现代工业管理统计方法

郑绍濂　吴立鹏　郑祖康　编

机械工业出版社

(京)新登字054号

本书比较全面地介绍了现代工业管理中应用的统计方法。全书分为两部分，第一章至第五章为基础部分，阐述了概率统计的基本理论，第六章至第十一章介绍了现代工业管理中常用的统计方法。

本书力求做到简明扼要，通俗易懂，读者阅读后能很快地掌握其主要方法，并用之于实践。在内容方面，本书介绍了一些较新的应用方法，对于可靠性统计、生存分析、统计质量控制、三次设计截断数据回归、岭回归以及非参数回归等方面都进行了阐述，希望能尽量满足我国日益发展的现代工业管理的需要。

本书可供工业管理人员、工程技术工作者以及工科院校的师生阅读，也可作为管理、工程等类大专院校的教科书或参考书。

图书在版编目（CIP）数据

现代工业管理统计方法/郑绍濂等编.—北京：机械工业出版社，1994.12

ISBN 7-111-04098-8

Ⅰ. 现…
Ⅱ. 郑…
Ⅲ. 工业统计
Ⅳ. F402.4

出 版 人　马九荣（北京市百万庄南街1号　邮政编码100037）
责任编辑：张淑琴　版式设计：胡金瑛　责任校对：肖新民
封面设计：方　芬　责任印制：路　琳
机械工业出版社印刷厂印刷・新华书店北京发行所发行
1994年9月第1版・1994年9月第1次印刷
787mm×1092mm 1/32・16.375印张・362千字
0 01—900 [illegible]册
定价 26.00 元

前　　言

随着改革开放的进一步深入，搞好现代工业管理具有重大的意义。如何有效地收集、整理和分析受随机影响的数据，并对所考虑的问题作出推断、决策与预测是一个十分重要的问题，直接影响到企业的经济效益与发展前景。数理统计则为之提供了必要的工具。编写这本书的目的是为了帮助工业管理人员、工程技术人员以及有关方面的工作者尽快地树立统计思想，掌握基本的统计方法，迅速地应用于实际工作。本书介绍了产品的抽样和检验方法，分析了生产过程的各种关系及其模型，给出了试验设计的优良方案，讨论了质量管理与控制的各种办法以及产品的可靠性与寿命估计，最后利用时间序列分析给出了若干预测方法。

在本书的编写中，我们力图做到简明扼要，理论联系实际，侧重实用，并引进一些较新的方法，希望对不同工作岗位上的读者都能有些益处。

在本书的编写中，武汉工业大学北京研究生院袁子仁教授仔细阅读了全书，并提出了不少宝贵的建议；本书的出版得到国家自然科学基金与高等学校博士学科点专项科研基金的资助；谨此一并致谢。

由于水平所限，编写工作又较匆促，因此一定存在不少缺点和错误，欢迎读者批评、指正。

编　者

目　　录

第一章　概率论基础知识

第一节　随机事件与概率

（一）随机现象与随机事件

在研究自然界和人类社会中各种事物运动和变化的规律时，会发现有两类不相同的现象。一类是在一定条件下有确定变化规律的现象，一旦认识了它的规律就可以事先作出准确的预言。例如太阳从东方升起；带电的物体间总是同性相斥，异性相吸；在理想的条件下，自由落体总是遵循着 $S=\frac{1}{2}gt^2$ 的运动规律；生物总是要经历生长、发育、衰老直至死亡的各个阶段等等。这一类现象我们称之为确定性现象。在一定条件下必然会发生的事情称之为必然事件；反之，在一定条件下必定不会发生的事情称之为不可能事件。

但是在自然界和人类社会中还广泛地存在着与确定性现象有着本质不同的另一类现象。例如，抛掷一枚硬币，事先并不能准确地预言结果是出现正面或反面；在一批产品中任意选一件，事先也不能断定是正品或次品；事先也无法确定未来一小时内电话交换台接到的电话呼叫次数；同一条生产线上用同样工艺生产出来的灯泡寿命也呈现出偶然性；在使用仪器测量物体的长度时，虽然物体的长度是确定的量，但由于测量仪器受到周围环境的影响以及观察者生理和心理上的偶然变化，使得观察结果中含有无法事先确定的测量误差；对于一切实验数据也同样包含着这种无法事先确定的误差；

