统计学入门很简单

田霞 著

有统计思维的分析，更清晰易懂
有数据支持的结论，更有说服力

更贴近现实的统计案例
用 Excel 即可实现的统计分析
帮你形成统计思维，从数据中找到真相

STATISTICS

日常生活中的统计学

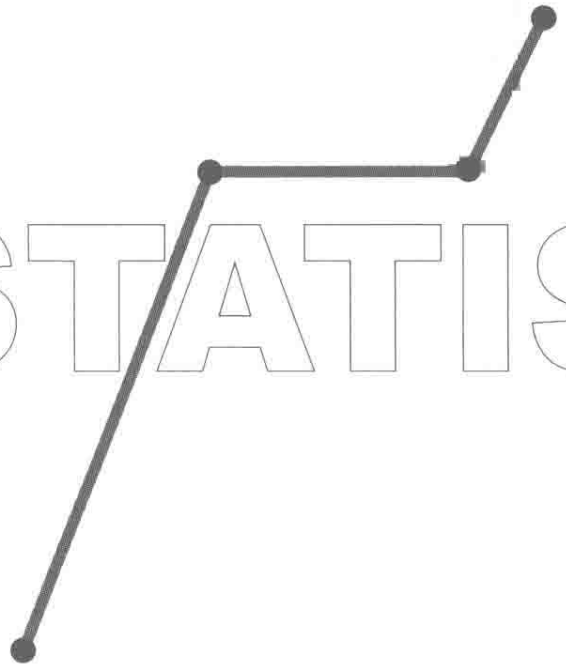
中国纺织出版社有限公司
国家一级出版社
全国百佳图书出版单位

统计学入门很简单

日常生活中的统计学

STATISTICS

田霞——著



中国纺织出版社有限公司

内 容 提 要

在生活中，我们经常会遇到需要进行数据分析的情景。无论是想要了解工资的平均值是否具有代表性，还是想要分析产品的质量是否稳定，或者想要研究影响销售的具体因素，数据分析都不可避免。而要进行分析，就需要对统计学的知识进行了解和学习。

本书分为三篇，分别介绍了概率的基础、统计的基础与统计的进阶，每章在介绍过预备知识后，都结合相应的统计知识给出了具体的案例，指导读者一步步进行运算，最后得出结论。除此之外，书中还介绍了如何运用Excel这一常见的软件进行简单的统计分析，让读者可以通过软件省略复杂的计算过程，在实际操作中更加得心应手。

图书在版编目（CIP）数据

统计学入门很简单：日常生活中的统计学 / 田霞著. --北京：中国纺织出版社有限公司，2023.10
ISBN 978-7-5229-0790-1

Ⅰ. ①统… Ⅱ. ①田… Ⅲ. ①统计学—基本知识 Ⅳ. ①C8

中国国家版本馆 CIP 数据核字（2023）第 141515 号

责任编辑：郝珊珊　　责任校对：高　涵　　责任印制：储志伟

中国纺织出版社有限公司出版发行
地址：北京市朝阳区百子湾东里 A407 号楼　邮政编码：100124
销售电话：010—67004422　传真：010—87155801
http://www.c-textilep.com
中国纺织出版社天猫旗舰店
官方微博 http://weibo.com/2119887771
天津千鹤文化传播有限公司印刷　各地新华书店经销
2023 年 10 月第 1 版第 1 次印刷
开本：880×1230　1/32　印张：7
字数：202 千字　定价：58.00 元

前言

我们生活中充满了各种数据。比如，在网上，我们会看到某城市或某省的平均工资，会看到他人购买商品后的评价等信息。某网站的点击量、在网页上的逗留时间……这些都是数据。那么，这些数据服从什么样的分布，有怎样的特点呢？在大数据时代，数据浩如烟海，如何从大量的数据中找出规律呢？这就需要统计的知识。统计和日常生活息息相关，我们可以使用统计的知识去理解和解决问题。本书通过案例的方式来讲述相关的统计知识。

