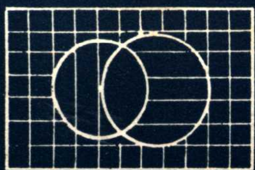

中学教学参考丛书

概率统计初步

山东教育出版社

中学教学参考丛书

概率统计初步

吴天滨 编

山东教育出版社

一九八三年·济南

中学教学参考丛书
概率统计初步
吴天滨 编
*
山东教育出版社出版
山东省新华书店发行
山东新华印刷厂印刷
*
787×1092毫米32开本 7.375印张 148千字
1983年1月新1版 1983年1月第1次印刷
印数：1—4,000
书号 7275·126 定价 0.59 元

前　言

概率统计是现代数学的一个重要分支，是一门研究随机现象数量规律的科学。它在自然科学、技术科学、国民经济各部门以及国防建设等许多方面都有着广泛的应用。

本书前五章基本上是参照全日制十年制学校《中学数学教学大纲》（试行草案）的内容和顺序编写的，第六章简略介绍了随机变数和概率统计应用的初步知识。书中凡超出大纲的内容，均以“*”号标明。

在编写中，力求做到文字通俗易懂，结合实例讲清重点和难点，并编入了较多的例题和练习题，书后附有练习题解答。本书除可供中学数学教师教学参考之外，也可供初学者自学。

编　者

一九七九年九月

目　　录

第一章　统计初步……………………………………………1
一　引言 ……………………………………………………1
二　总体、个体和样本 ……………………………………4
练习题一
三　频率分布 ……………………………………………13
练习题二
四　样本的均值 …………………………………………24
练习题三
五　样本方差和样本标准差………………………………31
练习题四
六　小结…………………………………………………42
第二章　事件及其概率……………………………………43
一　试验和事件 …………………………………………43
练习题五
二　频率和概率 …………………………………………47
练习题六
三　事件间的关系和运算…………………………………53
练习题七
四　概率的加法 …………………………………………73
练习题八
五　小结…………………………………………………76

第三章　等可能性事件的概率……………………………………77
一　关于排列、组合的一些知识……………………………77
二　等可能性事件的概率……………………………………79
三　概率的古典定义和统计定义的关系……………………83
四　等可能性事件的概率的性质和计算……………………85
练习题九
五　几何概率……………………………………………………93
练习题十
*六　较复杂的例子…………………………………………100
练习题十一
七　小结………………………………………………………111
第四章　条件概率和概率的乘法……………………………112
一　条件概率…………………………………………………112
练习题十二
二　概率的乘法………………………………………………118
练习题十三
三　全概率公式………………………………………………123
练习题十四
四　逆概率公式………………………………………………129
练习题十五
五　小结………………………………………………………136
第五章　事件的独立关系和独立试验………………………137
一　二事件的独立关系………………………………………137
练习题十六
二　多个事件的独立关系……………………………………142
练习题十七

三　简单事件独立试验n次恰好发生k次的概率…………150
练习题十八
*四　有关概率$P_{n,p}(k)$的补充知识…………158
五　小结…………167
*第六章　随机变数理论初步及统计方法应用简例…………169
一　随机变数及其概率分布…………169
二　二项分布、泊松分布和正态分布…………173
三　随机变数的均值和方差…………185
四　大数定律和参数定值估计简介…………192
五　二项分布应用简例——符号检验…………198
六　小结…………207
附表Ⅰ…………208
附表Ⅱ…………212
附表Ⅲ…………214
练习题解答…………218

第一章　统计初步

一　引　　言

自从有了人类社会，就有了社会性的生产和军事活动，由此也就出现了各种统计工作，例如需要统计人口，统计财富，统计武器，等等．据说我国在四千多年前的唐尧时代就已经有了关于人口的调查统计．随着社会的不断发展，人口越来越多；生产资料的种类和数量也越来越多；生产活动、社会活动和国家的管理工作越来越复杂；统计工作也必然越来越重要．一个国家要统计全国的人口、资源、国民经济总产值、各种收入和各种开支等等；一个工厂要统计生产的成本、产值、产品的数量质量情况等等；甚至一个家庭根据收入来安排开支也离不开统计工作．