再如某计算机下一次故障发生的时刻也是无法事先预计的；某商店次日的营业额，也是无法知道的。

上述种种现象有一个共同点：在基本条件不变的情况下，经过一系列试验或观察后所得到不同的结果，而各种结果的出现呈现出一种偶然性，这种现象或试验，称为随机现象或随机试验，简称试验。并约定随机试验总是在相同条件下重复进行的。

随机试验的各种结果称为随机事件，简称事件。对于随机事件，我们不能预言它在试验中一定发生或一定不发生，我们的兴趣在于找到一个量来描述它发生的可能性大小。一个随机事件发生的可能性大小是随机事件本身所固有的属性，是可以度量的，也就是说，对一个随机事件A，存在着一个与A相对应的数$P(A)$来刻划随机事件A发生的可能性大小。我们称$P(A)$为事件A发生的概率。

（二）样本空间与事件

为了研究随机试验，首先要知道的是这个试验可能出现哪些结果，这些结果称为样本点，一般用ω表示，它们在一次试验中不会同时出现。

随机试验的样本点全体称为样本空间，通常记作Ω

〔**例1-1**〕 从编号为1，2，…，n的n台电视机中任意选取一台，则$\Omega=\{1，2 \cdots n\}$

〔**例1-2**〕 考察明年上海地区梅雨季节的降雨量，则$\Omega=[0，\infty)$

由样本空间这一概念，容易知道，任一随机事件都是样本空间Ω的子集，或者说事件是由一些样本点所组成的集合。

在例1-1中，令A为“电视机号码不超过R”的事件，则

$A=\{1,2\cdots R\}$

在例1-2中，令A为“明年上海地区梅雨季节的降雨量在300mm以上”的事件，则$A=(300,\infty)$

当且仅当事件A所包含的某个样本点在试验中出现时事件A才发生，按这样的概念，样本空间Ω本身就是必然事件。

把不包含任何样本点的集合（即空集）叫做不可能事件记为ϕ。必然事件和不可能事件是随机事件的两个极端情形。

我们把事件看成是一些样本点的集合，有助于弄清事件之间的关系及其运算。

1．包含关系

若A、B都是随机事件，如果B所包含的全部样本点都属于A，则称事件A包含事件B，记作$A\supset B$或$B\subset A$，即若事件B发生，则必有事件A发生。

显然，对于任意事件A，均有$\Omega\supset A\supset\phi$。

如果$A\supset B$，$B\supset A$，则称A与B等价，记作$A=B$，即两事件为同一事件。

2．逆事件

A为一事件，在Ω中全体不属于A的样本点所组成的集合，称为A的逆事件，记作$\bar{A}$。当且仅当A不发生时$\bar{A}$才发生。显然，$\bar{\bar{A}}=A$。

3．交事件

A、B是两个事件，由同时属于A、B的全体样本点所组成的集合称为事件A、B的交事件，记作$A\cap B$或AB。当且仅当A、B同时发生时，交事件$A\cap B$才发生。

若$A\cap B=\phi$，则称事件A、B互斥或不相容。

$\bigcap_{i=1}^{n} A_i$表示事件$A_1, A_2 \cdots A_n$同时发生的事件。$\bigcap_{i=1}^{\infty} A_i$表示可列个事件$A_1$，$A_2$，…，同时发生的事件。

4．并事件

A，B是两个事件，由至少属于A，B中之一的样本点全体组成的集合称为A，B的并事件，记作$A \cup B$。当且仅当A，B中至少有一个事件发生时，事件$A \cup B$才发生。

$\bigcup_{i=1}^{n} A_i$表示n个事件A_1，$A_2 \cdots A_n$中至少有一个发生的事件，$\bigcup_{i=1}^{\infty} A_i$表示可列个事件$A_1$，$A_2 \cdots$中至少发生一个的事件。

若事件A，B互不相容、$A \cup B$，可记为$A + B$。

n个事件A_1，$A_2 \cdots$，A_n如两两互不相容，它们的并事件记为$\sum_{i=1}^{n} A_i \equiv A_1 + A_2 \cdots + A_n$。类似地可定义$\sum_{i=1}^{\infty} A_i = A_1 + A_2 + \cdots\cdots$

5．差事件

A、B是两个事件，由属于A但不属于B的全体样本点所组成的集合称为事件A与事件B的差事件，记作$A - B$。当且仅当事件A发生而事件B不发生时，差事件$A - B$才发生。显然，$A - B = A\bar{B}$