第一篇“概率基础与正态分布”先研究了“先下手为强”这个案例，利用独立性解决了几个问题，然后研究品牌效应的问题，最后讲述了神奇的数字“37%”的来由。如果你有选择困难症，不妨看看。

第二篇“统计基础”首先给出了数据的分类，然后针对数理统计的“部分推断总体”的特点，为了使抽取的样本具有代表性，而不会出现幸存者偏差这类错误，给出四种抽样方法：简单随机抽样、分层抽样、整群抽样和等距抽样。不要小瞧抽样方法，它对做出正确的决策来说是至关重要的；如果数据选得不合适，给出的结论一般也是不正确的。美国历史上就发生过杂志社为了预测总统的选举结果进行民意调查，但是因为选取的样本不具有代表性，花费了大量的人力、物力，却做出错误的预测，导致杂志社关门停刊的事情。最后介绍了常见的统计量，如样本均值、中位数、极差、样本方差、标准差和四分位数等。图形可以

展示数据的形态特点，“描述性统计”部分还给出了箱线图、茎叶图和直方图的画法。

第三篇“统计进阶”包括参数估计、假设检验、非参数假设检验、方差分析和回归分析等内容。“参数估计”中给出基金收益率的矩估计、电动汽车续航里程的区间估计、语音输入鼠标的识别正确率的区间估计。“假设检验”部分首先使用第二次世界大战时期统计学家研究面包重量的问题介绍假设检验的步骤，指出两类错误，并利用 Z 检验和 t 检验解决该问题。然后使用 Z 检验解决纸箱用纸厚度、降糖药重量和紫外线杀菌灯的寿命是否满足标准等问题；利用 t 检验解决安眠药的治愈率问题、饲料养鸡问题、饮料的容量问题，以及主动吸烟和被动吸烟有无区别，若有哪个危害性更大等问题。利用卡方检验解决手机电池的寿命、机床的精度等问题。利用 F 检验解决主动吸烟和被动吸烟的区别、哪个牛奶厂的牛奶更好等问题。“非参数假设检验”部分介绍了三种检验：卡方拟合优度检验、列联表的独立性检验和秩和检验。使用独立性检验研究吸烟和肺癌的关系、色盲和性别的关系等问题。使用秩和检验解决母亲的吸烟量对新生儿体重的影响等问题。

作者从事概率论与数理统计方面的教学和科研工作长达 17 年，具有丰富的概率论与数理统计的教学经验。编写案例尽量做到由简到难、通俗易懂，既保证有趣，又保证实用。读完这些案例，相信读者可以学会使用统计的知识去理解和解决问题。同时在此感谢中国纺织出版社有限公司的郝珊珊编辑，非常感谢她在这本书的编写过程中给予的大力支持和帮助。

田霞

目录

第一篇　概率基础与正态分布

第二篇　统计基础

第三篇　统计进阶

第一章　参数估计

第二章　假设检验

第一篇

概率基础与正态分布

概率基础

第一章

预备知识

1. 独立性

若事件 A 和 B 没有关系，称为独立，用数学的语言描述为 $P(AB) = P(A)P(B)$。只要满足这个式子，则 A 和 B 独立。

若事件 A，B，C 独立，则三个事件至少有一个发生的概率为 $P(A \cup B \cup C) = 1 - [1 - P(A)][1 - P(B)][1 - P(C)]$。

2. 组合

从 n 个不同的元素中任取 m 个，不考虑次序问题，不同的取法有 C_n^m 种，其中 $C_n^m = \dfrac{n!}{m!(n-m)!}$。

3. 全概率公式

若① $A_1, A_2, \cdots, A_n$ 为样本空间 Ω 的一个分割，即满足 $A_1, A_2, \cdots, A_n$ 两两互不相容（没有交集）且 $A_1 \cup A_2 \cup \cdots \cup A_n = \Omega$。

