

概率与统计导论

(美) H·L·阿尔德 E·B·罗斯勒 著

北京大学出版社

概率与统计导论

〔美〕 H.L. 阿尔德 E.B. 罗斯勒 著

胡崇能 李隆章 校译

北京大学出版社

内 容 简 介

本书用大量的实例,阐述了概率、排列与组合、二项分布、正态分布、随机抽样、大样本理论、非参数统计、回归与相关、卡方分布、时间数列、方差分析等概率与统计基本问题,从而介绍目前西方各种主要统计方法和具体应用。对于从事统计和概率教学、科研的大专院校的师生和实际统计工作者有一定的参考价值。

Introduction to Probability and Statistics

Henry L. Alder

Edward B. Roessler

W. H. Freeman and Company

San Francisco, 1976

概率与统计导论

北京大学出版社出版

(北京大学校内)

新华书店北京发行所发行

1 2 0 1 工 厂 印 刷

787×1092毫米 32开本 15印张 310千字

1984年6月第一版 1984年6月第一次印刷

印数 1—32,000册

统一书号: 4209·16 定价: 1.55元

翻 译 说 明

前些时间，我们和另些同志集体翻译了美国阿尔德、罗斯勒的《概率与统计导论》(1976年第6版)一书。该书深入浅出，简明扼要，从相互联系中，对目前西方各种主要统计方法的性质和具体应用，作了比较全面的介绍；而所用数学不深，便于广大读者阅读。现将全书统一校订。我们水平有限，讹误难免；译词也感到与其他书难以完全一致，如 alternative hypothesis 译为“变换假设”，set 译为“组”或“集合”，未作严格要求等；附录中参考书目和索引略去未译，并记于此。统祈惠予指正为幸。

第六版序(摘译)

本书原系作为一学期用的概率与统计基础教程的教科书而写作的。第一版系教师们用过数年的油印和石印本的修订。这些班的成百上千的学生来自自然和社会科学的几乎所有专业，少数来自人文科学领域。在典型的美国学院或大学里，统计导论课程开在很多不同的系科，面向各种专业，要求数学基础的水平差别很大，这种现象也许导致一种错误印象，即在各种应用领域，基本统计是根本不同的。在这方面，统计教学似乎占有一定独特的地位；差不多每一其他学科，都有相当数量明确规定了内容的基础教程，任何需要该学科某些基础知识的人都要选读本学科的基础教程。比如，差不多每一学院或大学都开有化学基本教程，而所有需要学习化学某种知识的学生就选读该课。为什么统计不作同样处理，除了纯粹行政或传统的原因之外，似乎再无别的。的确，将基本统计的教学集中于一个系科，这在大多数学校，既可以是数学系，也可以是统计系，这样会大有好处的。如果只开这一课程，很多不必要的重复就能避免，从而就可以保持考核的唯一标准，本书就是为这一课程而设计的。

本教科书所要求的数学基础，只是相当于高中(2 年级)的代数知识。在第十二章有一处用到三角，但读者缺乏这方面的知识也可略去这一段，而无损于连贯性，由于要求的数学基础有限，故有些定理只好述而不证，但这里由于假定的数学基础有限而未证的定理相对不多，读者可为之一惊；并且，当他发

现读了一个或两个初等微积分教程后，对这些略去了的证明的理解，提高并不大，由此亦稍可自慰。为了避免与极限过程相联的数学困难，各种讨论都尽可能地限于有限的情形，特别是所有总体都假定为有限的。

学习本课程的学生来自农业、生物、医学、商业管理、经济学、家政学、心理学、社会学、地质等。为了照顾这些不同的专业领域，例子和习题就选自所有这一切领域。

习题是本教程一个不可缺少的部分。我们的经验是统计基础教程，跟数学的其他基础课程一样，教起来最有效是在每讲后布置一定数量的家庭作业。特别注意了保证习题的计算量很小，（除第十五章外）不需用机器计算，这就不得不包括进虚拟的资料，虽则很多练习中的资料也反映了实际实验的结果。

