北京金属学会轧钢工程师技术培训班教材之一

试验研究的数理统计方法

北京金属学会

一九八三年五月

试验研究的数理统计方法

董德元　杨　节　苏敏文

江苏工业学院图书馆
藏　书　章

TG33 ⅹ-1

前　　言

试验研究是研究工作中根本手段之一，对于生产过程的研究来讲，它更是必经的途径。这是因为：第一，通过用物理、化学基本理论分析所建立的模型多数是理论结构型的，尚需用试验来确定其中某些系数或关系，还有，一般来讲，理论模型也要试验来验证；第二，因为生产过程的复杂性，建立严谨的理论型模型是困难的，必要的信息和关系需要用试验研究来获得。

试验安排和结果分析等问题都是属于试验设计的范畴。试验设计是统计学的一个重要分支。近５０年来它有了很大的发展，它的内容丰富，应用广泛，表现出比较强的生命力。

为了使读者较为深入地理解试验研究的数理统计方法理论和方法，在讲义的前面加了一部分概率统计知识，熟悉这方面内容的同志可以跳过去。

另外，讲义中的公式、符号、说明等方面都尽量与中国科学院数学研究所概率统计汇编《常用数理统计表》〔１〕相统一，同时，在这里也不再另外附加有关数表。这样做的目的很清楚，因为各书的数表有时不统一，对于非数学专业的人来讲，往往引起迷惑和差错。《常用数理统计表》编写的比较全，清楚简便，附有说明，是一本比较好的工具书籍。

本讲义不是一本数学书籍，也不是一本结合某些工业专题的研究介绍，而是一本有关理论的解说书籍，目的是为了给希望从事试验研究工作的，但又缺少有关理论知识及方法的人们提供一本入门书籍。因此，它的对象不是数学专业的学生，而是科技工作者和部分工科学生。

为了上述目的，讲义的侧重点在于解释和说明有关理论和方法，在方法上比较多地采用了课堂模拟试验的方法，即从抽样中建立理论和解释理论的方法。一般非数学专业的同志，在学习概率统计时，因为没有集合论等高等数学方面的知识，有跳跃性太大的感觉，同时也存在理论结合实际的困难。采用上述方法，对于目前在工作岗位的读者可能是有益的。

另外，为了每一部分有相对的独立性，便于阅读，在内容上保留了一部分重复。

目　　　　录

（概率统计知识）

此为试读，需要完整PDF请访问：www.ertongbook.com

第六章 极线定理及抽样

第七章 抽样分布

第八章　参数估计

第九章　假设检验

目　　录

10　一元线性回归

11　一元线性回归分析

12　子样相关系数r的检验（相关分析）

16 自适应算法

第一章 概　　率

1，1，随机试验

1.1.1 几个模拟的母体

科学试验的目的是探讨被研究对象。在这里把被研究的对象直观地叫做母体（或叫总体），从母体中抽取一部分个体、这一部分个体就叫子样。

为了理解课程内容和学者自己动手做一些试验，在这里提供几组数据，权且把这几组数据当做母体。

1 #母体：如果投一枚硬币，出现可能是正（1）面或反（0）面，一次一次投下去，并按次序纪录下来，比如0，1，0，0，1，1，0，1，1，1，0，0，…把“正”和“反”推广一下，比如，“正品”“次品”，这就成了工厂里产品的质量等级。

下面有100支晶体管，“1”代表正品，“0”代表次品，堆放在货架上，如表1，1，把这堆数据当做1#母体。

100支晶体管的等级（1#母体）

表1.1

	0	1	2	3	4	5	6	7	8	9
0	1	1	1	1	1	1	1	1	1	0
1	1	1	1	1	1	1	0	1	1	1
2	1	1	1	0	1	1	1	1	1	1
3	0	1	1	1	1	1	1	1	1	1
4	1	1	1	1	1	1	1	0	1	1
5	1	1	1	1	0	1	1	1	1	1
6	1	0	1	1	1	1	1	1	1	1
7	1	1	1	1	1	1	1	1	0	1
8	1	1	1	1	1	0	1	1	1	1
9	1	1	0	1	1	1	1	1	1	1

