高等工程专科教学用书

试验数据分析

邵振和　编

上海科学技术文献出版社

试验数据分析

邵振和 编

上海科学技术文献出版社

（沪）新登字301号

高等工程专科教学用书
试验数据分析
邵振和编
上海科学技术文献出版社出版发行
（上海市武康路2号）
上海崇明永南印刷厂印刷
开本787×1092 1/32 印张6.75 字数163,000
1994年9月第1版 1994年9月第1次印刷
印数：1—2,850
ISBN 7-5439-0614-4/O·102
定价：6.80元

内 容 提 要

国家教委指出，高等工程专科教育是高等教育的一个重要组成部分，明确其教学中要"以应用为目的，以必需、够用为度"，"不必强调理论的严密性和系统性，要以培养技能，强化应用为教学重点。"

试验数据的统计推断是一门有着广泛应用的学科，而适用于高等工程专科教学要求的该学科的教材，目前还较少，为此编写了这本以介绍对试验数据统计推断为主，兼顾必要的概率知识的教材。经过几年来教学试点，实用效果较好。

本教材内容有：概率知识简介、样本及其分布、假设检验、参数估计、回归分析、方差分析及正交试验设计简介。概率知识以统计中必需为度；其它内容以突出常用的基本方法为主；正交试验设计的方法，对于今后从事生产第一线上技术工作的学生来说是十分有用的，因此也简单的作了介绍。

本书系高等工程专科学校教材，也可供生产第一线上的广大工程技术人员使用。

前　　言

国家教委对高等工程专科教育十分重视，指出高等工程专科教育中的基础理论教学应“以应用为目的，以必需够用为度”，不必强调理论的严密性与系统性，不重理论推导而重结论和应用。针对以上要求，我们对原概率论与数理统计课程的教学，进行大胆的改革，经过两年多的实践，编写这本较为切合教学要求的教材《试验数据分析》，本书有如下几方面特点：

(1) 只须约30课时的时间，便可介绍以下内容：统计中所必需的概率知识、样本及其分布、假设检验、参数估计、回归分析，方差分析及正交试验设计简介。突出应用且课时少是其特点之一；

(2) 对本书的重点内容：假设检验、参数估计、回归分析、方差分析及正交试验设计，采用通俗讲解的方式，使学生了解主要原理及思想方法，重点掌握有关的结论及计算方法。由于在编写时，防止内容上不必要的重复及注意到各部分内容的内在联系，因此教材内容简明，易于接受。

(3) 只讲解统计中必需用到的概率知识，由易于接受的实例引入必需的某些结论。减少了内容，讲解中仅需6学时左右，试行后，效果良好。

(4) 考虑到正交试验设计的实用价值很大，它特别适合于生产第一线上的工程技术人员的需要，因此也作了简单的介绍。由于已有了方差分析的基础，所以直接介绍正交试验设计的方差分析，从而减少了教学时间，需9学时左右。

