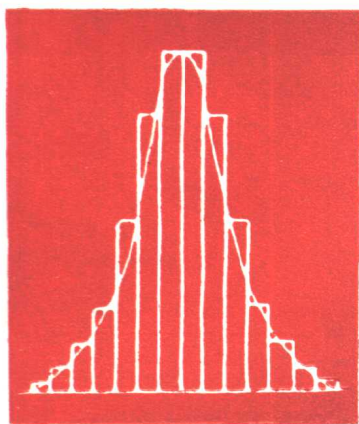
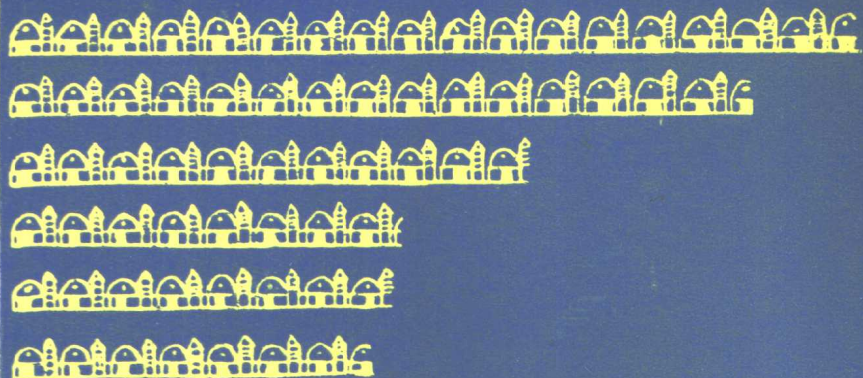
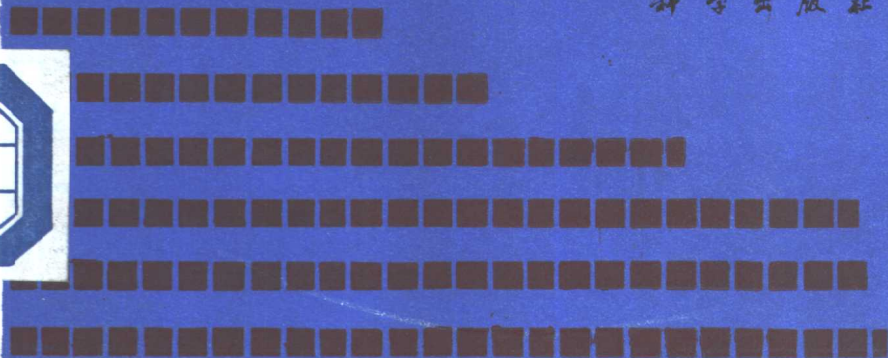




现代生活中的统计学 和概率论

韩冠堂 编

科学出版社



现代生活中的统计学 和概率论

韩冠堂 编

科学出版社

1987

内 容 简 介

概率统计不仅是近代自然科学的基础和有力的分析手段，而且也是现代企业质量管理的基础理论，采用概率统计方法进行设计、实验等所取得的经济效益已为实践所公认。

本书是一本入门书，它把概率统计与现代生活和科学技术紧密联系起来。书中所介绍的理论简明易懂，重点在于应用，密切联系实际，并以丰富的实例介绍了概率及其分布的计算和常用的统计检验方法。每章有适当练习题（书末附有答案）。

本书可供企业管理人员、领导干部，工业、企业科技人员以及有关专业学校的师生等参考阅读。

现代生活中的统计学和概率论

韩冠堂 编

责任编辑 徐一帆

科学出版社出版

北京朝阳门内大街137号

中国科学院印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

1987年8月第 一 版 开本：787×1092 1/32

1987年8月第一次印刷 印张：10 3/4

印数：0001—7,500 字数：244,000

统一书号：13031·3580

本社书号：4870·13-1

定价：2.25元

编者的话

概率论与数理统计是一门历史悠久但又新枝丛生的学科，它的主要原理和基本方法在现代生活和科学技术中已得到广泛的应用。其范围不断扩展、渗透日益深入，以致我们每个人的生活和工作都离不开它。尽管许多人还没有认识到这一点，但总是在受着它的支配和影响。