统计工作可分两大类：一类叫全面统计，一类叫抽样统计．全面统计就是对全体被统计对象全面地、没有遗漏地进行统计．抽样统计就是从全体被统计对象中抽取一部分作样子来推断整体的情况．

全面统计可以达到对被统计对象的全面了解，这是它的优点．但在实际工作中，全面统计有时是行不通的．例如在下面两种情况下实际上不能进行全面统计．

第一种情况是全面统计的工作量太大．例如铆钉厂生产了数以十万计的铆钉，这些铆钉必须合乎一定的规格，不合

格的就是废品．为了保证产品质量需要进行检验，统计铆钉合格的占多少，废品占多少．如果进行全面统计，就要对全部铆钉逐个检验，这样将花费大量人力、物力和时间，付出的代价太高．又如为了鉴定某种棉花的质量，需要统计棉纤维的长度，对每根棉纤维都进行测量统计（全面统计）是不可能的．第二种情况是有些产品质量需要检查，但检查统计有破坏性，全面统计等于全部破坏．例如要了解一批炮弹的效能，需要知道单发炮弹的杀伤半径，这可以通过试射测出数据进行统计，但试射一发就少一发，全面检查就要炸掉全部炮弹．又如灯泡质量的重要指标之一是耐用时间，若对每个灯泡都试验其耐用时间，试验和统计结束，灯泡也就全部报销了．再如要了解冬小麦的生长发育情况，需要知道小麦的分蘖和根的状态，全面检查统计就要把每株小麦都挖出来，那将破坏全部麦田．又要进行统计，又不能全面统计，那就只能进行抽样统计了，即从整体中抽出一部分做样子，来推断整体的情况．例如从大批铆钉中抽取100只进行检查统计，从大批灯泡中抽取10个进行检查统计．其实，劳动人民在实践中早就有了一些抽样统计的做法，象农民总是从一堆粮食中抓出一把来检查粮食的质量．

进行抽样统计，当然希望抽出的部分尽可能的少，同时推断整体情况尽可能的准确，但这里有着随机因素（也叫偶然因素）的影响．你抽取一个铆钉，直径可能偏大，也可能偏小；你抽取一个灯泡，耐用时间可能较长，也可能较短．怎样安排抽样使抽出的部分有更好的代表性？怎样利用抽样的结果对整体作出尽可能正确的判断？为了解决这些问题，人们在长期实践的基础上逐步建立了一套研究随机现象（亦

即偶然现象）数量规律的数学理论和方法，这就是概率统计的理论和方法。

除上述问题外，还有许多自然科学和技术科学都促进了概率统计理论的发展。随着工业化大生产的发展，科学技术不断地提出许多新问题；这些问题，本质上都联系到随机现象的数量规律，都类似于由部分推断（或阐明）整体的问题。

例如分子物理学中气体的一些宏观性状（温度、压强等）是大量分子运动状态的总和表现。我们不可能全面统计每个分子的运动速度、加速度、碰撞、……，但需要从单个分子运动的某些共同特点出发来阐明大量分子运动的宏观规律。

抽样统计的方法，由部分推断整体的方法就是概率论的方法，或称统计方法。如前所述全面统计按一般习惯的理解当然也应算是统计方法，但在概率统计及有关的科技领域中，统计方法一词专指概率论的方法。

由于随机现象广泛存在于自然界和生产实际中，而科学技术的深刻化和精确化必须考虑随机因素，因而概率统计的理论和方法在现代科学技术的各个部门都得到广泛应用。

近代物理的所有分支（尤其是原子核物理学）都必须考虑基本粒子的随机运动，因而都要用到概率论的方法。统计方法也大量应用于近代天文学、地质学。在气象学、地震学中用统计方法研究气候变化和地震发生的规律，进行统计预报，都获得了卓著的成效。在控制论的基础理论中统计方法占突出的地位，在农业科学试验、医学和生物遗传学中也广泛地应用着统计方法。