2 #母体：用碳化钨粉压制，烧结成的合金模具，按尺寸及性能分为“1”“2”和“3”，表中1，2，3是等级数，把第二组数据当做2 #母体。

1 #和2 #母体的数据都是整数，是可数的。

此为试读，需要完整PDF请访问：www.ertongbook.com

100块模具等级表（2#母体）

表1.2

	0	1	2	3	4	5	6	7	8	9
0	1	2	1	1	1	2	1	1	1	3
1	1	1	2	1	1	1	3	1	2	1
2	1	1	1	3	1	2	1	1	1	2
3	3	1	2	1	1	1	2	1	1	1
4	1	1	1	2	1	1	1	3	1	2
5	2	1	1	1	3	1	2	1	1	1
6	1	3	1	2	1	1	1	2	1	1
7	2	1	1	1	2	1	1	1	3	1
8	1	2	1	1	1	3	1	2	1	1
9	1	1	3	1	2	1	1	1	2	1

3 #母体：第三组数据是每班在初轧机上用固定工艺规程轧制名义尺寸为２５０×１３００的板坯。表中数据为“宽差”，即板坯尺寸与名义尺寸之差，比如，板坯的实际尺寸为１３０１，则其宽差为１３０１－１３００＝１，当然，这里是为了简化取了整数，因为１ ^{mm}可能是１．１ ^{mm}，也可能是１．１０５ ^{mm}等。

100个板坯宽差尺寸（3号母体）

表1.3

	0	1	2	3	4	5	6	7	8	9
0	－3	－7	5	3	－1	0	－5	－6	2	1
1	－1	1	－5	9	－3	7	－7	3	5	－9
2	1	－3	0	－9	11	－2	3	－5	－6	10
3	10	－3	1	－5	－11	5	－7	－2	4	－1
4	－13	1	－1	－2	1	2	4	－4	8	0
5	8	－4	－6	0	4	－2	0	2	－10	6
6	－4	－1	－8	2	－6	0	6	11	－3	8
7	6	－8	4	6	8	－4	3	0	0	－2
8	0	7	1	5	－5	3	－2	2	－7	－3
9	－3	0	9	4	2	－10	7	－9	6	－5

4 #母体：表1.4是误差的纪录，把它当做4 #母体，它与3 #都是占据了实数轴上的可能取值、

如果研究的对象的范围是某种工艺对晶体管、模具、宽差、误差的影响，那么100个数也是一部分数据，也就是一个子样，因为，生产还要继续下去。如果研究的范围就是限定在100个数据之内，那么就可以认为它们都是母体了。

表1.4

	0	1	2	3	4	5	6	7	8	9	
0	0.2	−2.0	0.6	−0.4	1.6	−0.2	0.4	−0.6	1.2	−0.6	0
1	−1.2	0.6	−0.4	0.2	−1.4	1.4	−0.2	0	0.4	0.6	1
2	−0.2	1.6	0.4	−2.4	0.6	−0.4	0	0.2	−1.0	1.2	2
3	0.6	−0.6	−2.0	1.3	−0.6	0.6	1.0	−0.6	0	0.2	3
4	0.6	0	−0.2	1.2	0.4	1.0	−1.2	−1.4	0.2	−0.6	4
5	0	−0.4	1.8	−1.0	0.6	−0.6	0.6	−1.6	−1.2	1.4	5
6	−1.0	1.0	0	−0.4	−0.4	0.2	−0.6	0.6	−0.2	0.6	6
7	−0.4	−0.6	1.0	0	1.2	−1.0	2.4	0.4	−1.2	−1.8	7
8	1.4	0.2	−0.2	0.6	−0.6	0.4	−1.8	1.0	−0.2	−1.0	8
9	−0.2	0.4	−0.6	0.2	−1.4	−1.6	−0.6	2.0	2.0	0	9
	0	1	2	3	4	5	6	7	8	9	