(5) 书上 * 号内容可不讲，总学时约为 28～32 左右。

本书由同济大学教授、全国应用概率委员会名誉理事王福保先生主审，天津职业技术师院姬振豫副教授对本书也提出宝贵意见，在此一并表示感谢。

由于编者业务水平不高，经验不足，虽经努力但仍存在许多问题，敬请批评指教。

编者　邵振和

于上海化工高等专科学校

数学教研室

1993 年 6 月

目　　录

第六章　方差分析

第七章　正交试验设计简介

第一章　概率知识简介

本章简单介绍后面几章中要用到的概率论的知识：随机事件及其概率、随机变量及其概率分布、随机变量的数学特征等内容。

§1.1　随机事件及其概率

一、随机事件的概念

在生产实践、科学试验和日常生活中，我们常常要讨论试验的各种各样结果。例如：

（1）掷一粒骰子的试验，其结果是出现的点数为1,2,3,4,5,6中的某一个。至于试验结果到底是哪一个，在试验之前无法预知，只有试验结束之后才知道；

（2）日光灯管使用寿命的试验，其结果是任意的某个非负实数。试验结果到底是多少，在试验之前无法预知；

（3）从一批含一定数量次品的产品中任意抽取四个检验的试验，观察出现的次品数，其试验结果是0,1,2,3,4中的某一个。试验结果在试验之前无法预知。

以上试验具有两个共同特点：

第一，可在相同条件下重复进行；

第二，每次试验的可能结果不止一个，究竟出现哪个结果，在试验之前无法预知。

我们把具有以上两个特点的试验，称为**随机试验**。（简称**试验**）

称按试验结果而确定发生或不发生的事件为**随机事件**（简

称**事件**），通常用大写英文字母$A,B,C,\cdots$表示。

随机试验中，由单个试验结果构成的事件，称为**基本事件**。

【例 1】 在掷骰子试验中，用A表示"出现 1 点"的事件，记为 A="出现 1 点"。同样的，设有 B="出现 2 点"、C="出现偶数点"、D="出现的点数不小于 4"，那么

A,B,C,D 均为事件，而事件 A,B 为基本事件。

在试验中，必然会发生的事件称为**必然事件**，通常用大写字母 S 表示；

在试验中，不可能发生的事件称为**不可能事件**，通常用大写字母Φ表示。

【例 2】 在掷骰子试验中，设 E="出现点数不大于 6"，F="出现点数大于 8"。那么E是必然事件，F是不可能事件。

必然事件和不可能事件都不是随机事件，但为了下面讨论问题方便起见，我们把必然事件和不可能事件当作特殊的随机事件。

二、事件的频率和概率的统计定义

在一次试验中，某事件是否发生是不能事先确定的，具有偶然性。但从大量的重复试验中，还是可以了解该事件发生的可能性的大小的，即具有"统计"的规律性。下面先介绍事件发生的频率的概念。

【定义 1】 在相同条件下进行 n 次重复试验，如果事件 A 发生了m次，称比值$\frac{m}{n}$为事件A的**频率**，记为

$$W(A)=\frac{m}{n}$$

频率具有如下性质：

(1) $0\leqslant W(A)\leqslant 1$；

(2) $W(S)=1$；

(3) $W(\Phi)=0$。

通常，当试验次数充分大时，可以发现，事件A的频率呈现出稳定性，即可在一个确定的数值附近摆动。

历史上有人做过如下试验：投掷硬币，观察出现正面的次数。表1·1列出了他们试验的记录及相应的频率。

表1.1

试验者	投掷次数n	出现"正面"次数m	频率$\frac{m}{n}$
蒲丰(Buffon)	4040	2048	0.5069
皮尔逊(Pearson)	12000	6019	0.5016
皮尔逊(Pearson)	24000	12012	0.5005
维尼	30000	14994	0.4998

从表1·1可看出，出现正面的频率接近0.5，并且投掷次数越多，频率越接近0.5。这个稳定值0.5是客观存在的，用它可以来反映随机事件发生的可能性的大小。这是我们下面定义事件的概率的客观基础。

【定义2】 在相同条件下进行大量重复试验，当试验次数充分大时，事件A的频率总在某一数值p附近摆动，则称p为事件A的**概率**，记为

$$P(A)=p$$

概率的上述定义称为**概率的统计定义**。

由概率的统计定义可知，概率有如下性质：

(1) $0\leqslant P(A)\leqslant 1$;

(2) $P(S)=1$;

(3) $P(\Phi)=0$。

三、古典概型和概率的古典定义

有些事件的概率，不必做大量重复试验去讨论，可以直接计算其概率，这就是所谓的古典概型。下面先看一个例子。

掷一粒均匀的骰子，观察向上一面出现的点数。

设 A_i="出现 i 点"，($i=1,2,3,4,5,6$) 那么 A_i 为基本事件。基本事件的个数为 6，有限；由于骰子是均匀的，故每个基本事件发生的可能性相同。