概率论是误差理论的基础．测量学、航空学、航海学也都离不开统计方法．军事技术的各部门更是应用统计方法的广阔场所．在射击论，投弹论，弹药论，瞄准论和火力控制论，防空论，以及其他军事科学中，统计方法都有重要应用．在军事和公安的侦察、监视、搜索、追踪等方面以及通讯理论和技术，各种社会服务事业，经济计划和生产管理，产品检查和质量控制等都广泛应用统计方法．

我们在统计初步这一章里要讲的是应用统计方法解决实际问题时经常用到的一些简单概念和整理数据、列表、作图、计算的简单技能．

二　总体、个体和样本

（一）总体和个体

总体（有的书上叫母体）是指我们在某一次统计分析工作中所要研究的对象的全体，“个体”是指所要研究的全体对象中的每一个单位．

例如要研究某工厂1978年9月20日生产的一批螺丝帽的内径大小情况．那么这批螺丝帽的内径的全体（比方说有20万个螺丝帽，就有20万个内径）就是总体，而每一个螺丝帽的内径则是一个个体．

又如要了解某砖瓦厂1978年9月24日产红砖的抗折强度情况，那么该厂这天生产的全部红砖的抗折强度就是我们所要研究的总体．比方说，这天生产了15万块砖，那么就有15万个抗折强度，总体就由这15万个抗折强度构成，而每个抗折强度都是一个个体．

又如我们要研究某种新型炮弹的杀伤半径（炮弹爆炸时杀伤范围近似球形，球的半径叫杀伤半径），那么，这种类型的全部炮弹（包括已生产出来的，也包括将要按同样规格，用同样原材料生产出来的这种类型的炮弹）的杀伤半径就构成总体，而每发炮弹的杀伤半径则是一个个体。

又如我们要研究某养鸡场一岁母来航鸡的年产蛋量，那么该鸡场所有的一岁母来航鸡的年产蛋量就构成一个总体，而每只一岁母来航鸡的年产蛋量则是一个个体。

在这里应注意，很可能在同一总体中两个或多个个体取同一数值，这时我们仍认为是两个或多个个体，例如，年产蛋量都是300个的两只母来航鸡的产蛋量看作两个个体300，300。

从上面所举的例子可以看出，什么是总体，什么是个体，要依研究任务而定．如果任务是研究1978年9月20日这一天某工厂所产螺丝帽的内径大小，那么如前所说，该厂在这一天所产全部螺丝帽的内径构成一个总体；但当我们的任务是研究该厂1978年全年日产螺丝帽平均内径的情况时，则9月20日该厂所产全部螺丝帽的平均内径就成了一个个体，1978年全年中每一个工作日都提供这样一个个体，它们的全体构成一个总体．由此可见，统计学中的个体和总体概念要比一块砖和一堆砖的概念抽象一些，应用于实际问题时务必要分辨清楚．

从上面例子还可看出，统计学中的个体一般不是笼统地指某具体事物，而是指该具体事物的某一数量指标，如螺丝帽的内径，红砖的抗折强度，炮弹的杀伤半径，母来航鸡的年产蛋量．这些都是具体事物的数量指标而不是笼统地指螺

丝帽、红砖、炮弹和母来航鸡.

从集合观点看来，很明显，总体是一个集合，个体是集合的元素.

（二）有穷总体和无穷总体

总体中包含的个体个数是有穷的，这种总体叫有穷总体. 例如1978年9月24日某砖瓦厂生产的砖有15万块，现在要研究这批砖的抗折强度. 我们知道，15万块砖就有15万个抗折强度，就是说这天该厂生产的砖的抗折强度的总体就包含15万个个体，因此是一个有穷总体. 又例如某养鸡场所有一岁母来航鸡的年产蛋量总体也是有穷的，有6万只一岁的母来航鸡就有6万个年产蛋量，我们所说的总体中就有6万个个体.