② $P(A_i) > 0, i = 1, 2, \cdots, n$，则有 $P(B) = \sum_{i=1}^{n} P(A_i)P(B \mid A_i)$。

当 $n=2$ 时，有 $P(B) = P(A)P(B \mid A) + P(\bar{A})P(B \mid \bar{A})$。

4. 数学期望和方差

设离散型随机变量 X 的分布律为：$P(X = x_n) = p_n, n = 1, 2, \cdots$，如果级数 $\sum_{n=1}^{\infty} x_n p_n$ 绝对收敛，则称该级数为 X 的数学期望，记作 $E(X)$，即 $E(X) = \sum_{n=1}^{\infty} x_n p_n$。

设 X 是一个随机变量，若 $E[X - E(X)]^2$ 存在，则称 $E[X - E(X)]^2$ 为 X 的方差，记为 $D(X)$，即 $D(X) = E[X - E(X)]^2 = E(X^2) - [E(X)]^2$。

5. 二项分布

若随机变量 X 的分布律为：$P(X = k) = C_n^k p^k (1-p)^{n-k}$（$k = 0, 1, 2, \cdots, n$），则称 X 服从参数为 n，p 的二项分布，记为 $X \sim b(n, p)$。

二项分布的判定：先考虑一次试验时，结果只有两种——事件 A 发生和 A 不发生，A 发生的概率为 p，将该试验独立重复地进行 n 次，则 n 次试验中事件 A 发生的次数服从二项分布。

如果试验次数 $n = 1$，二项分布就是 $0 \sim 1$ 分布，其期望和方差为 p 和 $p(1-p)$。

6. 泊松分布

若随机变量 X 的可能取值为 $0, 1, 2, \cdots, k, \cdots$，且 $P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$（$\lambda > 0, k = 0, 1, 2, \cdots$），则称 X 服从参数为 λ 的泊松分布，记为 $X \sim P(\lambda)$，其期望和方差都为 λ。

泊松分布指的是单位时间内事件发生的次数。比如，某十字路口一小时内发生交通事故的次数。

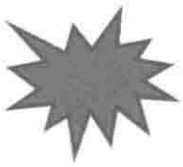
—01—
先下手为强

甲、乙两射手轮流对同一目标进行射击，谁先击中则得胜。每次射击，甲、乙命中目标的概率分别为0.4和0.6，甲先射，求甲得胜的概率。

关键词：独立性

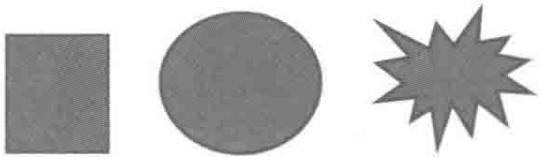
甲得胜，意味着甲有可能在第一轮中得胜，也可能在第二、第三、第四、第……轮中得胜。

若甲在第一轮中得胜，则甲第一次射击就击中目标，概率为0.4。



甲击中

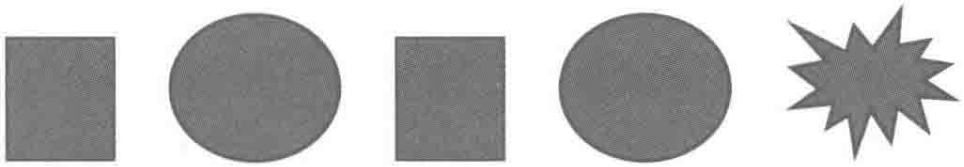
若甲在第二轮中得胜，则甲在第一轮中没有击中目标，乙也没有击中。甲在第二轮击中目标，所以甲得胜的概率为$0.6 \times 0.4 \times 0.4 = 0.24 \times 0.4 = 0.096$。



甲未击中　乙未击中　甲击中

若甲在第三轮中得胜，则甲在第一轮中没有击中目标，乙也

没有击中。甲在第二轮中没有击中目标，乙也没有击中目标。甲在第三轮中击中目标，所以甲得胜的概率为 $0.6 \times 0.4 \times 0.6 \times 0.4 \times 0.4 = 0.24^2 \times 0.4 = 0.02304$。