虽则读初等统计的大多数学生，是二年级生，但也有少数是新生。看来尽早读此课为佳。由于它并不比三角学困难，而据学生们反映，还更简单些，因此亦未尝不可在高中读此课程。早读此课就使学生能把他的知识用在很多现时应用了某种形式统计的课程中。及早讲授，除了别的明显好处外，统计学课程是复习高中代数的一个方便和有效的手段。因此，练习包括了涉及在两年高中代数中通常碰到的几乎所有项目的那些习题。——我看学生们将在解决实际有趣的问题之中，找到较之传统方法远为省力的复习他们的代数的方法。

一至十三章是使用统计的一切领域所要求的知识的基本核心。十四和十五章主要是为工商管理 and 学经济的学生准备的。

本版较第五版有四个大的改变。

1. 由于例子和习题中所使用数字一般反映实际资料，因此，只要可能，总以最新的可能得到的资料，使之合乎当前实际。

2. 有些章的习题增加了，而包括三到十三章的复习题列在新的附录中，这一套复习题与三到十三章的习题不同，是让读者在决定每一习题中使用三到十三章所学过的统计技术何者为适合，取得经验。

3. 有些章加了些材料，经验证明，这些材料是常常需要的或有用的，特别是，二项分布的讨论已予扩大，把这一分布处理为离散变量的概率分布的特例，同样地，把正态分布处理为连续变量分布的特例。在回归与相关一章，增加了一个得出回归直线的初级证明，并且增加了讨论指数曲线和幂曲线的一节，在卡方分布一章，增加了从列联表作结论的正确方法的讨论；在方差分析章增加了材料以弄清邓肯的新的多重极差检验法应予适用的条件。

4. 有少数改变是删去、增加或改变术语，以便符合现行统计实际和用语。

目 录

第一章 引 言	1
第二章 资料的整理	7
2-1 引 言	7
2-2 展示资料的表格方法和图形方法	7
2-3 频数分布及其表列与图示	14
第三章 总和记法	24
第四章 资料的分析	32
4-1 引 言	32
4-2 集中趋势的测度数	33
4-3 离中趋势测度数	45
第五章 基本概率、排列与组合	61
5-1 概率的定义	61
5-2 期 望	67
5-3 三个基本的概率定律	68
5-4 条件概率	78
5-5 组合与排列	82
5-6 重复试验	87
第六章 二项分布(和其它离散分布)	107
6-1 二项概率分布和理论频数分布	107
6-2 概率分布和一般理论频数分布	111
6-3 二项分布的平均数和标准差	114
6-4 二项分布的直方图	120

第七章 正态分布(与卜哇松分布)	123
7-1 正态分布作为二项分布的近似	123
7-2 卜哇松分布作为二项分布的近似	130
7-3 正态频数分布作为连续变量频数分布的极限	132
7-4 连续变量概率密度函数总论	134
7-5 正态曲线下各种特殊面积	136
7-6 按曲线分等级	137
第八章 随机抽样, 大样本理论	145
8-1 引言	145
8-2 样本平均数的分布	149
8-3 样本标准差的分布	153
8-4 两样本平均数之差的分布	154
8-5 标准误	158
第九章 假设检验, 显著水平、置信界限, 大样本方法	163
9-1 假设检验	163
9-2 置信界限	173
9-3 决定调查中样本的大小	176
第十章 学生氏 t-分布、小样本方法	184
10-1 总体标准差的估计	184
10-2 学生氏 t -分布	186
10-3 平均数的分布	187
10-4 平均数间差的分布	191
10-5 变值成对(配对)情形	195
第十一章 非参数统计	208
11-1 引言	208
11-2 对于非成对情形的维尔科克松两样 本检验法	208

