GB

中国国家标准汇编

中 国 国 家 标 准 汇 编

38

GB 4087 ~ 4111

中 国 标 准 出 版 社

1 9 8 9

中国国家标准汇编
38
GB 4087～4111
中国标准出版社总编室 编
*
中国标准出版社出版
（北京复外三里河）
中国标准出版社秦皇岛印刷厂印刷
新华书店北京发行所发行 各地新华书店经售

*
开本 880×1230 1/16 印张 45 1/4 字数 1 368 000
1989年 8 月第一版 1989年 8 月第一次印刷
印数 1—8,500〔精〕 2,800〔平〕 定价 24.90 元〔精〕 20.60 元〔平〕
*
ISBN 7-5066-0205-9/TB·067〔精〕
ISBN 7-5066-0206-7/TB·068〔平〕
*
标目 115—14〔精〕 115—13〔平〕

出 版 说 明

《中国国家标准汇编》是一部大型综合性工具书，从1983年起，分若干分册陆续出版。本汇编在一定程度上反映了我国建国以来标准化事业发展的基本情况和主要成就，是各级标准化管理机构及工矿企事业单位，农林牧副渔系统，科研、设计、教学等部门必不可少的工具书。

本汇编收入公开发行的全部现行国家标准，按国家标准号顺序编排。凡遇顺序号短缺，除特殊注明外，均为作废标准号或空号。

本分册为第38分册，出版精装本和平装本，收入国家标准GB 4087～4111。本分册以1988年12月底为限，收入了现行标准的最新版本。由于标准不断修订，请读者在使用和保存本汇编时，注意及时更换修订过的标准。

中国标准出版社除出版《中国国家标准汇编》外，还出版国家标准、专业（部）标准的单行本，及各种专业标准汇编，以满足不同读者的需要。

中国标准出版社
一九八九年二月

目　　录

中华人民共和国国家标准

UDC 519.25

GB 4087.1—83

数据的统计处理和解释
二项分布参数的点估计

Statistical interpretation of data
Point estimation of parameter in binomial distribution

1 引言

1.1 本标准所用统计学名词见国标GB 3358—82《统计学名词及符号》。

1.2 设总体中的部分个体具有某种特性，p是总体中具有此种特性的个体的比率。例如p可以是一批产品中不合格品的比率。从总体中随机地、独立地抽取若干个个体作为样本。本标准规定了基于这类样本，对总体的参数p作点估计的方法。

1.3 对有限总体，设其大小为N，样本大小为n。当抽取是有放回时，或当抽取是无放回的，但$\frac{n}{N}\leqslant 0.1$时，n次抽取可以认为是独立的。

2 经典估计法

2.1 样本的抽取方式

样本大小n是事先规定的。样本从总体中随机地、独立地抽取。此时样本中具有某种特性的个体的个数x是服从二项分布的随机变量X的一次观测值。X取值x的概率为

$$P\{X=x|n,p\}=\binom{n}{x}p^{x}(1-p)^{n-x}\quad, x=0,1,2,\cdots, n$$

2.2 估计量

p的估计量记为$\hat{p}$

$$\hat{p}=\frac{x}{n}\qquad\cdots\cdots(1)$$

式中：n——样本大小。其确定方法可参见附录A(参考件)公式（A1）。

x——样本中具有指定特性的个体的个数。

2.3 例子

为估计一批产品（约1000件）的不合格品率，从中随机地抽取40件作为样本，其中有5件不合格品。则

$$n=40$$
$$x=5$$
$$\hat{p}=\frac{x}{n}=\frac{5}{40}=12.5\%$$

3 序贯样本估计法

3.1 样本的抽取方式

国家标准局1983－12－21发布　　　　1984－10－01实施

事先不规定样本大小，而是从总体中顺序地、随机地、独立地抽取样本。每抽取一个样本，立即检查该个体是否具有指定的特性，不断累计具有指定特性的个体的个数，当具有指定特性的个体数达到事先规定的数c（大于或等于2的整数）时，停止抽样。此时，累计抽取的个体的总个数n是一个服从负二项分布的随机变量，n取值k的概率为