今天，希望学习和了解概率与统计的人数在与日俱增。而且，随着时代的前进，这将从主观要求变成客观需要。正象十九世纪的预言家 H. G. 威尔斯所预见的：“统计的思想有朝一日将象读和写的能力一样，是每一个工作的公民所必需的。”

现在，我们已经看到，概率统计不仅是近代自然科学的基石和有力的分析手段，而且成为现代企业中质量管理的基础理论。同时，用概率统计方法进行设计、试验以及配备能源等，往往会带来很大的经济效益，这一点已为许多人所重视，并取得明显的实际效果。一个盲目的试验者，虽然很辛苦，却不自觉地在那里浪费时间和物力。但若用正交表等科学方法进行试验设计则会收到事半功倍之效。今天，应用统计学本身已发展成为一个有趣的新领域。

正是基于上述考虑，编者想把概率统计与现代生活和科学技术联系起来，因而编写了这本书。本书是概率统计方面的一本入门书。理论力求简明，重点在于应用，密切联系实际，以使读者掌握概率及分布的计算和常用的统计检验方法，这是本书的基本宗旨。因此，对于一些定理和推论，略

AYA-2114

去了严格的证明，但指出了参考的书目。在数学基础方面只要求读者具有微积分的知识，如若对矩阵有些了解，则更令人满意。

书中选取的大量例题，除极少数是与概率的产生有关的经典例子外，其余绝大部分都来自现实生活，以期引起读者进一步研究的兴趣。每章末尾都安排有练习题，用以巩固所学的内容，初学者亦可用之检查自己实际掌握的程度。

以上所述，仅仅是编者的愿望。但由于水平所限，书中一定会有许多不妥之处，错误亦在所难免，恳请读者提出宝贵意见。

本书在编写过程中得到魏淑秋、陈伟侯等同志的热情帮助。他们审阅了全书，并提出了许多宝贵意见，谨此一并表示衷心的感谢。

编者

1985年5月

目 录

编者的话

第一章 引论	(1)
1.1 概率论与数理统计的研究对象	(1)
1.2 现代生活中的统计学和概率论	(1)
1.3 几个基本概念	(4)
练习 1	(7)
第二章 样本数据的描述	(8)
2.1 频数分布	(8)
2.2 频率分布	(13)
2.3 频率及其分布的随机性	(14)
2.4 其它的图示技巧	(18)
练习 2	(24)
第三章 概率及概率分布	(29)
3.1 概率的定义及计算	(29)
3.2 随机变量的概率分布	(48)
练习 3	(61)
第四章 统计特征数及各种分布型式	(66)
4.1 统计特征数的意义	(66)
4.2 位置特征数	(68)
4.3 离散特征数	(75)
4.4 其它特征数	(80)
4.5 总体特征数	(81)
4.6 统计量的分布及其特征数	(94)
4.7 常见分布型式及其关系	(99)

4.8 大样本的分布——两个重要定理	(102)
练习 4	(113)
第五章 统计推断	(118)
5.1 参数估计	(118)
5.2 统计假设检验概述	(136)
5.3 双侧检验与单侧检验	(138)
5.4 统计结论的两类错误	(140)
练习 5	(143)
第六章 分布参数的统计检验	(145)
6.1 单总体或双总体的参数检验	(145)
6.2 多总体的平均值的检验	(154)
练习 6	(167)
第七章 分布型式的检验	(171)
7.1 χ^2 适度检验	(171)
7.2 K 检验(柯氏准则)	(176)
练习 7	(179)
第八章 秩和检验法——非参数统计检验	(182)
8.1 秩和检验法的应用范围	(182)
8.2 秩和检验法的基本思路	(183)
8.3 秩和检验的内容及步骤	(186)
8.4 离散变量的秩和检验	(192)
8.5 秩和检验法的可靠性	(195)
练习 8	(197)
第九章 多元随机变量的相关及回归分析	(199)
9.1 相关关系	(199)
9.2 一元线性回归分析	(202)
9.3 相关系数及其检验	(204)
9.4 回归效果的鉴别	(209)
9.5 多元线性回归方程	(217)
9.6 三个重要系数及其应用	(224)

9.7 非线性回归分析	(231)
练习9	(235)
第十章 试验设计及抽样方法简介	(240)
10.1 试验设计	(240)
10.2 抽样方法简介	(260)
练习10	(273)
练习答案	(276)
附表	(321)
参考书目	(335)