母体的实际情况有时为人们所熟知，有时人们不了解它，而且大多数母体是我们不清楚的。为了今后使用这些母体，下面对它们加以说明：

把1号母体中所有个体直接列数（按0，1）就可以知道“0”是10个，“1”是90个，就是说，在1 #母体中“0”占10%，“1”占90%，或者与$P=0.9$，$q=0.1$，

$$p+q=1 \qquad (1.1)$$

可以设想，投掷一枚均匀硬币，将为$p=0.5$，$q=0.5$，$p+q=1$、

2 #母体也可以很容易用列数方法知道：$p_1=0.70$，$p_2=0.20$，$p_3=0.1$。即

等级	1	2	3
p	0.7	0.2	0.1

（1.2）

3 #和4 #母体也可以用上述列数的办法求得各种“宽差”和“误差”所占的比率。但有时会遇到困难。特别是数据少时，数据将过分分散，因为它们可以在全实数轴取值，较好的办法是采用分组的办法。比如，−13.5～−11.5算一组，并且用该组的上下

限的平均数Xi当作该组的代表值——叫做组中值，比如1用

$$[-13.5+(-11.5)]/2=-12.5$$

表1—5

(事件)组号	区间 $a_i \sim a_{i+1}$	组中值 x_i	频数	频数 n_i	比率 n_i/n
1	−13.5～−11.5	−12.5	一	1	0.01
2	−11.5～−9.5	−10.5	下	3	0.03
3	−9.5～−7.5	−8.5	正	5	0.05
4	−7.5～−5.5	−6.5	正下	8	0.08
5	−5.5～−3.5	−4.5	正正	10	0.10
6	−3.5～−1.5	−2.5	正正下	13	0.13
7	−1.5～0.5	−0.5	正正正	15	0.15
8	0.5～2.5	1.5	正正下	13	0.13
9	2.5～4.5	3.5	正正一	11	0.11
10	4.5～6.5	5.5	正正	9	0.09
11	6.5～8.5	7.5	正丅	6	0.06
13	8.5～10.5	9.5	正	4	0.04
	10.5～12.5	11.5	丅	2	0.02
				100	1.00

作为落在−13.5～−11.5之间所有数据的代表。

在这里原数据为−13.0，用的是提高一位的−13.5，这是为了避免正好是−13.0，−11.0等数据分组时不好决定归属的问题。

表1.5是3＃母体落在各区间的数字频数（n_i）和比率（n_i/n）、n是母体内的总数，

用$\frac{n_i}{n}/\triangle X \approx f(X)$作为纵坐标，$X_i$作为横坐标画出母体的分布图形（图1.1）、图中$\triangle X$是组距，它是每个方块的底边。

$\frac{n_i}{n}$是方块的面积，它代表了X_i在母体中所占比率。

$\frac{n_i}{n}/\triangle X$是方块的高度。当$\triangle X \to O$时，图形将趋于一条园滑曲线，即

$$P = \lim_{\triangle x \to o} \frac{n_i}{n}/\triangle f = f(x) \qquad 1.8$$

很明显，当$\triangle X \to O$时，

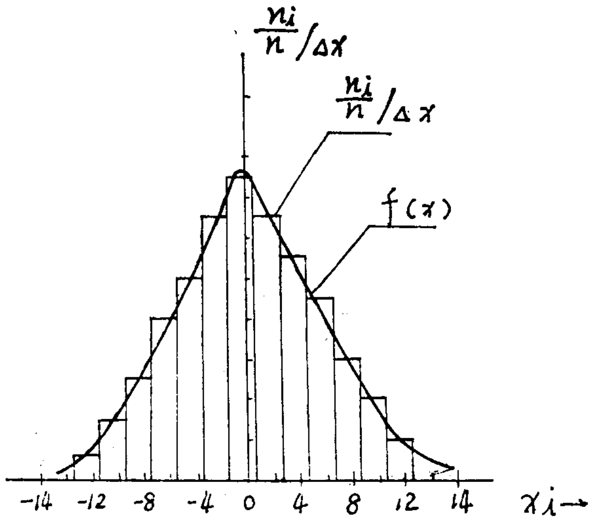

图 1.1

$$\frac{\left(\frac{n_i}{n}\right)}{\Delta X} \cdot \Delta X \rightarrow O \qquad (1.4)$$

