陈幼松　杨位钦　编著

实用数理统计方法及应用题详解

北京科学技术出版社

实用数理统计方法及应用题详解

实用数理统计方法及应用题详解
陈幼松　杨位钦　编著
*
北京科学技术出版社出版
（北京西直门外南路19号）

新华书店首都发行所发行　各地新华书店经售
通县马驹桥印刷厂印刷
*
850×1168毫米　32开本　10.875印张　324千字
1988年9月第一版　1988年9月第一次印刷
印数1—2,500册

ISBN7-5304-0100-9/N.2　　定价 3.10 元

内 容 简 介

本书分为统计分布（正态、χ^2、t，F、二项、泊松、均匀等分布）及其应用、统计推断（均值、方差、比例、拟合优度及非参数检验等）、统计分析（回归、相关、质量控制、方差等分析）三部分．不做理论推导、公式证明，着重介绍其应用方法、应用条件．结合科学实验、工农业生产、经营管理等实际问题，为读者提供了大量例题和习题（占全书三分之二左右）．解题步骤层次分明、思路清晰，力求使抽象概念具体化．文字通俗浅显，便于读者理解掌握．本书特别适合各行各业中从事实际工作而需用数理统计方法的同志使用，也可供各类学校师生在概率和数理统计方法课程中作为参考书．

目　　录

第一章　统计分布及其应用

第一节　正态分布

一、概述、参数和特性

正态分布是一种连续分布。客观世界中许多随机产生的误差，都服从正态分布。根据中心极限定理知道，许多独立的随机变量共同起作用时，不管它们各自服从什么分布，只要它们所占的比例均匀地小，则它们的总和是服从正态分布的。这正是许多随机现象都服从正态分布的原因。因此，正态分布也就成为最重要的分布。

服从正态分布的随机变量X常记作$X\sim N(\mu,\sigma^2)$。其中，μ为均值、σ^2为方差。它的概率密度函数（pdf）$f(x)$为

$$f(x)=\begin{cases}\dfrac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/(2\sigma^2)} & -\infty\leqslant x\leqslant\infty;\ \sigma>0;\ -\infty\leqslant\mu\leqslant\infty\\ 0 & \text{其他}\end{cases}\tag{1.1}$$

μ表示分布曲线的对称轴在横坐标上的位置，是分布的位置参数标准差σ则表示分布曲线的散布程度，是分布的尺度参数。这两个参数的最佳估计量分别为

$$\hat{\mu}=\text{样本均值}=\overline{X}\tag{1.2}$$

$$\hat{\sigma}=\text{样本标准差}=S\tag{1.3}$$

无偏的 $S=\left[\dfrac{\sum X_i^2}{n-1}-\dfrac{n}{n-1}\overline{X}^2\right]^{1/2}$，而有偏的 $S=\left[\dfrac{\sum X_i^2}{n}-\overline{X}\right]^{1/2}$，式中$X_i$为第$i$次观察到的随机变量$X$的取值，$n$为样本容量。当$n\geqslant 30$时，有偏和无偏的$S$差别很小。

在进行统计假设检验时，经常假定总体服从正态分布。例如，在产品制造中，相对于标称尺寸、重量或容积的误差，以及动物、植物或人所具有的体态特点等均服从正态分布。

二、标准正态分布（SND）

具体情况下的正态分布，其参数μ、σ往往各不相同，为了便于利用统计表进行计算，需要将具体的正态分布化为标准的正态分布。这就要定义一个随机变量Z，使

$$Z=\frac{\text{同均值之差}}{\text{标准差}}=\frac{x-\mu}{\sigma} \tag{1.4}$$

这样，当x等于均值μ时，$z=0$；当$x-\mu=\pm\sigma$时，$z=\pm1$。故随机变量Z一定是$\mu=0$、$\sigma=1$，它的分布就称为标准正态分布，记作$Z\sim N(0,1)$或Z～SND。Z的取值范围和X一样，为$-\infty$至$+\infty$而Z的概率密度函数可由式（1.1）、（1.4）导出

$$f(z)=\begin{cases}\dfrac{1}{\sqrt{2\pi}}e^{-z^2/2} & -\infty\leqslant z\leqslant\infty\\ 0 & \text{其他}\end{cases} \tag{1.5}$$

图1.1为SND曲线，曲线下面的面积表示Z处于两给定值之间的概率，图中标明了±1、±2、±3之间的面积大小。它们分别同$N(\mu,\sigma^2)$时，对应于$\mu\pm1\sigma$、$\mu\pm2\sigma$、$\mu\pm3\sigma$之间的面积相等。例如，图1.1中±2之间的面积为0.9546，它是式（1.5）的如下积分值

$$\int_{-2}^{2} f(z)dz = \int_{-2}^{2} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{z^2}{2}\right] dz = 0.9546$$