第一章 引论

1.1 概率论与数理统计的研究对象

概率论所研究的是随机现象*的数量规律，它与描述所谓决定性现象**的数学相比，有其本身的特点。概率论的中心课题就是要给“可能性”以确切的描述，并给出科学的估计方法。它是一门从数量的角度研究随机现象内部隐藏的必然规律的学科。

数理统计是一门年轻而又有趣的学科，它为科研工作者提供所必需的科学方法，利用这些方法去收集数据，并用来确定所需的数据量，通过样本进行推断，并且量度它们的不确定性，从而做出有意义的决定。

在近代，概率论与数理统计的应用领域不断扩展，并出现了一些相应的边缘学科。并且数理统计已成为现代企业管理的重要理论基础。图 1.1 仅列出了它们的主要应用，而其影响范围比这要广泛得多。

1.2 现代生活中的统计学和概率论

概率论和统计学在我们的生活中到处起着重要的作用，我们许多人虽然不熟悉它，甚至没有意识到这一点，但往往

• 随机现象：在一定的条件下不能预先断言出现哪种（这种或那种）结果的现象。

•• 决定性现象：在一定的条件下必然发生某一事件的現象。

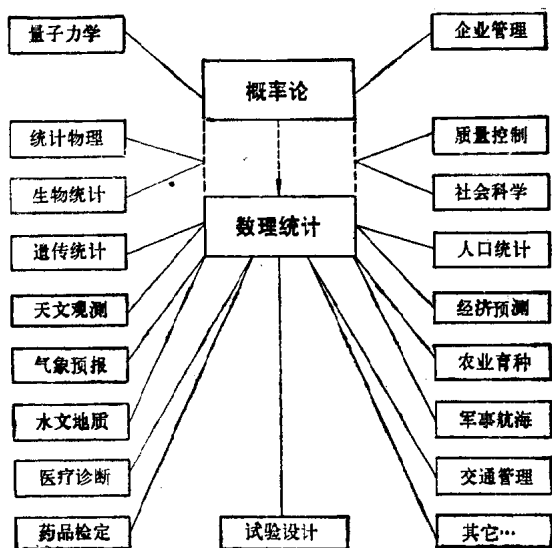


图 1.1

都是不自觉地利用了它而做出某些决定的。

概率论与统计学不仅可应用于社会学和人口问题等方面的研究，而且也被应用于自然科学方面。如物理学中的分子运动理论、放射性原子的平均寿命以及工业生产中产品质量的控制等等。概率与统计之所以能如此广泛地应用于各个方面，是因为客观世界中普遍地存在着一些现象，它们似乎有偶然性，但我们不能用因果关系很明确的简单定律加以概括，而必须通过大量观测后加以综合分析，归纳出“大量现象”的规律来。下面举一些例子说明这一学科的广泛应用。

(1) 由某妇产医院的综合分析表明：对孕妇使用某种麻醉药品，胎儿就可能有生命危险。

(2) 由人口寿命统计资料表明：平均来说，当代人的平均寿命比他们的父母要长些。

(3) 由世界人口流动情况的统计数据表明：过去美国的移民中，大多数来自西欧和加拿大，现在的新移民中绝大多数来自亚洲和拉丁美洲。这些统计对于经济学家和社会学家是很有用的。

(4) 对于准确地预报江河的汛期和水位，粗看似乎很难预测，但如果分析了历史上长期的水文观测资料，就可以掌握统计规律，做出较准确的预报。这对于水利工程建设具有重要的意义。

(5) 在医学上，由于人体之间的差异，同一种药物或治疗方法对不同人可能产生不同的效果，因此需要经过大量的临床实践，找出药物或治疗方法的统计规律，才能获得比较好的疗效。

(6) 地震的发生也是一种随机现象，为了预报地震，就需要从历史资料中探求各种条件与发生地震之间的联系，以及这些联系的统计规律。

(7) 一批产品中有正品，也有次品，我们随意取一件来检验，就有“正品”或“次品”这两种可能，这是带有偶然性的。但是，大批产品中次品的多少是符合一定的统计规律的，怎样根据这些统计规律，合理确定抽样检验方式，进行产品质量控制，这是生产实际中经常遇到的问题。