【定义 3】 我们把仅含有限个基本事件且每个基本事件发生的可能性都相同的随机试验，称为**古典概型**。

对于古典概型，事件的概率可以直接计算。

【定义 4】 如果试验结果有 n 个基本事件组成，每个基本事件出现具有相同的可能性，而事件 A 由其中某 m 个基本事件所组成，则 A 的概率为

$$P(A)=\frac{m}{n}$$

概率的上述定义称为概率的古典定义。

显然有如下性质：

(1) $0\leqslant P(A)\leqslant 1$；

(2) $P(S)=1$；

(3) $P(\Phi)=0$。

【例 3】 掷一粒均匀骰子，观察向上一面出现的点数。设 B="出现偶数点"，C="出现的点数不小于 3"，那么有

$$P(B)=\frac{3}{6}=\frac{1}{2}$$

$$P(C)=\frac{4}{6}=\frac{2}{3}$$

这是因为该试验属古典概型，含有 6 个基本事件。事件 B

由3个基本事件："出现2点"、"出现4点"、"出现6点"所组成；事件C由4个基本事件："出现3点"、"出现4点"、"出现5点"、"出现6点"所组成。

【例4】 将一枚硬币连抛两次，观察正、反面出现情况。求下列事件的概率：

A="恰有一次出现正面"

B="至少有一次出现正面"

【解】 该试验所有可能情况为

(正,正)，(正,反)，(反,正)，(反,反)

因此，基本事件总数$n=4$。

事件A由2个基本事件：(正,反)，(反,正)所组成，故有

$$P(A)=\frac{2}{4}=\frac{1}{2}$$

事件B由3个基本事件：(正,正)，(正,反)，(反,正)所组成，故有

$$P(B)=\frac{3}{4}$$

【例5】 某盒中装有100只集成电路，其中有96只正品，4只次品。随机地从盒中任取4只，求下列事件的概率：

A="4只全是正品"

B="4只中恰有一只次品"

【解】 基本事件总数为$n=C_{100}^4$

组成事件A的基本事件数$m_A=C_{96}^4$，故有

$$P(A)=C_{96}^4/C_{100}^4=0.847$$

组成事件B的基本事件数$m_B=C_4^1\cdot C_{96}^3$，故有

$$P(B)=C_4^1\cdot C_{96}^3/C_{100}^4=0.146$$

【例6】 一个三位数字的号码锁，每位上有0,1,2,…,9

十个数字码。如果不知道该锁号码，问开一次锁就能把该锁打开的概率有多大？

【解】 每位数字上有十个数码，三位数字所构成的号码有 10^3 个，因此基本事件总数 $n=10^3$

而能打开的号码仅有一个，所以 $m=1$，因此能打开**该锁的**概率为 $\frac{1}{10^3}=0.001$，这是一个可能性很小的随机事件。

通常，我们把概率很小(比如为 0.1，0.05，0.01 或更小等)的事件，称为**小概率事件**。且认为：**在一次随机试验中，小概率事件几乎是不会发生的。**

四、事件的运算

【定义 5】 两事件 A, B 中至少有一个发生，是一个事件，称其为事件 A 与 B 的并(或和)，记为 $A+B$。

【定义 6】 两事件 A 与 B 同时发生，是一个事件，称其为事件 A 与 B 的交(或积)，记为 AB。

上述两定义可以推广到有限个事件的和与积。

【例 7】 掷一粒均匀骰子，观察向上一面的点数，设

$A_i=$“出现 i 点”　　　　$(i=1,2,3,4,5,6)$

$A=$“出现偶数点”

$B=$“出现 1 点或 3 点”

$C=$“出现的点数小于 3”

那么有

$$A=A_2+A_4+A_6$$

$$B=A_1+A_3$$

$$C=A_1+A_2$$

$$AB=\Phi$$

$$AC=A_2$$

$$BC = A_1$$

【定义 7】 如果两事件A与B不能同时发生（即$AB=\Phi$），称事件A与B是互斥的。

可以有如下结论：

（1）基本事件间是互斥的；

（2）互斥事件A与B的和的概率等于两事件的概率之和，即

$$P(A+B)=P(A)+P(B)$$

【例 8】 掷一粒均匀的骰子，观察向上一面出现的点数。

设 A_i = "出现 i 点" $(i=1,2,3,4,5,6)$

A = "出现偶数点"