在SND时，概率密度分布曲线对称于$\mu=0$，故可通过求从0算起的一侧面积，然后乘以2而得到两侧对称的面积。

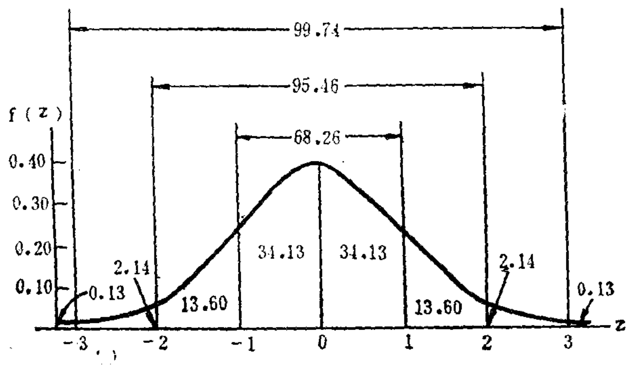

图 1.1　SND曲线下的面积百分比

在SND曲线下，从$-\infty$至给定值z所限定的面积，见附表1。表中给出概率值

$$P(-\infty \leqslant z \leqslant z) = \int_{-\infty}^{z} N(0, 1)\, dz$$

利用附表1求概率值的步骤如下：

1．写出想求的随机变量取值范围的概率表达式，例如

$P(5 \leqslant x \leqslant 10)$

2．用式（1.4）将其转换为SND时的概率表达式。例如，当$\mu=20/3$、$\sigma=5/3$时，上例可化为$P(-1 \leqslant z \leqslant 2)$。

3．考虑到附表1只给出累积分布函数的值，如果步骤2得出的表达式中含有上限和下限，或者呈P（$z \geqslant$数值）的形式，则应改写。例如，应使

$$P(-1 \leqslant z \leqslant 2) = P(z \leqslant 2) - P(z \leqslant -1)$$

$P\ (z \geqslant 2) = 1 - p\ (z \leqslant 2)$

4. 由附表1查出概率值。如

$P\ (z \leqslant 2) = 0.9773$

$P\ (z \leqslant -1) = 0.1587$

5. 进行步骤3的算术运算。如

$P\ (-1 \leqslant z \leqslant 2) = 0.9773 - 0.1587 = 0.8186$

6. 得出想求的概率。如

$P\ (5 \leqslant x \leqslant 10) = 0.8186$

三、正态分布的应用

正态分布可用以确定设计时的要求，保证产品以一定的百分比满足设计要求。

例 1.1 一种新型下料机试验结果说明，它系以 $\sigma = 0.07$ 厘米围绕均值作正态分布，试求 (*a*) 要使下料的不合格品不超过2%，对称的公差范围应当怎样确定？(*b*)规定的尺寸为12±0.02厘米时，合格品的百分比多大？

解 (*a*) 这时z值应对称地位于0的两侧，使下限 z_L和上限z_U所限定的SND曲线下面的面积为98%，即应使$P\ (z_L \leqslant z \leqslant z_U) = 0.98$。由附表1，可求得

$$z_U = 2.326 \text{和} z_L = -2.326$$

如果X为表示尺寸的随机变量，则其标称尺寸即为μ，利用SND时的变量 $Z = (x - \mu)/0.07$，将z_L、z_U化为x的上、下限。如由

$$-2.326 = (x_L - \mu)/0.07$$

得下限 $x_L = \mu - 0.163$

同样，得上限 $x_U = \mu + 0.163$

这样，有了要求的均值（即标称尺寸），便可求出 x_L 和x_U 的尺寸。

(*b*) 这时均值为12.0，而尺寸范围为$x_L = 11.98$厘米和$x_U =$

12.02厘米。利用前面介绍的由附表1求概率的方法，求

$$P(11.98 \leqslant x \leqslant 12.02) = P(-0.286 \leqslant z \leqslant 0.286)$$

最后，得 $2P(0 \leqslant z \leqslant 0.286) = 2 \times 0.1126 = 0.2252$

即尺寸处于12±0.2厘米范围内的合格品占22.5%。

在实际工作中，我们只能从无限的总体中取得有限的样本数据，需要根据这些数据来推断总体是否符合某种分布。通常，这要用下一章所介绍的统计检验方法来完成，但也可以通过拟合对比的方法来进行直观的判断。

进行拟合前，要对所获得的数据进行合理的分组，得出组距、组中值和每组数据的频数。然后按下列步骤进行同$N(\mu, \sigma^2)$曲线的拟合。

1. 用各组的组中值估计出参数$\hat{\mu}$和$\hat{\sigma}$。

2. 计算各组的$z = (x - \mu)/\sigma$，式中x为该组的上界。

3. 由z求出各组面积应占的百分比，将其作为该组的概率P。

4. 由$E =$（样本容量）×（组概率）$= nP$，求出各组的期望频数E。

5. 用数值法或图示法对各组观察到的频数和期望频数进行比较，以作出判断。

流速(升/秒) 组中值	频数
7.65	5
7.95	21
8.25	35
8.55	15
8.85	3
9.15	1
	80

例 1.2 今由计算机监控搜集到某种流程的80天流速，经分组整理得数据如右上表所列。今要 (a) 拟合成正态分布；(b) 在同一张图上画出期望频数和观察到的频数。