11-3	对于成对情形的符号检验法	219
11-4	对于成对情形的维尔科克松检验法	222
第十二章	回归与相关	231
12-1	引 言	231
12-2	线性回归方程	233
12-3	估计的标准误	243
12-4	相关系数	244
12-5	相关系数的意义	250
12-6	变量改变时的线性回归方程	257
12-7	Y对X与X对Y的回归直线的比较	258
12-8	过原点的各回归直线	259
12-9	指数曲线和幂曲线	260
12-10	多项式的回归	261
12-11	多元回归	262
第十三章	卡方分布	269
13-1	定 义	269
13-2	卡方分布	270
13-3	在遗传学上的应用	272
13-4	在列联表上的应用	273
13-5	在正态性检验上的应用	279
13-6	校正的卡方	281
第十四章	指 数	294
14-1	引 言	294
14-2	基期的选择	294
14-3	简单指数	295
14-4	加权指数	297
第十五章	时间数列	302

15-1	引 言	302
15-2	长期趋势	304
15-3	季节变动	312
15-4	循环波动	322
第十六章	F-分布	330
16-1	定 义	330
16-2	两方差同质性的检验	333
第十七章	方差分析	339
17-1	一个分类标准的方差分析(方差的一个方向 分析), 引言	339
17-2	平方和的分割, 相等的样本大小	340
17-3	平方和的分割。样本大小不相等	346
17-4	方差分析表中方差的对比, 样本大小相等	348
17-5	最小显著差异	352
17-6	邓肯新的多重极差检验法	355
17-7	方差分析表中方差的对比, 样本大小不等	357
17-8	两样本的特殊情况	359
17-9	一个分类标准的方差分析中的假定	361
17-10	两个分类标准(两个方向的方差分析)。随机 完全区组设计	362
17-11	变 换	367
	复习题	380
	单号习题答案	398
附 录	404
	读物选(略)	
	表 I 正态概率曲线下的面积	406
	表 I a 正态概率曲线的纵坐标	409

表 II	卜哇松分布	412
表 III	学生氏 t -分布	414
表 IV	维尔科克松分布(非成对)	416
表 V	维尔科克松分布(成对)	416
表 VI	将 r 变换为 Z	420
表 VII	卡方分布	421
表 VIII	F -分布	423
表 IX	邓肯的新多重极差	430
表 X	百分数变换为反正弦 $\sqrt{\text{百分数}}$	433
表 XI	平方和平方根	439

第一章 引言

统计是一门关于数量资料的搜集、整理、分析和解释的科学。

“搜集资料”是取得量数或计数的过程。正确的结论只能来自适当搜集的或具有代表性的资料。这虽然是统计程序中的一个很重要部分，我们为了精简，将不予讨论，而将只对已取得的资料的处理，加以阐述。

“整理资料”是以适当形式展示已收集的量数或计数的工作，以便导出合乎逻辑的结论。关于用表格和图形整理和展示资料的一些有代表性的方法，将在第二章讨论。

“分析资料”是从已给的量数和计数中，抽出有关信息的过程，从而使一个概括的、全面的数量描述能以形成。供这个目的用的最重要的量数如平均数、中位数、极差、标准差以及其他，将在第四章论述。

“解释资料”是从分析资料中导出结论，并且一般涉及从搜集的小量对象所可能有的信息，对同类的大量对象作出预测。这项任务是本书所要讲的主要部分。

所以，统计是一门科学，它讨论在一定程度上能用数量信息，即由计数或测度得来的信息加以回答的问题。至于究竟是为生物研究而计数昆虫数，还是调查工厂的工人数或工时数，这都关系不大。统计工作者的责任，首先是选定所需信息的种类，其次是力求适当和有效地搜集与加工此项信息，最后是解释结果。在对资料作解释，特别是所根据的资料不完整时，统