（三）古典概型与概率的基本性质

在各种随机试验中，有一类随机试验是最简单的，它满足如下两个条件：

（1）试验的全部样本点为有限的，不妨设为$\omega_1, \omega_2, \cdots$,

ω_n；

（2）每一样本点的出现都是等可能的，即 $P(\omega_i)=\frac{1}{n}$　$i=1,2,\cdots n$。

满足上述两个条件的随机试验模型称为古典概型。

〔例1-3〕 箱中有外形相同的球 R 个，从中随意抽取一个，此时每个球被摸到的概率为 $\frac{1}{k}$，故它是古典概型。这个例子就是产品抽样检验的标准模型。

在古典概型中，事件 A 的概率为

$$P(A)=\frac{A\text{所包含的样本点数}}{\text{样本点总数}} \tag{1-1}$$

对于古典概型概率，

（1）对于任意事件 A，有 $P(A)\geqslant 0$；

（2）$P(\Omega)=1$；

（3）若 A，B 互不相容，则 $P(A+B)=P(A)+P(B)$

容易将（3）推广到

（4）若 A_1，$A_2\cdots$，A_n 两两互不相容，则

$$P\left(\sum_{i=1}^{n} A_i\right)=\sum_{i=1}^{n} P(A_i) \tag{1-2}$$

从上述基本性质出发，可以证明其它一系列概率公式。

（5）　$$P(\overline{A})=1-P(A) \tag{1-3}$$

（6）加法公式：

A，B 是两个事件，则

$$P(A\cup B)=P(A)+P(B)-P(AB) \tag{1-4}$$

（7）若 $A\supset B$，则

$$P(A-B)=P(A)-P(B) \tag{1-5}$$

上述概率公式对于任何概率模型同样成立。在一般概率模型下，尚要用到：

（8）若事件 A_1，A_2……是可列个两两互不相容的事件，则

$$P\left(\sum_{i=1}^{\infty} A_i\right)=\sum_{i=1}^{\infty} P(A_i) \tag{1-6}$$

性质（8）称为概率的可列可加性。

（四）条件概率

我们称在事件 B 发生的条件下，事件 A 发生的概率为事件 A 关于事件 B 的条件概率，记作 $P(A|B)$。

在古典概率的情形下，$P(A|B)$的计算公式很容易导出。若 A 所包含的样本点为 k 个，B 所包含的样本点为 l 个（$l\neq 0$）。AB所包含的样本点为 r 个，在事件 B 发生的情形下，试验的结果一定为 B 所包含的 l 个样本点中的某一个，就是说在 B 发生的条件下，可能的样本点总数为 l 个，在这种情形下，要发生事件 A，当且仅当试验的结果为 AB 所包含的 r 个样本点中的某一个，就是说有利场合的个数为 r。于是，我们有

$$P(A|B)=\frac{r}{l}=\frac{r/n}{l/n}$$

从而条件概率 $P(A|B)$的计算公式为

$$P(A|B)=\frac{P(AB)}{P(B)} \tag{1-7}$$

这里要求 $P(B)>0$。

显然

$$P(\overline{A}|B)=1-P(A|B) \tag{1-8}$$

条件概率的概念，在实际中使用得十分广泛。例如，用自动化仪表检验某电子元件。以D表示“元件是好的”，$\overline{D}$表示“元件是不好的”，A表示“仪器检测后认为元件是好的”，$\overline{A}$表示“仪器检测后认为元件是不好的”。这时仪器的质量和水平就取决于下面两个条件概率：$P(A|D)$与$P(\overline{A}|\overline{D})$。前者表示好的元件经检测后被接受的概率，后者表示坏的元件经检测后被拒绝的概率。这两个概率越接近于1，仪器水平就越高，否则就表示仪器不理想。

由条件概率的定义知

$$P(AB)=P(B)P(A|B) \tag{1-9}$$

这就是概率的乘法公式。推广到多个事件：

$$P(A_1A_2\cdots A_n)$$
$$=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1A_2\cdots A_{n-1}) \tag{1-10}$$

全概率公式：若$A_1, A_2, \cdots A_n$是两两互不相容的n个事件，且$\sum_{i=1}^{n} A_i=\Omega$，$P(A_i)>0$，$i=1, 2, \cdots n$。则对于任意事件$B$有

$$P(B)=\sum_{i=1}^{n} P(A_i)P(B|A_i) \tag{1-11}$$

〔**例1-4**〕 把四个工厂生产的同型号电子元件混合在一起，已知一厂的元件占20%，二厂的元件占30%，三厂的元件占45%，四厂的元件占5%，又一厂元件的次品率为0.1，二厂的次品率为0.05，三厂的次品率为0.02，四厂的次品率为0.04，从混合后的元件中任取一件恰为次品的概率是多少？