总体中包含的个体个数是无穷的，这种总体叫无穷总体. 例如研究太平洋的深度，因为太平洋中有无穷多个地点，每个地点有一个深度，所以太平洋中一切深度的总体是一个无穷总体. 又例如研究一个蒸馏车间里的温度情况，我们可以用温度计在车间中各个位置测量温度. 车间里的位置有无穷多个，所以车间各位置的温度的总体也是一个无穷总体.

包含个体数量很大的有穷总体往往近似地看成无穷总体. 例如15万块砖的抗折强度的总体，6万只鸡的产蛋量的总体，1万个初生兔的体重的总体，20万个螺丝帽的内径的总体等都可以近似地看成无穷总体. 但若总体含有个体数量只有几百个或更少，一般就不能当作无穷总体看待了.

很明显，从集合的观点看来，有穷总体就是有穷集合，

而无穷总体是无穷集合.

(三) 样　　本

从总体中取出的一部分个体的集合叫做这个总体的样本. 样本中含有个体的个数叫做样本的大小. (样本也叫子样)

很明显, 样本就是总体的子集合. 在理论上当然可以考虑含无穷多个个体的样本, 也可以把总体本身看成一个样本. 但有实用价值的还是含有穷多个个体的样本, 今后如不特别声明, 我们所说的"样本"都是指作为总体的真子集的, 只含有穷多个个体的那种样本.

例1　某糖厂用自动打包机包装糖, 每包标准重量是100斤, 某日开工后测得9包重量如下 (单位: 斤):

99.3, 98.7, 100.5, 101.2, 98.3,

99.7, 99.5, 102.1, 100.5

在本例中, 某日该糖厂用自动打包机打的全部糖包的单包重量是一总体, 每包的单包重量是一个个体, 而测得的9包重量是一个大小为9的样本.

下面介绍抽取样本的方法.

1. 随机抽样和随机样本

抽取样本的方法和过程叫"抽样". 怎样抽样是一个重要问题. 比方说专挑大的或专挑小的, 专挑质量好的或专挑差的都是一些抽样方法. 但我们的目的是通过样本来判断总体情况, 因此要求抽出的样本具有代表性. 专挑某种个体得到的样本难以反映出总体的全面情况, 代表性很差, 因此专挑大的、小的、坏的、好的都不是合理的抽样方法. 得到广泛

使用的抽样方法是随机抽样法，随机抽样即抽样时不加任何主观偏向，使所有个体处于平等地位，每一个个体被抽到的可能性都是相等的.

今后如不特别声明，我们所说的抽样总是指随机抽样.

用随机抽样方法得到的样本叫随机样本. 今后如不特别声明，我们所说的样本也都是指随机样本.

2. 不放回抽样和放回抽样

抽样时如果抽出第 1 个不放回去；再抽第 2 个，不放回去；……；一直抽到第 n 个，得到一个大小为 n 的样本. 这种抽样叫做不放回抽样.

另一种抽样法是抽出第 1 个个体，记下数据后放回总体中去；再抽第 2 个，记下数据后再放回去；……；一直抽到第 $n-1$个，记下数据再放回去；最后抽出第 n个. 这样也得到一个大小为 n 的样本，这种抽样叫做放回抽样.

两种抽样的区别在于，在不放回抽样中，每个个体至多被抽到一次；而在放回抽样中，每个个体被抽到有可能多于一次. 例如说第一次抽到一只直径 13.50 毫米的铆钉，对不放回抽样来说，以后各次就抽不到这只铆钉了；而对放回抽样来说，以后各次还可能抽到这只铆钉.

今后如不特别声明，我们所说的抽样总是指不放回的抽样.