计工作者必须应用能产生有效的结论的原理和技术。他往往被期待面对着不确定的情况，作出明智的判断。

统计一词有两种大不相同的含义。当用如前一段所指示的意义时，统计是用来研究和评价数量资料的一种科学程序。统计若作复数用，则是“数量资料”的同义语，例如说《世界年鉴》或《美国统计摘要》中有统计，这就是说其中有数量资料。这是这个词比较古老和普通的意义。最初，统计是为了政府首脑管理国家而收集的资料，这类用数字表示的信息可以上溯到亚里士多德和他的《国家事务》的文章时代。有事实可以证明“统计”和“国家”都是来自同一字源，从古老的时代起，最文明的国家已编集有大规模的“统计”，以弄清一国的人力和物力，供军事和财政的需要。在《圣经》中可读到许多这样的调查，在罗马帝国各领域内常有为税捐而汇编的簿册。

概率论的研究开始于意大利文艺复兴时代。其时，赌徒们力求找出掷骰子取胜的一套办法，乃商询这样一些学者，如格·卡丹诺(Giroleo Cardano) (1501—1576)、著名的数学天文学家伽利略·伽利力(Galileo Galilei) (1564—1642)；伽利略写了一篇短文，提出了概率论的基本原理，奠定了整个统计科学的基础。

十六、十七世纪时，在富人中赌博盛行，赌法复杂，赌注量大，从而在各种赌博中，如何能合理地算出取胜的机会，日益迫切。一位热中赌博的法国知识分子，司·莫瑞(Chevalier de Méré)商询名数学家与哲学家伯·巴斯克(Blaise Pascal) (1623—1662)，这促使巴斯克同他的几位数学朋友，尤其是同皮·费末特(Pierre de Fermat) (1601—1665)开始通讯。这项通讯形成为近代概率论与集合分析的起源。

其他积极研究概率法则的闻名数学家有高维·莱布尼兹

(Gottfried Wilhelm Leibniz)(1646—1716) 和翟克伯·贝努里(Jakob Bernoulli)(1654—1705)。翟克伯·贝努里是有名的贝努里家族九个数学家中的第一个。九人声望均高，而翟克伯，他的兄弟 J·贝努里(Johann Bernoulli)(1667—1748)，他的侄子尼古拉(Nikolaus Bernoulli)(1687—1759)、丹尼尔(Daniel Bernoulli)(1700—1782) 举世闻名。把概率论作为一个整体广泛加以阐述的第一部著作就是翟克伯写的，他阐明了大数定律原则。尼古拉把概率的概念应用到法律问题方面，丹尼尔则把概率微积分应用于传染病学和保险学研究方面。

与此同时，人口统计资料的收集与现在称为统计的整个知识的发展、都有重大进步。在英国，约翰·格兰特(John Graunt)(1620—1674)对生命统计、保险统计和经济统计进行了半数理性的研究。他的工作为威廉·配第(William Petty)(1623—1687)所继承光大，配第研究了伦敦人口的生命统计，象这样的工作以往还未曾有过；继起的还有埃德蒙·哈莱(Edmund Halley)(1656—1742)，他发展了死亡表，被誉为是生命统计科学的创始人。

阿伯拉罕·隶美弗(Abraham De Moivre)(1667—1754)阐明了复合事件概率的方法，从概率的原理导出了排列与组合的原理，创造了生命偶然性科学。1733年他发现了常态曲线方程，归纳统计理论很多方面都是以这个方程为根据。这同一钟形曲线也往往称为“拉普拉斯”曲线，“高斯”曲线，或“高斯—拉普拉斯曲线”，以纪念马奎斯·达·拉普拉斯(Marquis de Laplace)(1749—1827)和卡尔·弗里德里希·高斯(Karl Friedrich Gauss)(1777—1855)，因为他们二人也各自独立地发现了这个方程。高斯从重复测量一个数量误差的研究中导出了这个方程。他还首创了最小平方法，并发展了观察误差理论。拉

