第二册

《统计学》辅导

曹毓侯　主编

中国展望出版社

《统计学》辅导

第二册

曹毓侯　主编

吕月瑞　李目珍　刘兰亭　任若恩

中国展望出版社

一九八四年五月

《统计学》辅导（第二册）
曹毓侯　主编
吕月瑞　李目珍　刘兰亭　任若恩
*
中国展望出版社出版
（北京西城区太平桥大街4号）
太原新华印刷厂印刷
北京新华书店发行

开本850×1168毫米 1/32　10　印张
260千字　1984年5月　山西第1版
1984年5月　第1次印刷　1—55,000册

统一书号：4271·041　　定价：1.20元

内 容 提 要

本书是《（统计学）辅导》第二册，包括胡孝绳著《统计学》原书第十章至第十九章的内容，主要是阐述各种统计分配的基本原理、函数推导及其实际应用，分为分配函数（包括二项分配、波阿松分配、常态分配），大样本统计分配、和小样本统计分析方法（包括χ^2分配、t分配、F分配、Z分配的推导及其应用等）；其次是阐述变易数的分析方法，以及各种统计指数的编制和应用。每章后面附有思考题与练习题，书后附有思考题及练习题解答，可供大专院校师生及统计工作者学习参考。

前　　言

本书是对香港中文大学胡孝绳先生所著《统计学》的辅导资料，共分上、中、下三册，上册已于1983年3月份出版。本书第十章由任若恩提供初稿，第十一、十二、十三、十四、十五、十六、十七章由刘兰亭提供初稿，第十八、十九章由曹毓侯提供初稿，其中，第十章至第十四章由曹毓侯、李目珍、吕月瑞共同审编，第十五章由吕月瑞修改，第十六章由曹毓侯修改，第十七章由李目珍修改，最后全部由曹毓侯主笔修改定稿。思考题与练习题解答，由任若恩、王治中、庞风仙等同志编写。在编写过程中，得到山西省统计学会的大力支持和帮助，借此表示感谢。

一九八三年五月十五日

目　　录

第十章 二项分配

本章主要是阐述统计总体中，由性质变数所形成的二项分配的有关问题，其内容包括以下几个方面：

一、二项分配的涵义、性质及其公式。

二、二项分配的平均数、标准差和中位数。

三、理论或然率与观察或然率的区别和联系，以及统计或然率的概念和意义。

一、二项群体（二项总体）

统计总体可分为两大类：一类是由性质（质量）变数组成的统计总体，例如男性或女性，生存或死亡，已婚或未婚，合格品或废品等等。这种总体的各个组成单位，按其性质可归为不同性质类别中的一类。另一类是由数量变数组成的统计总体，例如温度、体重、物价、产量、产值等等。这些总体的各个组成单位，按其数量大小顺序排列，成为一条直线上的各点。

性质变数总体中的每个单位，分别属于两类性质中的一类，因此称为二项总体（即二项群体）。在二项总体中，具有某一种性质的单位占总体单位的比率（即比例），用p代表，具有另一种性质的单位所占的比率用q(即 1 − p)代表。例如，将某地出生婴儿依性别分类，男性的比率p为0.49，女性的比率q为0.51（1 − 0.49）。如将某地新生婴儿中任意抽取一个，则男孩的或然率就为49%。

二、二项分配公式

在二项总体中，每一独立事件出现或成功的或然率是P，不出现或失败的或然率是q。在两个独立事件的场合，此事件出现

正面或出现反面，共有4种情况可能发生。把两个事件分别用a与b代表，那么正反两面不同的4种组合情况如下：

ab	ab	ab	ab
反反	反正	正反	正正

第一种情况反面连续出现两次的或然率是$q\cdot q=q^2$，第四种情况正面连续出现两次的或然率是$p\cdot p=p^2$，第二和第三种情况，出现一正一反与一反一正的或然率都为pq，相加起来是$2pq$。把这4种情况的或然率相加为：