B = "出现 1 点或 3 点"

C = "出现的点数小于3"

试验证 $P(A+B)=P(A)+P(B)$，

而 $P(A+C)\neq P(A)+P(C)$

【证明】 基本事件总数$n=6$，

组成事件A的基本事件数$m_A=3$，同理有$m_B=2$，$m_C=2$。

组成事件$(A+B)$的基本事件数$m_{(A+B)}=5$，这是因为事件$(A+B)$由基本事件A_1,A_2,A_3,A_4,A_6所组成。

同理 $m_{(A+C)}=4$

因此有 $P(A+B)=\frac{5}{6}$ $P(A)=\frac{3}{6}$ $P(B)=\frac{2}{6}$

所以有 $P(A+B)=P(A)+P(B)$

由于 $P(A+C)=\frac{4}{6}$ $P(C)=\frac{2}{6}$

所以 $P(A+C)\neq P(A)+P(C)$

（3）所有基本事件A_i $(i=1,2,\cdots,n)$之和构成必然事件，

所以有 $$\sum_{i=1}^{n} A_i = S$$

于是有 $$P\left(\sum_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) = 1$$

（4）如果事件A与B满足 $A+B=S$ 且 $AB=\Phi$，则称事件B与事件A是对立的，记为 $B=\bar{A}$，那么有

$$P(\bar{A}) = 1 - P(A)$$

五、条件概率及随机事件间的独立性

【定义 8】 在事件B已发生的条件下，事件A发生的概率，称为A对B的**条件概率**，记为 $P(A|B)$。

【例 9】 某口袋中装有18只白球和2只红球，随机地从袋中接连取两球，每次取一只。如果设

$$A = \text{“第二次取得红球”}$$
$$B = \text{“第一次取得红球”}$$

讨论事件A的概率。

【解】 事件A发生的概率与取球的方式有关。

（1）如果取球时，第一次取球后放回，那么在事件B发生的条件下，事件A的概率为

$$P(A|B) = C_2^1/C_{20}^1 = \frac{1}{10}$$

而不论事件B发生与否，事件A的概率为

$$P(A) = (C_2^1 \cdot C_2^1 + C_{18}^1 \cdot C_2^1)/(C_{20}^1 \cdot C_{20}^1)$$
$$= \frac{4+36}{400} = \frac{1}{10}$$

它们是相同的，即 $P(A|B) = P(A)$

（2）如果取球时，第一次取球后不放回，那么

$$P(A|B) = C_1^1/C_{19}^1 = \frac{1}{19}$$

$$P(A)=(C_2^1\cdot C_1^1+C_{18}^1\cdot C_2^1)/(C_{20}^1\cdot C_{19}^1)$$
$$=\frac{2+36}{20\times 19}=\frac{1}{10}$$

它们是不同的。即 $P(A|B)\neq P(A)$

前者，称为**放回取样**。事件B的发生与否，不影响事件A的概率；后者，称为**不放回取样**。事件B的发生与否，影响事件A的概率。

【定义8】 如果事件B的发生与否，不影响事件A的概率，即 $P(A|B)=P(A)$（可以推出，它等价于$P(AB)=P(A)P(B)$），**称事件A对事件B是独立的。**

可以证出，如果事件A对事件B是独立的，那么事件B对事件A也是独立的，即它相互独立。

我们经常要将某个随机试验在相同条件下重复进行n次，如果任何一次试验结果的概率都不依赖于其它各次试验的结果，则称这n次试验为**n次重复独立试验**。比如，有放回的抽检n次产品，便是n次重复独立试验。可以从大量重复独立试验中发现事件的统计规律性。

§1.2 随机变量及其概率分布

上节我们讨论了随机事件及其概率，为了进一步研究随机现象，我们将引入随机变量的概念。

一、随机变量

【引例1】 设有产品10件，其中有3件次品，7件正品，现从中随机抽取4件，向“抽得的次品件数”（以下简称“次品数”）是多少？

记X表示“次品数”，X的可能取值为0,1,2,3，因此X是变量，但X的值在抽样前无法确定，而在抽样后又是完全确定