3. 大样本和小样本

样本的大小是相比较而言的，在实际应用中常把大小在50以下（一般是大小在10左右）的样本叫小样本；而把大小超过50的样本叫大样本.

例 1 中的样本大小 $n=9$，这就是一个小样本.

例2 研究某机床所产铆钉的质量，从该机床所产铆钉中随机抽取200个，测量其直径得200个数据如下：

表 1

以毫米计的200个铆钉的直径

13.39	13.43	13.54	13.64	13.40	13.55	13.40	13.26
13.42	13.50	13.32	13.31	13.28	13.52	13.46	13.63
13.38	13.44	13.52	13.53	13.37	13.33	13.24	13.13
13.53	13.53	13.39	13.57	13.51	13.34	13.39	13.47
13.51	13.48	13.62	13.58	13.57	13.33	13.51	13.40
13.30	13.48	13.40	13.57	13.51	13.40	13.52	13.56
13.40	13.34	13.23	13.37	13.48	13.48	13.62	13.35
13.28	13.59	13.47	13.46	13.62	13.54	13.20	13.38
13.43	13.35	13.56	13.51	13.47	13.40	13.29	13.20
13.46	13.44	13.42	13.29	13.41	13.39	13.50	13.48
13.53	13.34	13.45	13.42	13.29	13.38	13.45	13.50
13.55	13.33	13.32	13.69	13.46	13.32	13.32	13.48
13.29	13.25	13.44	13.60	13.43	13.51	13.43	13.38
13.24	13.28	13.58	13.31	13.31	13.45	13.43	13.44
13.34	13.49	13.50	13.38	13.48	13.43	13.37	13.29
13.54	13.33	13.36	13.46	13.23	13.44	13.38	13.27
13.66	13.26	13.40	13.52	13.59	13.48	13.46	13.40
13.43	13.26	13.50	13.38	13.43	13.34	13.41	13.24
13.42	13.55	13.37	13.41	13.38	13.14	13.42	13.52
13.38	13.54	13.30	13.18	13.32	13.46	13.39	13.35
13.34	13.37	13.50	13.61	13.42	13.32	13.35	13.40
13.57	13.31	13.40	13.36	13.28	13.58	13.58	13.38
13.26	13.37	13.28	13.39	13.32	13.20	13.43	13.34
13.33	13.33	13.31	13.45	13.39	13.45	13.41	13.45
13.40	13.36	13.45	13.48	13.29	13.58	13.44	13.56

这是一个大小为$n=200$的大样本。

4．利用随机数表抽样

前面说过，为了使抽出的样本有代表性，我们一般是利用随机抽样方法来取得随机样本，当总体较小而且每个个体体积不太大时，直接进行随机抽样是比较容易的，但当总体很大或单个个体体积较大时，直接进行随机抽样就困难了．比方说从100只铆钉中随机抽出10只很容易，只要把100只铆钉放在盒里摇一摇，然后不加挑选地随手取出10只就行了．但要从20万只铆钉中随机抽出10只或200只来就不容易保证抽样的随机性，因为20万只堆在一起，靠里边的就很难被抽到．又如从100个学生中随机抽出10人考查学习情况也不容易保证抽样的随机性，抽样者闭上眼睛摸10个人，施行起来也不方便．站在不同位置的学生被摸到的可能性也不相同．

为了保证抽样的随机性，可以把每个个体编上号码，然后随机地抽号码，例如把100个学生编为1～100号，在100张同样的小纸片上分别写上1～100这100个数码，然后从这100张小纸片中随机抽取10张，由此确定10个学生．这个办法当总体很大时仍不够方便，由此发展出利用随机数表抽样的办法．

先介绍随机数表．把0、1、2、3、4、5、6、7、8、9这十个数码分别写在十张同样的小纸片上，放在盒里或袋里，随机取一张，记下号码，放回去；再取一张，记下号码，放回去；如此进行10000次，把所记下的10000个数字依次列在表上就得到一个常用的随机数表．