$$p\{n=k\,|\,c,p\}=\binom{k-1}{c-1}p^{c}(1-p)^{k-c}\ ,\quad k=c,c+1,\cdots$$

3.2 估计量

$$\hat{p}=\frac{c-1}{n-1}\qquad\qquad(2)$$

式中：c——事先规定的具有指定特性的个体达到的个数。其确定方法可参见附录A（参考件）公式（A2）。

n——达到c个具有指定特性的个体时，累计抽取的个体总数。

3.3 例子

为估计一批产品（约1000件）的不合格品率，顺序地逐个抽样检查，规定发现5件不合格品时停止抽样，当抽到第35件时，发现了第5件不合格品，则

$$c=5$$

$$n=35$$

$$\hat{p}=\frac{c-1}{n-1}=\frac{4}{34}=11.8\%$$

附 录 A
根据对准确度的要求确定n及c的方法
（参考件）

当希望差不多以$1-\alpha$的概率保证所得的点估计$\hat{p}$与被估值 p 的绝对差不超过δ，即

$$P\{|\hat{p}-p|\leqslant\delta\}\approx 1-\alpha$$

时，可根据以往的记录或经验确定一个p的粗估值p_0。此时，经典估计法中的n可如下确定：

$$n\approx\left(\frac{u_{1-\alpha/2}}{\delta}\right)^2 p_0(1-p_0) \qquad \cdots\cdots(A1)$$

序贯样本估计法中的c可如下确定：

$$c\approx\left(\frac{u_{1-\alpha/2}}{\delta}\right)^2 p_0^2(1-p_0) \qquad \cdots\cdots(A2)$$

式中：α——由所要求的保证概率$1-\alpha$确定；

$u_{1-\alpha/2}$——标准正态分布的$1-\alpha/2$分位数。

附 录 B
其它几种估计方法
（参考件）

除标准正文中的估计方法之外，这里再给出两种估计方法。在使用者各方协商一致和主管部门同意的情况下，可以采用这些估计方法。

B.1 贝叶斯估计

B.1.1 使用条件

掌握p的先验知识：p服从β分布，其概率密度函数为

$$f(x)=\begin{cases}\dfrac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1} & \text{当}0<x<1\\ 0 & \text{当}x<0\text{或}x>1\end{cases}$$

式中a，b为未知参数，并且已知p的经验的均值u与方差v。例如当有大批以往的可靠的p的数值记载时，可根据这些历史资料算出p的经验均值u与方差v。

B.1.2 样本的抽取方式

样本大小是事先规定的。样本从总体中随机地、独立地抽取。

B.1.3 估计量

由u与v计算a、b的数值如下

$$a=u\left[\frac{u(1-u)}{v}-1\right] \quad\cdots\cdots\cdots\cdots (B1)$$

$$b=(1-u)\left[\frac{u(1-u)}{v}-1\right] \quad\cdots\cdots\cdots\cdots (B2)$$

则

$$\hat{p}=\frac{x+a}{n+a+b} \quad\cdots\cdots\cdots\cdots (B3)$$

式中：n——样本大小；

x——样本中具有指定特性的个体的个数。

B.2 极小极大估计

B.2.1 使用条件

当样本大小n比较小，并且p值在1/2左右时，经典估计会引起较大的均方误差。这时采用极小极大估计能使极大均方误差达到极小值。

B.2.2 样本的抽取方式

样本大小n是事先规定的。样本从总体中随机地、独立地抽取。

B.2.3 估计量

$$\hat{p}=\frac{x+\frac{1}{2}\sqrt{n}}{n+\sqrt{n}} \qquad \text{(B4)}$$

式中：n——样本大小；

x——样本中具有指定特性的个体的个数。

附加说明：

本标准由全国统计方法应用标准化技术委员会提出。

本标准由全国统计方法应用标准化技术委员会数据的处理和解释分委员会工作组起草。

本标准主要起草人孙山泽、于秀林、郑忠国。

中华人民共和国国家标准

UDC 519.25

数据的统计处理和解释
二项分布参数的区间估计

GB 4087.2—83

Statistical interpretation of data
Interval estimation of parameter in binomial distribution