下面的两个例子说明，统计资料有可能被滥用，以致做出错误的决定，应引起足够重视。

(1) 在某城市中过去没有小儿麻痹症的病例，但1975年出现了2例，1976年又有4例。若由此得出：1975—1976年期间，该病发病率增加100%。这样的统计结果会使每一个做父母的感到震惊！然而经过细致的分析就会发现，这样的病例数是在500万居民中调查的结果，因此，相对值是很小的。

(2) 在电视广告节目中, 某一外国公司声称: “过去 11 年中, 该公司销售的汽车, 10 辆中有 9 辆仍在道路上行驶”。该厂历年的汽车销售量如图 1.2 所示。你认为应该如何评价这种宣传呢? 注意: 该公司最近两年的销售量较大。你认为广告中的宣传能说明该公司的产品质量好、寿命长吗?

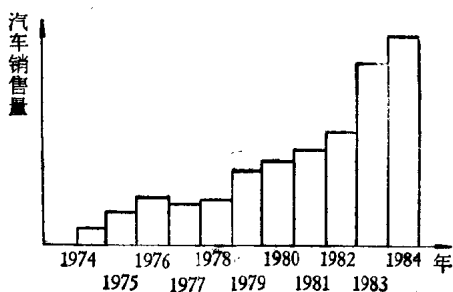


图 1.2

1.3 几个基本概念

一、随机事件与随机试验

如前所述, 象纯水在一个大气压下加热到 100°C 必然沸腾, 这类现象称为决定性现象; 而投掷一枚硬币时, 可能正面向上, 也可能反面向上, 而且不能预先断定哪面向上, 这类现象称为随机现象。

(1) 随机事件 随机现象中的事件在一定条件下可能发生, 也可能不发生, 故称之为这个条件下的随机事件。例如, “正面向上”是“投掷一枚硬币”这个条件下的随机事

件。

(2) 随机试验 观察一定条件下发生的现象通常称为试验，如果一个试验。

(a) 可以在相同的条件下重复进行；

(b) 每次试验可以出现不同的结果，但不能预先断定出现哪一种结果；

(c) 一次试验中必有且仅有“一切可能结果”中的一个结果出现。

我们则将满足上述条件的试验称之为随机试验。

很明显，随机试验观察的对象实际上就是一个随机现象。为了简单起见，今后我们也常常将随机事件称为事件，将随机试验简称为试验。

随机事件在一次试验中，可能发生也可能不发生，有其偶然性，但在多次重复试验中却会呈现出明显的规律性。

二、总体与样本

(1) 总体 数理统计中把欲观测的、满足指定条件的全部个体的集合称为总体。若总体包括无限个个体时，称为无限总体；若总体只包括有限个个体时，则称为有限总体。

(2) 样本 从一个总体中抽取 n 个个体，这 n 个个体则称为该总体的一个容量为 n 的样本。

例如，从一批总数为1000件的零件毛坯中，取出20件零件毛坯称其重量，其数据如下（单位为公斤）：

108	113	109	96	103	104	107	109	102	100
93	92	101	108	98	106	103	101	104	105

这里，总体包括 $N=1000$ 件毛坯，是一个有限总体。样本包括20件毛坯重量，故其容量 $n=20$ 。其中每一零件毛坯重量

都是个体。

如果我们要研究总体的特性，我们不可能花费大量时间去测定每一个体，那么如何利用这一样本中的 20 个数据得到所需要的总体信息呢？这就要用数理统计方法才能获得正确的结论。

如果读者具有集合论的基本知识，可将上述概念与集合中有关术语相对照，就能够更好地理解它们。

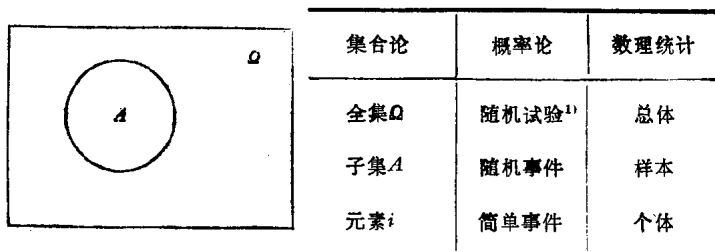


图 1.3

1) 这里指随机试验的所有可能结果，亦称样本空间，详见第三章。

三、描述统计与统计推理

(1) 描述统计 将所收集的数据通过一定的方法列成图表，并需运用某种技巧进行简缩才能得出有意义的统计结论。

(2) 统计推理 根据某一样本进行估值或进行推断的程序，这是数理统计学的主要理论部分，其内容主要包括参数估计与统计假设检验两方面，今后将分别予以讨论（详见第五·一九章）。

