

<<统计思维>>

图书基本信息

书名：<<统计思维>>

13位ISBN编号：9787115317377

10位ISBN编号：7115317372

出版时间：2013-5

出版时间：图灵教育

作者：Allen B.Downey

译者：张建锋,陈钢

版权说明：本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问：<http://www.tushu007.com>

<<统计思维>>

内容概要

## <<统计思维>>

### 作者简介

Allen B. Downey是富兰克林欧林工程学院的计算机科学副教授，曾执教于韦尔斯利学院、科尔比学院和加州大学伯克利分校。

他先后获麻省理工学院计算机科学硕士学位和加州大学伯克利分校计算机科学博士学位。

Downey已出版十余本技术书，内容涉及Java、Python、C++、概率统计等，深受专业读者喜爱。

他的最新Think系列书还有Think Complexity: Complexity Science and Computational Modeling、Think Python。

## &lt;&lt;统计思维&gt;&gt;

## 书籍目录

前言 xi 第1章 程序员的统计思维 1 1.1 第一个孩子出生晚吗 2 1.2 统计方法 3 1.3 全国家庭成长调查 4 1.4 表和记录 5 1.5 显著性 9 1.6 术语 10 第2章 描述性统计量 13 2.1 均值和平均值 13 2.2 方差 14 2.3 分布 15 2.4 直方图的表示 16 2.5 绘制直方图 17 2.6 表示概率质量函数 19 2.7 绘制概率质量函数 21 2.8 异常值 22 2.9 其他可视化方法 23 2.10 相对风险 24 2.11 条件概率 24 2.12 汇报结果 25 2.13 术语表 26 第3章 累积分布函数 29 3.1 选课人数之谜 29 3.2 PMF的不足 31 3.3 百分位数 33 3.4 累积分布函数 34 3.5 CDF的表示 36 3.6 回到调查数据 37 3.7 条件分布 38 3.8 随机数 39 3.9 汇总统计量小结 40 3.10 术语表 40 第4章 连续分布 43 4.1 指数分布 43 4.2 帕累托分布 47 4.3 正态分布 49 4.4 正态概率图 52 4.5 对数正态分布 54 4.6 为什么需要模型 57 4.7 生成随机数 58 4.8 术语 58 第5章 概率 61 5.1 概率法则 62 5.2 蒙提霍尔问题 65 5.3 庞加莱 67 5.4 其他概率法则 68 5.5 二项分布 69 5.6 连胜和手感 69 5.7 贝叶斯定理 72 5.8 术语 75 第6章 分布的运算 77 6.1 偏度 77 6.2 随机变量 79 6.3 概率密度函数 81 6.4 卷积 82 6.5 正态分布的性质 85 6.6 中心极限定理 86 6.7 分布函数之间的关系框架 88 6.8 术语表 89 第7章 假设检验 91 7.1 均值差异的检验 92 7.2 阈值的选择 94 7.3 效应的定义 96 7.4 解释统计检验结果 96 7.5 交叉验证 98 7.6 报道贝叶斯概率的结果 99 7.7 卡方检验 100 7.8 高效再抽样 102 7.9 功效 103 7.10 术语 104 第8章 估计 107 8.1 关于估计的游戏 107 8.2 方差估计 109 8.3 误差 110 8.4 指数分布 111 8.5 置信区间 111 8.6 贝叶斯估计 112 8.7 贝叶斯估计的实现 114 8.8 删失数据 116 8.9 火车头问题 117 8.10 术语 121 第9章 相关性 123 9.1 标准分数 123 9.2 协方差 124 9.3 相关性 125 9.4 用pyplot画散点图 127 9.5 斯皮尔曼秩相关 130 9.6 最小二乘拟合 132 9.7 拟合优度 135 9.8 相关性和因果关系 137 9.9 术语 139 作者及封面简介 141 索引 142

## &lt;&lt;统计思维&gt;&gt;

## 章节摘录

版权页：插图：习题3—1按照院长的方法构建这些数据的PMF，并计算均值。

因为数据是分组的，所以可以用每组的中点值。

然后再从学生的角度来构建选课人数的分布，并计算均值。

假设想要得到学校每门课程选课人数的分布情况，但又无法从院长那里得到可信的数据。

其中一种解决办法是随机选择一组学生，然后询问他们所选课程的上课人数。

然后可以根据调查的结果计算出PMF。

这个结果是有偏差的。

因为选修人数多的课程会被过采样，所以在估计选课人数真实分布时要对观察到的分布做一个合适的变换。

编写一个UnbiasPmf函数，参数是观察值的PMF，返回据此估计出的表示选课人数分布的Pmf对象。

答案可以从<http://thinkstats.com/class-size.PY>下载。

习题3—2 在大部分的田径比赛中，选手都是同时出发的。

如果跑得快，那么在比赛刚开始的时候会超过很多人，但在跑出几英里后你就会发现，周围都是跟你速度差不多的选手。

我第一次参加长跑（209英里）接力时，注意到一个奇怪的现象：当我超过其他选手时，我会跑得更快；当其他选手超过我时，他们通常也会跑得更快。

一开始，我觉得速度的分布是两级分化的：速度快和速度慢的人都很多，但跟我速度差不多的人应该不多。

但随后我发现我的选择是有偏差的。

这个比赛有两个特点：分阶段出发，不同的队伍出发时间也不同；此外，同一个队伍中选手的水平也参差不齐。

因此，选手在比赛道路上所处的位置与其速度和名次没有什么关系。

在我开始跑时，我周围的参赛选手基本上是随机的。

那这其中的偏差来自何处？

在整个比赛过程中，超过其他选手或者是被其他选手超过的概率跟选手间速度差异的大小是有关的。为什么？

想想最极端的情况。

如果我跟另外一个比赛选手的速度完全一样，那我们就不能超过对方，也不可能被对方超过。

如果某个选手跑得特别快，在我跑的过程中跑完了全程，那这位选手肯定会在某个地方超过我。

写一个BiasPmf函数，其参数是表示选手速度实际分布的Pmf和观察者的速度，返回值是一个新的Pmf，表示其他选手相对观察者的速度分布。

用一般的道路比赛（不是接力赛）的数据测试函数。

我写了一个程序读取马萨诸塞州Dedham的James Joyce Ramble一万米比赛的数据，并将每个选手的速度单位转换成m/h。

可以从<http://thinkstats.com/relay.PY>下载这个程序。

运行该程序，看看速度的PMF。

## <<统计思维>>

### 编辑推荐

大数据革命风起云涌。

数据分析成为每个浪尖上的舞者的必杀技。

而统计思维是数据分析和数据挖掘的根基。

每个程序员都应该具备统计思维，看到统计思维：程序员数学之概率统计你已经比别人先行一步。

这是一本极为独特的统计思维入门图书。

独特的编程视角。

对于主要的概率统计概念，作者都给出了开源的代码示例，其新颖独特的讲解方法绝对可以让程序员对概率统计产生更深刻的认识。

幽默风趣的示例。

你是否一直无法理解蒙提霍尔问题？

庞加莱是怎样发现面包商的企图的？

作者援引经典问题，帮你打开统计思维。

公共开源数据。

拿来美国全国家庭成长调查（NSFG）与行为风险因素监测系统（BRFSS）中的数据，重用参考代码，立即让自己的代码跑起来。

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:<http://www.tushu007.com>