1 引言

1.1 本标准所用统计学名词见国标GB 3358—82《统计学名词及符号》。

1.2 设总体中部分个体具有某种特性，p是总体中具有此种特性的个体的比率。例如p可以是一批产品中不合格品的比率。从总体中随机地、独立地抽取若干个个体作为样本。本标准规定了基于这类样本，对总体的参数p作区间估计的方法。

1.3 对有限总体，设其大小为N，样本大小为n。当抽取是有放回时；或当抽取是无放回的，但$\frac{n}{N}<0.1$时，n次抽取可以认为是独立的。

1.4 在n个随机地、独立地抽取的个体中，具有某种特性的个体的个数x是服从二项分布的随机变量X的一次观测值。X取值x的概率为

$$P\{X=x \mid n, p\} = \binom{n}{x} p^x (1-p)^{n-x}, \quad x=0,1,2,\cdots,n$$

2 比率 p的双侧置信区间和单侧置信区间

p的双侧置信区间是(p_L, p_U)，这里$0 \leqslant p_L < p_U \leqslant 1$。$p_L$称为下置信限，$p_U$称为上置信限。

p的单侧置信区间有两种形式：

a. 仅有下置信限p_L（$0 \leqslant p_L < 1$）的置信区间（p_L，1〕。

b. 仅有上置信限p_U（$0 < p_U \leqslant 1$）的置信区间〔0，p_U）。

这些区间必然包含p的点估计$\hat{p}=\frac{x}{n}$。

选用哪种类型的区间，要根据具体问题的性质而定，不依赖于观测值x。

本标准采用的置信区间所满足的条件为：所求得的置信区间以给定的$1-\alpha$的概率包含真正的p值。具体地说，双侧区间应满足

$$P\{p_L < p < p_U\} = 1-\alpha$$

单侧区间应满足

$$P\{p_L < p\} = 1-\alpha$$

$$\text{或} \quad P\{p < p_U\} = 1-\alpha$$

概率$1-\alpha$称为置信水平。根据不同的要求，$1-\alpha$的数值通常从0.90，0.95，0.99中选取。

由于二项分布的离散性，不一定都能求得使上述等式成立的p_L，p_U，这时可取p_L，p_U满足

$$P\{p_L < p < p_U\} \geqslant 1-\alpha \quad \text{（双侧情形）}$$

$$\text{或} \quad P\{p_L < p\} \geqslant 1-\alpha \quad \text{（单侧下限）}$$

$$\text{或} \quad P\{p < p_U\} \geqslant 1-\alpha \quad \text{（单侧上限）}$$

这个置信区间包含真值p的实际概率称为这个区间的"置信水平的实际值"。

国家标准局1983－12－21发布　　　　1984－10－01实施

·3 置信区间的求法

3.1 当$n<10$时，置信区间一般太宽，无实际使用价值。

3.2 当$n\geqslant 10$时，下面的公式给出了置信水平为$1-\alpha$的置信限。

$$p_L=\frac{\nu_2}{\nu_2+\nu_1 F_{1-\alpha}(\nu_1,\ \nu_2)}\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\ (1)$$

式中：$\nu_1=2\ (n-x+1)$，$\nu_2=2x$；

$$p_U=\frac{\nu_2}{\nu_2+\nu_1/F_{1-\alpha}(\nu_2,\ \nu_1)}\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\ (2)$$

式中：$\nu_1=2\ (n-x)$，$\nu_2=2\ (x+1)$。

此处的$F_{1-\alpha}(\nu_1,\ \nu_2)$是自由度为$\nu_1$，$\nu_2$的$F$分布的$1-\alpha$分位数，它的值可由$F$分布表（见国标GB 4086.4—83《统计分布数值表 F分布》）中查得。