$$q^2+2qp+p^2$$

这实际上是二项式$(q+p)^2$的展开式，即：

$$(q+p)^2=q^2+2qp+p^2$$

因为$q+p=1$，而$(q+p)^2=1$，所以上面$q^2+2qp+p^2=1$。用一个掷硬币的例子说明这个规律，把硬币正面向上看做成功，反面向上看做失败，如果硬币是均匀对称的，那么正面向上与反面向上的或然率$p=q=\frac{1}{2}$。连续投二次硬币，而不出现正面（连续出现两次反面）的或然率为$(\frac{1}{2})^2$，出现一次正面和一次反面的或然率为$2(\frac{1}{2})(\frac{1}{2})$，而出现两次正面的或然率为$(\frac{1}{2})^2$。计算出来分别为$\frac{1}{4}$，$\frac{1}{2}$，$\frac{1}{4}$相加的结果正好是1。用上面推导出的公式表示：

$$(\frac{1}{2}+\frac{1}{2})^2=\frac{1}{4}+\frac{1}{2}+\frac{1}{4}$$

在三个独立事件的场合，也就是把硬币投掷三次的场合，投掷的结果按上面的方法排列出来共有8种情况：

abc	abc	abc	abc	abc	abc	abc	abc
反反反	反反正	反正正	反正反	正反反	正反正	正正反	正正正

上面八种情况可归纳为：有一种是三个全为反面，有三种为二反一正，有三种为一反二正，有一种三个全为正面。它们的或然率分别为：q^3，$3q^2p$，$3qp^2$，p^3，这种组合的或然率之和则是二项式3次方的展开式，即：

$$(q+p)^3=q^3+3q^2p+3qp^2+p^3$$

从以上分析可以看出，二项式的乘方指数所代表的就是独立事件的个数。在展开式中q与p的指数，代表出现各种正反情况的组合。各项数值则等于该项组合的或然率。

例如：投掷三个硬币，那么各种组合情况出现的或然率如下：

三个反面的或然率为：$q^3=(\frac{1}{2})^3=\frac{1}{8}$

二个反面一个正面的或然率为：$3q^2p=3(\frac{1}{2})^2(\frac{1}{2})=\frac{3}{8}$

一个反面二个正面的或然率为：$3qp^2=3(\frac{1}{2})(\frac{1}{2})^2=\frac{3}{8}$

三个正面的或然率为：$p^3=(\frac{1}{2})^3=\frac{1}{8}$

这四种情况出现的或然率之和为1，表明所有可能出现的各种情况的或然率总和应等于1。即：$(\frac{1}{2}+\frac{1}{2})^3=(\frac{1}{2})^3+3(\frac{1}{2})^2(\frac{1}{2})+3(\frac{1}{2})(\frac{1}{2})^2+(\frac{1}{2})^3=\frac{1}{8}+\frac{3}{8}+\frac{3}{8}+\frac{1}{8}=1$。

把上述的分析推广到n个事件的场合，则二项式的乘方指数应该是n次，而展开式的各项组合所发生的或然率为：

$$(q+p)^n=q^n+nq^{n-1}p+\frac{n(n-1)}{1\times 2}q^{n-2}p^2+\cdots\cdots+p^n$$

$$=C_0^nq^n+C_1^nq^{n-1}p+C_2^nq^{n-2}p^2+\cdots\cdots+C_n^n p^n$$

如用X代表成功的件数，并用连加符号Σ表示，上式可简化为

$$(q+p)^n=\sum_{x=0}^{n}C_x^nq^{n-x}p^x=\sum_{x=0}^{n}\frac{n!}{x!(n-X)!}q^{n-x}p^x\cdots\cdots(1)$$