解 以A_i表示元件是i厂生产的，$i=1, 2, 3, 4$，B表示“从混合后的元件中任取一件恰为次品”。

显然，$A_1+A_2+A_3+A_4=\Omega$，又$P(A_1)=0.2$，$P(A_2)=0.3$，$P(A_3)=0.45$，$P(A_4)=0.05$，$P(B|A_1)=0.1$，$P(B|A_2)=0.05$　$P(B|A_3)=0.02$，$P(B|A_4)=0.04$。由全概率公式

$$\begin{aligned}P(B)&=\sum_{i=1}^{4}P(A_i)P(B|A_i)\\&=0.2\times0.1+0.3\times0.05\\&\quad+0.45\times0.02+0.05\times0.04\\&=0.046\end{aligned}$$

由全概率公式容易推得另一公式。

贝叶斯公式：若$A_1, A_2, \cdots A_n$两两互不相容，且$\sum_{i=1}^{n}A_i=\Omega$，$P(A_i)>0$ $i=1, 2, \cdots n$，B是任意事件，$P(B)>0$，则

$$P(A_i|B)=\frac{P(A_i)P(B|A_i)}{\sum_{i=1}^{n}P(A_i)P(B|A_i)} \qquad (1\text{-}12)$$

贝叶斯公式常常被用来分析原因，从而作出适当的决定和决策。$P(A_i)$称为先验概率，$P(A_i|B)$称为后验概率。

（五）统计独立性

1．两个事件的统计独立性

若A、B是两个随机事件，如果事件A的发生与否对于事件B发生的概率没有影响，即：$P(B|A)=P(B)$，则称事件A与B彼此独立的，这里要求$P(A)>0$，独立性可定义为

$$P(AB)=P(A)P(B) \qquad (1\text{-}13)$$

2. 三个事件的相互独立性

A、B、C三个事件，如果

$P(AB)=P(A)P(B)$，$P(AC)=P(A)P(C)$，$P(BC)=P(B)P(C)$，$P(ABC)=P(A)P(B)P(C)$

成立则称事件A、B、C是相互独立的。

3. 多个事件的相互独立性

若A_1，A_2，$\cdots A_n$是n个事件，满足

$$P(A_iA_j)=P(A_i)P(A_j) \quad 1\leqslant i<j\leqslant n \qquad (1\text{-}14)$$

$$P(A_iA_jA_k)=P(A_i)P(A_j)P(A_k) \quad 1\leqslant i<j<k\leqslant n$$

$$P(A_iA_jA_kA_l)=P(A_i)P(A_j)P(A_k)P(A_l)$$

$$1\leqslant i<j<k<l\leqslant n$$

$$\cdots$$

$$P(A_1A_2\cdots A_n)=P(A_1)P(A_2)\cdots P(A_n)$$

则称事件A_1，A_2，$\cdots A_n$相互独立。

4. 随机试验的独立性

若有n个随机试验E_1，E_2，$\cdots E_n$，以A_1表示属于试验E_1的任意事件，A_2表示属于试验E_2的任意事件，……，A_n表示属于试验E_n的任意事件，成立

$$P(A_1A_2\cdots A_n)=P(A_1)P(A_2)\cdots P(A_n) \qquad (1\text{-}15)$$

则称试验E_1，$E_2\cdots E_n$是相互独立的。

试验的独立性通常用于重复试验的场合，即同一试验重复进行n次，例如重复投掷一枚硬币n次，同一台机床加工n只同样的零件等等。

试验的独立性通常是根据直观的背景，或者作为近似的假定，实用上很少运用式(1-15)来验证试验是否独立。相反，倒是在直观的背景或近似的假定下来使用式(1-15)，这一点在数理统计中十分重要。

（六）贝努里试验概型

在重复独立试验中最简单也最为重要的随机试验模型就是所谓贝努里试验模型。它满足两个条件

（1）每次试验只有两个结果，即事件A（成功）或$\overline{A}$（失败），$P(A)=p$在每次试验中保持不变；

（2）试验是相互独立的。

贝努里试验概型虽然简单却十分重要，因为在实际问题中，通常我们只对某一特定的事件感兴趣，这时就可把随机试验的结果看成只有两个：发生或不发生。例如工厂里的全部产品可分成各种等级，但如果我们关心的只是次品，则工厂的全部产品就可分成两类：正品和次品。