练 习 1

1. 甲乙两个歌舞团在一年中分别举行了三次音乐会，甲团售出的票数分别是 1580、940 和 1620 张；乙团售出的票数分别为 1160、1420 和 1340 张。下列结论中哪一个需要统计推理，哪一个只需描述统计呢？

(1) 平均来说，甲团比乙团更吸引人；

(2) 两团音乐会的听众总数之间有相当大的差别；

(3) 在以后任何一次音乐会上，甲团将比乙团吸引更多的听众。

2. 仔细地阅读一份报刊，看看它所用的统计是描述的，还是推理的。

3. 如果你有兴趣想确定最终会比他父亲高的男孩所占的百分数，你应该怎样选择样本呢？你能够选择你的朋友和亲属作为一个随机样本吗？

4. 某市关于交通情况的年度报告中指出：有 234 次交通事故与饮酒后开车的司机有关；有 15763 次交通事故涉及到饮酒后的行人。我们是否能由此得出结论：在该城市，饮酒后开车的司机比酒后的行人更安全呢？

第二章 样本数据的描述

如前所述，数理统计工作总是从收集资料开始的，但由生产和实验过程中收集到的资料和数据往往是分散的，而且从表面上看不出有什么规律性，也不能说明任何问题。必须经过整理和归纳后，这一批数据所遵循的规律才能显露出来，方可得出有意义的统计结论。本章的内容主要是讨论对于资料和数据怎样进行整理和归纳，这是统计工作的第一步，也是重要的基本技巧，因此必须很好地掌握并能熟练地应用。

下面以日常生活中的实际问题为例，说明描述统计中常用的两种处理方法：列表法与图示法。此外，频率曲线和频率函数是数理统计学中的重要概念，又是实际中应用最多的，因此，对其性质也必须充分了解。

2.1 频数分布

我们通过一些具体例子来说明列表和图示的方法。

〔例 2.1〕服装职业学校共有 9000 名学员，为了确定这类学校中女学员的平均年龄，市教育局决定抽选 150 人作为一个样本，然后利用由这个样本所得到的认识，做出关于所有女学员平均年龄的推理。

在选取样本中最重要的是，它必须是一个随机样本。这意味着，总体中的每一个体（即该类学校的每个女学员）被抽选的机会都是同样可能的。如果不满足这个条件，人们就不能根据这个样本做出关于总体的正确的推理（关于随机抽

样的问题可参见第 10.2 节)。

对于所抽选的 150 人的年龄按调查的顺序列出, 如表 2.1 所示。

表 2.1 150 名女学员的年龄

20	17	24	18	16	23	21	17	21	27	23	17	16	19	14
23	26	38	33	20	33	26	26	26	33	26	35	28	35	24
21	18	21	23	22	21	19	23	19	18	19	15	23	21	18
27	40	27	26	27	25	27	25	25	25	30	19	26	32	22
24	19	22	19	24	15	20	22	23	26	23	18	21	20	21
34	41	35	20	29	20	29	27	29	32	29	29	32	28	31
19	22	23	25	23	23	18	19	24	24	21	20	24	22	20
30	31	39	44	38	37	30	37	33	30	36	34	36	32	26
25	23	17	24	18	24	24	21	16	20	18	22	25	24	17
28	29	34	31	25	34	25	36	28	27	31	27	28	30	28

由表 2.1 我们只能看出: 大多数女学员的年龄是 20 多岁, 最小的一个是 14 岁, 即 $x_{\min}=14$; 最大的一个是 44 岁, 即 $x_{\max}=44$, 但得不出任何结论。

显然, 为了有所发现, 必须重新组织数据。我们采用一种方便的办法, 即按从小到大的顺序进行排列, 并统计每一年龄出现的次数 (称频数), 如表 2.2 所示。

表 2.2 年龄与频数

年 龄	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
频 数	1	2	3	5	8	9	9	10	10	12	10	9	8	7	6	6
年 龄	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	
频 数	5	4	4	4	4	3	3	2	2	1	1	1	0	0	1	