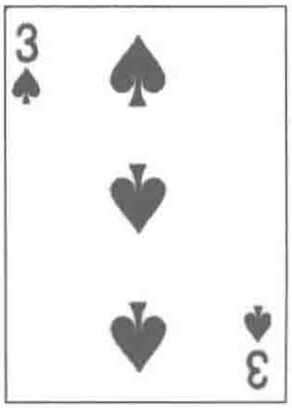
前三轮甲得胜的概率为 $0.4 + 0.096 + 0.02304 = 0.51904$。

若比赛继续进行下去，甲得胜的概率总会大于或等于 0.51904。

这个案例说明先手的重要性，尽管甲单次命中目标的概率为 0.4，但是因为他先射击，所以最后甲得胜的概率大于 0.5。如果想做什么事情，只要考虑清楚，有具体的思路，就要先下手为强，一直旁观犹豫是不可能取得胜利的。

—02—
什么样的扑克牌是独立的

从一副去掉大小王的扑克牌中任取 1 张，以 A 记事件“取到黑桃”，以 B 记事件“取到 3”，考虑 A，B 是否独立。



关键词：独立性

事件 A 表示取到黑桃，事件 B 表示取到的是

3，则 A 与 B 的交集 AB 表示取到的是黑桃3。那么这两个事件 A 和 B 有没有关系呢？先求事件 A 的概率。去掉大小王后的扑克牌有52张，其中黑桃有13张，所以取到黑桃的概率为13/52，即 $P(A)=13/52$。再考虑事件 B 的概率。总共52张扑克牌，印有数字3的牌共4张，则取到数字3的概率为4/52，即 $P(B)=4/52$。最后求事件 AB 的概率。黑桃3有1张，所以取到黑桃3的概率为1/52，即 $P(AB)=1/52$。

此时 AB 的概率等于 A 的概率与 B 的概率的乘积，即 $P(AB)=P(A)P(B)$，满足独立的定义，所以 A 与 B 独立，即事件 A 和 B 毫无关系。

—03—
破译密码

CTF（Capture The Flag）中文一般译作夺旗赛，在网络安全领域中指的是网络安全技术人员之间进行的一种技术比赛形式。CTF起源于1996年DEFCON全球黑客大会，以代替之前黑客们通过互相发起真实攻击进行技术比拼的方式，已经成为全球范围网络安全圈流行的竞赛形式。2013年，全球举办了超过五十场国际性CTF赛事。

有三位同学甲、乙、丙想参加CTF比赛，为了能在比赛中拿到名次，做些适当的练习是必要的。三个人找到一道古典密码的题目，题目为：

密文内容如下：{79 67 85 123 67 70 84 69 76 88 79 85 89 68 69 67 84 78 71 65 72 79 72 82 78 70 73 69 78 77 125 73 79 84 65}，

请将其解密。

因为是练习，三个人便约定分开研究该题目，约定时间为一小时，看谁能解密成功。

假设甲、乙、丙能破译该密码的概率分别为1/3，1/4，1/5，则这三位同学能破译密码的概率为多少？

关键词：独立性

三个人分开破译密码，每个人能否破译密码是独立的。密码能被破译，有可能是三个人中的一个成功破译密码，也有可能是三个人中的两个成功破译密码，也可能三个人都成功破译密码。总之，只要三个人中至少有一个能成功破译密码，则该密码就可以被破译。由预备知识中的独立性公式，三个人能破译密码的概率为 $1-(1-1/3)(1-1/4)(1-1/5)=3/5$。

三个人都参与破译密码，成功的概率要大于单个人成功的概率，这就是人多力量大。

—04—
逆向思维的重要性

口袋中有9个黑球、1个白球，每次从口袋中随机地摸出一球，并换入一个黑球，求第10次取到黑球的概率。