在多重贝努里试验概型中，我们着重介绍两个概率，即二项分布与几何分布。

二项分布　在n次贝努里试验中事件A恰好发生k次的概率。

把n次贝努里试验的结果记录到n个格子里，恰好发生k次A的事件就是得到这样的记录：有k个A出现在k个格子里，而其余的$n-k$个格子出现$\overline{A}$，由试验的独立性知，出现每一个可能的有利记录的概率为p^kq^{n-k}，其中$q=1-P$，容易知道每一个有利记录都是互不相容的。全部有利记录个数共有$\binom{n}{k}$个，这样，如以$b(k,n,p)$表示在n次贝努里试验中A恰好出现k次的概率，则有

$$b(k,n,p)=\binom{n}{k}p^kq^{n-k}$$

$$k=0,1,\cdots n \tag{1-16}$$

几何分布　在贝努里试验序列中首次成功恰好出现在第k次试验这一事件的概率分布称为几何分布。若以A表示成

功，则上述事件发生当且仅当出现这样的试验记录：$\underbrace{\bar{A}\cdots\bar{A}}_{k-1}A$，

若以$g(k,p)$表示几何分布，则有

$$g(k,p)=q^{k-1}p,\quad k=1,2\cdots \tag{1-17}$$

k称为首次成功的等待次数。

求解二项分布的方法，可以通过查二项分布表得到，但实际中，不能想象能造出过于庞大的表，所以常运用近似计算公式。一种近似计算，运用了极限定理，我们将在稍后一点加以叙述；另一种近似公式，即所谓二项分布的普阿松逼近。

由级数的收敛性，当n很大时

$$b(k,n,p)=\binom{n}{k}p^k q^{n-k}\approx\frac{(np)^k}{k!}e^{-np} \tag{1-18}$$

式(1-18)称为二项分布的普阿松逼近，在应用上述近似公式时，要求n较大，p较小，np大小适中。

记 $$p(k,\lambda)=\frac{\lambda^k}{k!}e^{-\lambda}\qquad k=0,1,2\cdots \tag{1-19}$$

称$p(k,\lambda)$，$k=0,1,2,\cdots$为普阿松分布，$\lambda>0$为其参数。

在人类的社会生活以及自然界中许多随机现象都服从普阿松分布。尤其是一类称为随机流的随机现象。例如电话交换台到达的电话呼唤数；公共汽车站到达的乘客批数；一定范围内的车祸发生次数等都在一定条件下近似服从普阿松分布。理论上已经证明，如果随机流满足以下条件：

（1）平稳性　在任意时间区间内到达k个质点的概率只与k和时间区间的长度有关，而与初始时刻无关；

（2）无后效性　在时刻t_0之前到达的质点数与时刻t_0

之后到达的质点数，是两个相互独立的随机事件；

（3）普遍性　在充分小的时间间隔中，最多到达一个质点。

可以证明这样的随机流服从普阿松分布，即在单位时间内到达 k 个质点的概率具有形式：

$$\frac{\lambda^k}{k!}e^{-\lambda},\quad \lambda>0,\quad k=0,1,2\cdots$$

第二节　随机变量与分布函数

随机试验的结果往往与数值发生关系，有一类随机试验的样本点直接与数值对应，例如，产品检验中的次品数；还有一类随机试验与上面略为不同，例如，考察射击的命中环数时，样本点是取射击弹落点在平面上的坐标，即 $\omega=(x,y)$，但每一个样本点还是与一个命中环数值相联系。下面给出随机变量的定义。

定义　随机变量就是其取值由随机试验的样本点所确定的变量。通常以 X，Y 或 ζ，η，ξ 等表示

下面考察两类随机变量，一类称为离散型随机变量，另一类称为连续型随机变量。

（一）离散型随机变量

如果随机变量的全部可能取值只有有限个或可列个数，则称为离散型随机变量。

对于离散型随机变量 X，除了要知道它可能取哪些值之外，更重要的是要知道它取这些值的概率。其一般形式可表示为

$$\begin{pmatrix} x_1 & x_2\cdots x_n\cdots \\ p_1 & p_2\cdots p_n\cdots \end{pmatrix} \tag{1-20}$$