



所有的 0 和 5 清楚地显示了学生的报告偏差。