为了方便使用，本标准给出了当$10\leqslant n\leqslant 30$时，相应于不同置信水平的置信限的数值表〔见附录A（补充件）〕。

单侧置信区间或双侧置信区间的上置信限p_U的值可从表中直接读出。

对于单侧置信区间或双侧置信区间的下置信限p_L，只需用$x'=n-x$代替x，从表中读出相应于x'的p值，$(1-p)$就是相应于x的p_L值。

例：取$n=20$，$x=8$，$1-\alpha=0.95$，

a. 求单侧置信区间〔0，p_U）。

$n=20$，$x=8$；

从表中读出与n，x相应的$p=0.606$，

$p_U=p=0.606$

单侧置信区间为〔0，0.606）。

b. 求单侧置信区间（p_L，1〕。

$n=20$，$x=8$，$x'=n-x=20-8=12$；

从表中读出与n，x'相应的$p=0.783$，

$p_L=1-p=1-0.783=0.217$

单侧置信区间为（0.217，1〕。

c. 求双侧置信区间（p_L，p_U）。

$n=20$，$x=8$，$x'=n-x=12$；

从表中读出与n，x相应的$p=0.639$，

$p_U=p=0.639$；

从表中读出与n，x'相应的$p=0.809$，

$p_L=1-p=0.191$

双侧置信区间为（0.191，0.639）。

3.3 当$n>30$，且$0.1<\dfrac{x}{n}<0.9$时，置信水平为$1-\alpha$的置信限由下面的近似公式给出

$$p_L=p_*-u\sqrt{p_*\ (1-p_*)\ /\ (n+2d)}\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\ (3)$$

$$p_U=p^*+u\sqrt{p^*\ (1-p^*)\ /\ (n+2d)}\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\ (4)$$

式中：

$$p_*=\frac{a+d-0.5}{n+2d}$$

$$p^*=\frac{a+d+0.5}{n+2d}$$

u，d为常数。对不同的置信水平，u，d的取值见下表。

表1

置信水平 $1-\alpha$	单侧		双侧	
	u	d	u	d
0.90	1.282	0.7	1.645	1
0.95	1.645	1	1.960	1.5
0.99	2.326	2	2.576	2.5

例：取$n=40$，$x=12$，$1-\alpha=0.95$，

a. 求单侧置信区间〔0，p_U）。

$1-\alpha=0.95$ 查表1得 $u=1.645$，$d=1$，

$n=40$，$x=12$，

$$p^*=\frac{x+d+0.5}{n+2d}=\frac{12+1+0.5}{40+2\times 1}=0.321$$

$$\sqrt{p^*(1-p^*)/(n+2d)}=0.072$$

$$p_U=p^*+u\sqrt{p^*(1-p^*)/(n+2d)}$$
$$=0.321+1.645\times 0.072=0.439$$

单侧置信区间是〔0，0.439）。

b. 求单侧置信区间（p_L，1〕。

$1-\alpha=0.95$，查表1得 $u=1.645$，$d=1$，

$n=40$，$x=12$

$$p_*=\frac{x+d-0.5}{n+2d}=\frac{12+1-0.5}{40+2\times 1}=0.298$$

$$\sqrt{p_*(1-p_*)/(n+2d)}=0.071$$

$$p_L=p_*-u\sqrt{p_*(1-p_*)/(n+2d)}$$
$$=0.298-1.645\times 0.071=0.181$$

单侧置信区间是（0.181，1〕。

c. 求双侧置信区间（p_L，p_U）。

$1-\alpha=0.95$，查表1得 $u=1.960$，$d=1.5$

$n=40$，$x=12$

$$p_*=\frac{x+d-0.5}{n+2d}=\frac{12+1.5-0.5}{40+2\times 1.5}=0.302$$

$$p^*=\frac{x+d+0.5}{n+2d}=\frac{12+1.5+0.5}{40+2\times 1.5}=0.326$$

$$\sqrt{p_*(1-p_*)/(n+2d)}=0.070$$

$$p_L=p_*-u\sqrt{p_*(1-p_*)/(n+2d)}$$
$$=0.302-1.960\times 0.070=0.165$$

$$\sqrt{p^*(1-p^*)/(n+2d)}=0.071$$

$$p_U = p^* + u\sqrt{p^*(1-p^*)/(n+2d)}$$
$$= 0.326 + 1.960 \times 0.071$$
$$= 0.466$$