所以,在X_i处的比率也就趋于O,意思是说，当数据在实数轴上取值时，某一个定值(比如，$X_i = 4.5$)在母体中所占的比率是零。比率是由面积给定的，而不是由高度$f_{(x)}$给定的。

1.1.2 必然性和偶然性

我们做试验，比如从1#母体中随便取一数。这个数是“0”，是“1”，并不一定，或者说，结果是偶然的。

科学主要是研究在重复试验中的不变性质。因此，在这里必须把偶然性和必然性的辨证关系给予说明。

恩格斯深刻地指出：“凡是断定为必然的东西，都是用纯碎偶然构成的。而凡被认为偶然的东西，则是一种有必然性隐藏在里面的形式。”这个科学的论断揭示了必然和偶然的辨证关系，它是认识事物的有力武器。

世界上的事物按其“发生的可能性”来看，可以分为：必然发生的和偶然发生的两大类。

如果，在某一定条件下，某个事件在试验中一定会发生，就称之为“必然事件”。如水在大气压力下，100°C时会沸腾，这就是一个必然事件。还有一些是在一定条件下不可能发生的事件，其实也是属于必然事件籍畴的。比为“等速度直线运动在无外力作用下，改变运动状态”是一个不可能事件，其实与下列说法是等价的：“在无外力作用下，不可能改变等速直线运动状态”是一个必然事件。“人不可能长生不死”也是一个必然事件。

但是在自然界中，偶然性是永远伴随着一切事物的。但在偶然性中又蕴藏着必然的规律。这是因为母体中情况所决定的。投一枚硬币出现“正面”或“反面“是不能事先预料的。但是投掷次数很大时，则将发现“正面”和反面，将是近乎相等。很明显这是因为在一枚硬币的“正面”和“反面”是各占一半，$p = P = 0.5$。

我们可认定，偶然事物的规律性主要表现在各种事物都具有自己出现的“可能性”

上。比如，1 #母体中“1 ”占90%，“0 ”占10%，那么，当由里面随便地，不受主观愿望左右地抽取一个样，虽然“1 ”，“0 ”出现是偶然的，但是可确知，在同一条件下，重复抽取子样，在偶然中会发现必然性的规律，就是出现“1 ”的比率约占90%，出现“0 ”的比率约占10%。

我们过去曾大量接触过“确定性规律”，即“动态规律”。例如：欧姆定律 $U=IR$，牛顿第二定律 $F=ma$，在这些定律中，各参数之间呈现出一定的函数关系。另外，也存在着很多的“统计规律”，例如，由无数气体分子所组成的总体中，尽管每个分子的运动轨迹、速度等都是带有偶然性的，但从总体来观察，将发现，气体的压力，则是稳定在一定值上。

实际上，在做研究“动态规律”的试验时，必然伴有偶然现象。比如，按欧姆定律做试验观测时，每次试验结果也必有一定的波动，这是因为在试验过程中，不可避免地伴有其它因素影响的缘故。不过，在误差不大时，一般是牺牲了偶然性误差所影响的精度而已。但必须指出，有时这种牺牲是不能允许的，有时要求我们把这种误差分离出来，并作必要的估计。

1.1.3 随机试验

通过前面的介绍，可以对母体、随机试验、随机事件等作为如下的直观陈述。

母体　母体是指研究对象的全体，它所包含的各种个体，是按一定机会而出现。

随机试验　在同一组条件实现的情况下，所得到的结果，不一定是相同的，然而，每一个可能的实验结果都有一定的出现机会，这种试验叫作“随机试验”。

随机事件　随机试验的结果叫做“随机事件”。

比如，从2 #母体中，任意抽取一个子样，它可能是“2 ”或“3 ”，其结果是不定的。但是，在母体中的“1 ”，“2 ”，“3 ”都有各自的比率，因而它们的出现必然是按一定机会的，因此，从2 #母体中，随机地抽取子样，便是随机试验。