在$p\neq q$时，上面的公式仍然是成立的。这种分配，目前在国

注：原书中186页第11行$(\frac{1}{2})^4$应改为$(\frac{1}{2})^3$

内外关于或然率论和数理统计的文献中，都叫做二项分配（布），该分配是贝努里（Benoulli）创立的，故二项分配也叫做贝努里分配，上述公式亦称为贝努里公式。

这个展开式是代表所有可能成功事件数的或然率之和，也就是从o、1、2，…一直到n件成功的或然率之和。如果只计算n个独立事件中x件成功的或然率，其一般公式为：

$$Y_x = C_x^n q^{n-x} p^x$$

将组合展开即为：

$$Y_x = \frac{n!}{x!\ (n-x)!} q^{n-x} p^x \quad \cdots\cdots (2)$$

式中：Y_x代表n个独立事件中X件成功的或然率，

X代表变数，即成功件数，取值范围为0，1，2，3，……n。

上式计算结果都是相对次数，即某一成功次数占总次数的百分数。例如，投掷5个均匀的骰子一次，要求计算只有2个1点出现的或然率和至多有2个1点出现的或然率。一骰子共有6面，所以出现1点的或然率$p=\frac{1}{6}$，没有1点出现的或然率$q=\frac{5}{6}$。投掷5个骰子只有2个1点的或然率为：

$$Y_2 = \frac{5!}{2!\ (5-2)!} \left(\frac{5}{6}\right)^{5-2} \left(\frac{1}{6}\right)^2$$

$$= 10 \left(\frac{5}{6}\right)^3 \left(\frac{1}{6}\right)^2 = 0.16$$

以上计算过程表明，出现2个1点的或然率是$\left(\frac{1}{6}\right)^2$，其他3个不是1点的或然率为$\left(\frac{5}{6}\right)^3$，系数10代表出现2个1点，3个非1点的不同组合数，所以出现2个1点的或然率应为16%。

另外，还可进一步计算至多出现2个1点的或然率，其方法

就是分别将正好出现2个1点的或然率、出现1个1点的或然率和不出现1点的或然率求出，然后相加即得。计算过程如下：

$$Y_0 = \frac{5!}{0!\ 5!}\left(\frac{5}{6}\right)^5\left(\frac{1}{6}\right)^0 = 0.4$$

$$Y_1 = \frac{5!}{1!\ 4!}\left(\frac{5}{6}\right)^4\left(\frac{1}{6}\right)^1 = 0.4$$

至多仅有2个1点的或然率：

$$Y_{(X\leq 2)} = Y_0 + Y_1 + Y_2 = 0.4 + 0.4 + 0.16 = 0.96$$

如果要求至少有2个1点（包括2个1点和超过2个1点）的或然率，就应该从所有可能发生1点的或然率中，减去出现0个1点和1个1点的或然率。因为所有可能发生1点的或然率之和为1，所以至少有2个1点的或然率为：

$$Y_{(X\geq 2)} = 1 - Y_0 - Y_1 = 1 - 0.4 - 0.4 = 0.2$$

例：一个工厂的产品有95%是没有缺陷的，现在某一顾客买了25件这个工厂的产品，问25件产品都没有缺陷的或然率是多大？至少有20件没有缺陷的或然率有多大？设P为废品率，q为合格率。

25件产品都没有缺陷的或然率为：

$$Y_{25} = C_{25}^{25}(0.05)^0(0.95)^{25}$$

$$= \frac{25!}{0!\ 25!}(0.05)^0(0.95)^{25}$$

$$= 0.2774$$

至少有20件没有缺陷的或然率为20件、21件、22件、23件、24件、25件没有缺陷的6种情况的或然率之和：

$$Y_{20} = C_{20}^{25}(0.05)^5(0.95)^{20} = 0.006$$

$$Y_{21} = C_{21}^{25}(0.05)^4(0.95)^{21} = 0.0269$$

注：原书中185页第5行 $\sum_{n=0}^{n}$，$\sum_{n=x}^{n}$ 应改为 $\sum_{x=0}^{n}$ $\sum_{x=0}^{n}$