双侧置信区间是（0.165，0.466）。

3.4 当 $n>30$，且 $\frac{x}{n}\leqslant 0.1$ 或 $\frac{x}{n}\geqslant 0.9$ 时，可采用泊松近似。这种近似需要利用 χ^2 分布表（见国标 GB 4086.2—83《统计分布数值表 χ^2分布》）。这时下置信限为

$$p_L = \begin{cases} \dfrac{2\lambda}{2n-x+1+\lambda} & \text{当}\dfrac{x}{n}\text{接近于 }0 \\ \dfrac{n+x-\lambda'}{n+x+\lambda'} & \text{当}\dfrac{x}{n}\text{接近于 }1 \end{cases} \qquad \cdots\cdots\cdots\cdots (5)$$

对于单侧置信区间，式中

$$\lambda = \frac{1}{2}\chi^2_{\alpha}(2x)$$

$$\lambda' = \frac{1}{2}\chi^2_{1-\alpha}[2(n-x)+2]$$

对于双侧置信区间，式中

$$\lambda = \frac{1}{2}\chi^2_{\alpha/2}(2x)$$

$$\lambda' = \frac{1}{2}\chi^2_{1-\alpha/2}[2(n-x)+2]$$

上置信限为

$$p_U = \begin{cases} \dfrac{2\lambda}{2n-x+\lambda} & \text{当}\dfrac{x}{n}\text{接近于 }0 \\ \dfrac{n+x+1-\lambda'}{n+x+1+\lambda'} & \text{当}\dfrac{x}{n}\text{接近于 }1 \end{cases} \qquad \cdots\cdots\cdots\cdots (6)$$

对于单侧置信区间，式中

$$\lambda = \frac{1}{2}\chi^2_{1-\alpha}(2x+2)$$

$$\lambda' = \frac{1}{2}\chi^2_{\alpha}[2(n-x)]$$

对于双侧置信区间，式中

$$\lambda = \frac{1}{2}\chi^2_{1-\alpha/2}(2x+2)$$

$$\lambda' = \frac{1}{2}\chi^2_{\alpha/2}[2(n-x)]$$

这里，$\chi^2_{\alpha}(\nu)$ 表示自由度为 ν 的 χ^2 分布的 α 分位数。

例：取 $n=50$，$x=5$，$1-\alpha=0.95$，

a. 求单侧置信区间 $[0, p_U)$。

$n=50$，$x=5$，$\frac{x}{n}=0.1$，用接近于零的公式。

$$\lambda = \frac{1}{2}\chi^2_{1-\alpha}(2x+2) = \frac{1}{2}\chi^2_{0.95}(12)$$

$$= \frac{1}{2}\times 21.026 = 10.513$$

$$p_U=\frac{2\lambda}{2n-x+\lambda}=\frac{2\times10.513}{2\times50-5+10.513}=0.199$$

单侧置信区间是〔0，0.199）。

b. 求单侧置信区间（p_L，1〕。

$n=50$，$x=5$，$\frac{x}{n}=0.1$，用接近于零的公式。

$$\lambda=\frac{1}{2}\chi^2_{\alpha}(2x)=\frac{1}{2}\chi^2_{0.05}(10)$$

$$=\frac{1}{2}\times3.940=1.970$$

$$p_L=\frac{2\lambda}{2n-x+1+\lambda}=\frac{2\times1.970}{2\times50-5+1+1.970}=0.040$$

单侧置信区间是（0.040，1〕

c. 求双侧置信区间（p_L，p_U）。

$n=50$，$x=5$，$\frac{x}{n}=0.1$用接近于零的公式。

对p_L，有

$$\lambda=\frac{1}{2}\chi^2_{\alpha/2}(2x)=\frac{1}{2}\chi^2_{0.025}(10)$$

$$=\frac{1}{2}\times3.247=1.624$$

$$p_L=\frac{2\lambda}{2n-x+1+\lambda}=\frac{2\times1.624}{2\times50-5+1+1.624}=0.033$$

对p_U，有

$$\lambda=\frac{1}{2}\chi^2_{1-\alpha/2}(2x+2)=\frac{1}{2}\chi^2_{0.975}(12)$$

$$=\frac{1}{2}\times23.337=11.669$$

$$p_U=\frac{2\lambda}{2n-x+\lambda}=\frac{2\times11.669}{2\times50-5+11.669}=0.219$$

双侧置信区间是（0.033，0.219）。