从2 #母体中，每次抽一个数（子样含量为1）则出现结果是“1 ”，“2 ”，“3 ”等都是随机事件，如果每次抽2个数，那么它的结果必然是（1，2）、（2，3）、（3，1）、（1，1）、（2，2）、（3，3）六者之一，（1，2）、（2，3）或（3，1）等都是随机事件（在这里没有要求出现先后的次序）。

还有一点需要说明，就是“在一组条件实现的情况下”，意思是说，试验是在相同条件下（可重复的），第一次试验不影响第二次试验的结果（随机独立的）。因此，可以把随机试验归结为恒等的，可重复的并随机独立的试验。

可重复的随机独立试验的可能性是科学中和概率论中一个基本假定：任何一个试验都可以一次一次地重复进行，已有的和现在结果的任何意识都不影响未来试验的结果。

实际上，为了模型化，我们是把上述随机试验理想化了。在现实生活中，真正的重复试验难做到，同样，一个真正均匀的硬币也是不可能的，正如现实生活中没有一个实

物绝对符合园的方程是一样的。

下一步，我们将把事件概括为“一群点”来加以描述。

1.2 事 件

人类对自然界的认识过程中，最基本的概念是事件的概念。它概括了一切事物的结果。虽然事件的定义和术语在概率论中是不够确切的一个，但它却是由随机试验的结果过渡到数学抽象的一个重要的概念。同时事件的抽象概念只是涉及到它的发生或者不发生。

为了描述概率的数学模型，下面将对事件进一步加以说明，并且使它抽象化。说明事件的根本原则是设法将一切可能的结果都要包括到一个空间里去。

1.2.1 点集

基本事件是指在一定研究范围内不能再分的随机试验结果。从数学上说，可以把它当做空间一个点 ω，或者叫做一个元素。

基本事件空间是指基本事件所构成的集合。抽象来讲，它是全部点 ω 所构成的集合。它记做$\Omega=\{\omega\}$

例1.1 投掷一枚硬币，它有两个结果，一个是“正”，一个“反”。“正”和“反”都是基本事件，即 $\omega_1=$正，$\omega_2=$反。由“正和反”两个点组成一个集合，就是投一枚硬币的基本事件空间，记做$\Omega=\{$正，反$\}$或者写成$\Omega=\{\omega$：正、反$\}$

例1.2 从2#母体中抽取一个数，它的不可能再分的结果只可能是“1”，“2”或者“3”。“1”，“2”，“3”都是基本事件，并且由这三个点集合成一个基本事件空间$\Omega=\{1,2,3\}$

例1.3 设有几个试样，每个试样上都标号，即1，2，…，n等。任取其一，其结果必为1，2，…n中一个数，这n个数分别都是基本事件，而且1，2，…，n个点就形成一个集合，基本事件空间$\Omega=\{1,2,3,\cdots,n\}$

例1.4 上述的例子的基本事件空间都是由有限数目的点所组成的集合。下面是一个由无限个点组成的例子。我们要合计一个季度内某自动生产线上仪表发生故障的次数，它可能0次，1次，2次，…，∞次。在这里把不可能次数包括进去并不防碍，重要的是每个可能的基本事件都包括进去。实际上这个基本事件空间将由与正整数的数目相同的无限个点所组成。

例1.5 下面这个例子是连续性的。从生产线上抽取一块板坯，测量它的宽差，其可能结果是由$-\infty$到∞的实数轴上的任何一个点，即实数轴上任何一个点都是一个基本事件，而实数轴上的无限个全部的点的集合便是基本事件空间，$\Omega=\{\omega: -\infty\sim\infty\}$

例1.6 某随机试验的基本事件空间必须包括所有的点。$\Omega=\{\omega: 1,2\}$，它不是从2#母体中抽取子样的基本事件空间，因为它没有把点“3”包括进去。

此为试读，需要完整PDF请访问：www.ertongbook.com