$$Y_{22}=C_{22}^{25}(0.05)^{3}(0.95)^{22}=0.093$$

$$Y_{23}=C_{23}^{25}(0.05)^{2}(0.95)^{23}=0.2305$$

$$Y_{24}=C_{24}^{25}(0.05)^{1}(0.95)^{24}=0.365$$

$$Y_{25}=C_{25}^{25}(0.05)^{0}(0.95)^{25}=0.2774$$

$$0.9988$$

以上计算的是指n个独立事件，在一次试验中，各种成功件数的或然率，还可以进一步推广到多次试验。这只要用试验次数N去乘各种成功件数的或然率就可得出相应成功的可能次数。一般公式是：

$$Y_x = N\frac{n!}{x!\,(n-x)!}q^{n-x}p^{x}\cdots\cdots(3)$$

公式中的Y_x与公式（2）中的Y_x涵义不同。在公式（2）中Y_x代表或然率，即相对次数，是一个在0与1之间的数。公式（3）中的Y_x则为绝对次数，是一个小于N的正整数。

例：投硬币三枚，共投32次，则出现无正面、1正面、2正面，和3正面的次数分别计算如下：

$$Y_0=32\times\left(\frac{1}{2}\right)^{3}=4\text{次}$$

$$Y_1=32\times 3\left(\frac{1}{2}\right)^{2}\left(\frac{1}{2}\right)=12\text{次}$$

$$Y_2=32\times 3\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^{2}=12\text{次}$$

$$Y_3=32\times\left(\frac{1}{2}\right)^{3}=4\text{次}$$

各种成功、失败事件组合的可能次数之和为N，结合上例即4＋12＋12＋4＝32次，并可用下式表示：

$$N(q+p)^{n}=N\sum_{x=0}^{n}\frac{n!}{x!\,(n-x)!}q^{n-x}p^{x}\cdots\cdots(4)$$

以上是指理论或然率，在实际试验中，如果所需要的条件能够满足，例如骰子与硬币等等都很均匀，没有人为的主观因素影响，那么在试验次数很多时，实际发生的或然率与理论或然率十分接近，而且随着试验次数的增多而接近的程度也相应提高。

三、二项分配的性质

二项分配有以下五种性质：

（1）二项分配是不连续分配（离散分配），在这个分配中，变数 X只能取整数。用图形表示是一个次数多边形而不是一条光滑的曲线。

如：$P=\frac{1}{3}$，$n=6$，$N=729$，根据公式（3）：

$$X=0,\quad Y_0=729\frac{6!}{0!\ 6!}\left(\frac{2}{3}\right)^6=64$$

$$X=1,\quad Y_1=729\frac{6!}{1!\ 5!}\left(\frac{2}{3}\right)^5\left(\frac{1}{3}\right)^1=192$$

$$X=2,\quad Y_2=729\frac{6!}{2!\ 4!}\left(\frac{2}{3}\right)^4\left(\frac{1}{3}\right)^2=240$$

$$X=3,\quad Y_3=729\frac{6!}{3!\ 3!}\left(\frac{2}{3}\right)^3\left(\frac{1}{3}\right)^3=160$$

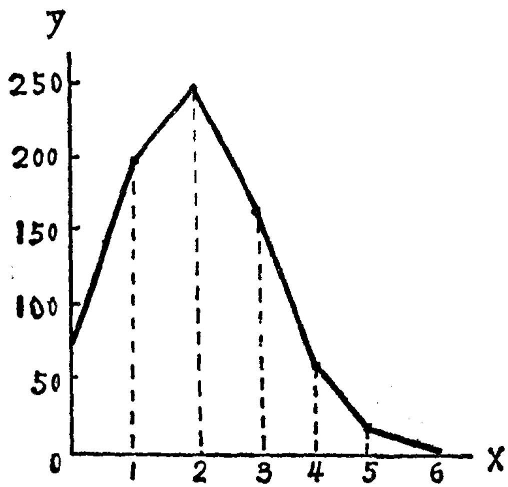
$$X=4,\quad Y_4=729\frac{6!}{4!\ 2!}\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right)^4=60$$