普拉斯对把统计应用于天文学作出了重大贡献，并且同阿瑞·玛利·勒根德(Adrien Marie Legendre)(1752—1833)一起，把偏微分方程应用在概率论的研究方面。“机误”一词首先出现在1815年弗里德里希·威廉·贝塞尔(Friedrich Wilhelm Bessel)(1784—1846)的著作中，他还推进了工具误差理论。

对这个理论有贡献的人还有詹姆士·斯特林(James Stirling)(1692—1770)，他发展了 $n!$ 的近似值；马奎·狄·康多塞(Marquis de Condorcet)(1743—1794)把概率与统计应用在社会问题方面；托马斯·贝叶斯(Thomas Bayes)(1702—1761)第一个从归纳上使用了概率；里昂海·尤拉(Leonhard Euler)(1707—1783)首先采用希腊字母“西格玛”作为总合的符号；托马斯·辛普生(Thomas Simpson)(1710—1760)将连续原则引进数理概率论。任·果·庸·阿莱伯特(Jean Le Rond d'Alembert)(1717—1783)在研究概率中运用了气象资料；约赛夫·鲁意斯·拉格朗日(Joseph Louis Lagrange)(1736—1813)应用了微积分；培瑞·雷蒙·狄蒙特莫特(Pierre Rémond de Montmort)(1678—1719)引进了差分演算。孔德·狄·蒲丰(Comte de Buffon)(1707—1788)预见到近代遗传学和概率微积分的若干方面；西蒙·旦尼斯·普哇松(Siméon Denis Poisson)(1781—1840)创立了以他的名字命名的分布。

在1835年与1870年间，比利时科学家兰伯特·凯特勒(Lambert A.J. Quetelet)(1796—1874)在发展与运用概率与统计方面作了重大贡献。他表明了生物学与人类学的量数紧密地遵循常态曲线。他不仅将统计方法应用于生物学，也应用于教育学与社会学。他表现出非常广泛的兴趣，被誉为是认识大数守恒的第一个人，也是表明在一个研究领域发展起来的统计技术可应用于其他大部分领域的创始人之一。

在德国，乔治·弗里德里希·克拉普 (George Friedrich Knapp) (1842—1926) 根据凯特来的原则对死亡统计作了广泛的研究，威·来克西司 (Wilhelm Lexis) (1837—1914) 发展了一种今天称为方差的单向分析的方法。

在十九世纪最后四分之一的时期中，弗兰西斯·高尔顿爵士 (Francis Galton) (1822—1911)，英国优生学派的创始人，他以无限的热忱，对他所能积累足够资料的每一生物变量的系统性变异的原理，进行了验证。生物变异有序原理的揭示，构成为生物学研究一个新时代的开端。高尔顿同他的伟大继承人卡尔·皮尔森 (Karl Pearson) (1857—1936) 使用遗传学中的问题，发展了回归与相关的概念。此后，皮尔森同查理士·埃德华·司皮尔曼 (Charles Edward Spearman) (1863—1945) 推广了这个理论、把它应用于社会科学的研究中。皮尔森还广泛地研究了抽样误差的影响，发展了卡方检验法，并在文献中引进了“平均差”和“标准差”两名词。

本世纪初期，威廉·司来·哥司特 (William Sealy Gosset) (1876—1937)，爱尔兰贵尼斯酿酒厂的一位统计工作者，他以笔名“学生”发表了许多篇关于解释抽样得来资料的文章。他第一个认识到，阐述从小样本抽取可靠资料方法、有其重要性。他的方法不久就在英国为朗奈德·阿·费暄 (Ronald A. Fisher) (1890—1962) 及其同事们所推广。他们对科学，尤其是人口遗传学，作出了很多贡献，大大地扩展了实验理论，提高了统计方法及其在各个研究领域中应用的兴趣。引进现在广为流行的“零假设”这一名词和发展方差分析的统计技术的，就是费暄。

在二十世纪中，杰出的统计学者不胜枚举，他们都在发展新理论和应用方面，起了积极作用。电子计算机的应用大大地促进了这些发展。当今的研究人员无不视统计为最有用的工具