$$X=5,\quad Y_5=729\frac{6!}{5!\ 1!}\left(\frac{2}{3}\right)^1\left(\frac{1}{3}\right)^5=12$$

$$X=6,\quad Y_6=729\frac{6!}{6!\ 0!}\left(\frac{1}{3}\right)^6=1$$

根据以上资料，可绘制二项分配图如下，

图1．二项分配（$P=\frac{1}{3}$，$n=6$，$N=729$）



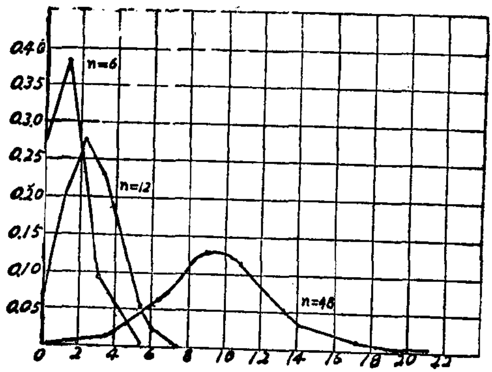
上图中横座标上X的取值，代表成功的件数，纵座标上Y的取值，代表729次试验中，成功件数为0到6件所对应的试验次数。例如次数多边形最高点横座标为2，纵座标为240。这就代表这样一种结果：由于硬币是不均匀的，正面向上的或然率只有$\frac{1}{3}$，每次投6个硬币正面向上或然率都为$\frac{1}{3}$，所以在729次试验中，2个正面4个反面的情况一共出现了240次。

（2）二项分配函数的图形，受参数p和n两个数值的变化影响。这里所说的函数是指$Y_x = C_x^n q^{n-x} p^x$所代表的函数关系。在这个函数中，自变量X取正整数，代表成功的件数，因变量Y_x代表某一种事件组合发生的或然率。只要n、p数值确定，其次数多边形的形状也就确定了。

（3）当$p=q$时，也就是$p=\frac{1}{2}$时，分配是对称的，当$p\neq q$时，分配是不对称的。当$p>q$时，右方的纵座标较高，也就是在X取值较大时，P成功的或然率较大，这种情况称为向左偏

态，即分配的图形左边拖着一个尾巴。反之，当$p<q$时，左方的纵座标较高，即在X的值较小时，P成功的或然率较小，这种情况称为向右偏态，即分配的图形右边拖着一个尾巴。但是，即使p与q的值不相等，只要n的数值增加，分配的偏态也在逐渐减弱，而变得接近对称。当n的值较小时，图形的不连续状态十分明显。这种性质，可图示如下：

图2．二项分配（0.8+0.2）^{2}代表图，当n＝6，12，48



上图画有三条折线，每一条折线，分别代表一种p与n的取值所决定的二项分配函数关系。在这三个关系中，p的取值相同，都是0.2，因此q也相同，都为0.8。n代表试验次数，分别取值为6，12、48。函数的图形随着n值的增大而逐渐平缓。

（4）二项分配的平均数，$\overline{X}=np$，标准差$\sigma=\sqrt{npq}$，公式推导过程见下一节。

（5）当n很大、p很小时，二项分配接近普阿松（Poisson）分配，而当n很大、而p取值适中时，二项分配接近常态分配。关于普阿松分配和常态分配，以及利用这两个分配计算二项分配的近似值等问题，将在后面各章加以介绍。

例：有10个工人操作机床间歇性地使用电力，由于都是间歇性地使用电力，所以就不必要按10个工人同时使用电力来规定总的负荷，而只要规定一个保险程度较高的总负荷就可以了。假定这10个工人的工作是相互独立的，而且都以同样的或然率p需