解 按上述步骤进行拟合，下面所提到的列次为表1.1中的列次。

1. 由式 (1.2) 求 $\hat{\mu} = (7.65 \times 5 + 7.95 \times 21 + \cdots + 9.15 \times 1)/80 = 8.224$，由式 (1.3) 求 $\hat{\sigma} = [(7.65^2 \times 5 + 7.95^2 \times 21 + \cdots + 9.15^2 \times 1)/80 - \hat{\mu}^2]^{1/2} = 0.293$

2. 由各组上界x（列2），求出其z值（列3）：

$z=(x-8.224)/0.293$。

3．由附表/求出SND时由$-\infty$到z所包含的面积（列4），减去上一次得到的面积，便得到概率P（列5）。

4．$E=80P$就是期望的频数（列6）。

5．比较的结果说明，观察到的流速同正态曲线拟合很好(图1.2）。

表 1.1 例 1.2拟合时用到的数据

组中值 (1)	上界x (2)	z (3)	z左边的面积 (4)	该组的面积 (5)	频数 期望值 (6)	频数 观察值 (7)
7.65	7.80	−1.45	0.0735	0.0735	5.88	5
7.95	8.10	−0.42	0.3372	0.2637	21.10	21
8.25	8.40	0.60	0.7257	0.3885	31.08	35
8.55	8.70	1.62	0.9474	0.2217	17.74	15
8.85	9.00	2.65	0.9960	0.0486	3.89	3
9.15	9.30	3.69	0.9999	0.0039	0.31	1
					80.00	80

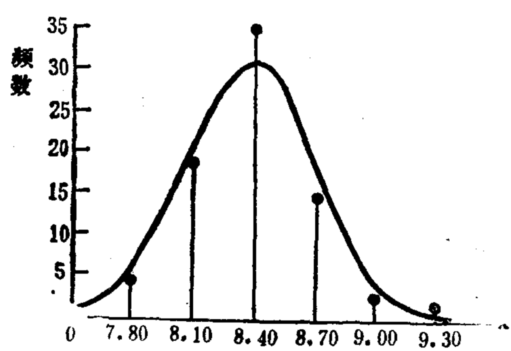

图 1.2 例1.2中频数的观察值同正态分布拟合的比较

正态分布只要做一些假定，例如随机变量X不能取负值，便可用来近似其他分布。最常见的是用以近似离散的二项分布（二

项分布详见本章第五节）。可以证明，服从二项分布的变量$X \sim b(x;\ n,\ p)$ 其$\mu = np$而$\sigma = \sqrt{npq}$，当n增大时，则变量$Z = (X - np)/\sqrt{npq}$的累积分布函数（cdf）$F(z)$的极限$[F(z) = P(-\infty \leqslant Z \leqslant z)]$，可用SND的概率密度表示如下：

$$\lim_{n \to \infty} F(z) = \int_{-\infty}^{x} N\ (0,\ 1)dx$$

在实际应用中，只要$np \geqslant 5$且$nq \geqslant 5$，则按正态分布计算的结果同真正按二项分布计算的结果便非常接近。

由于用正态分布近似二项分布时，系用连续的概率密度函数去计算离散变量的概率值，所以要±0.5作为连续性修正。例如，求二项分布时x落在a和b之间的概率，因二项分布离散变量相邻两值的间隔为1，这些离散值被当作连续正态分布的组中值，故利用正态分布计算时，对应连续变量的范围应取$a-0.5$到$b+0.5$。在实际中，还应化为SND来求解，如求二项分布时x落在a和b之间的概率，则应化为求

$$P\left(\frac{a-0.5-np}{\sqrt{npq}} \leqslant z \leqslant \frac{b+0.5-np}{\sqrt{npq}}\right) \qquad (1.6)$$

例 1.3 在一次观察中，发现250袋化肥有32袋在运输中破损。过去资料表明破损数服从二项分布且破损率为10%，今要求（a）计算破损数不超过期望值的可能性；（b）恰好破损32袋的可能性。

解 （a）先求出二项分布时的期望值$np = 250 \times 0.10 = 25$。再求破损数不超过25时的概率，用二项分布时这一概率为

$$P\ (d \leqslant 25) = \sum_{d=0}^{25} C_{250}^{d}\ \ 0.10^{d}\,0.9^{250-d}$$

真正计算非常冗烦，而且n太大无法查表。但由于$np = 25 > 5$且$nq = n(1-p) = n - np = 250 - 25 = 225 \geqslant 5$，所以可用SND近似。这时$a = 0$、$b = 25$、$\mu = np = 25$，$\sigma = \sqrt{npq} = 4.74$，故求

$$P\left(\frac{0-0.5-25}{4.74}\leqslant z\leqslant\frac{25+0.5-25}{4.47}\right)$$
$$=P\ (-5.38\leqslant z\leqslant 0.11)$$
$$=P\ (z\leqslant 0.11)\ -P\ (z\leqslant -5.38)$$
$$=0.5438$$

即250袋化肥实际破损率不超过10%的可能性为54.4%。

(b) 同样可用正态分布近似，不过这时系求d就等于32的概率。考虑了连续性修正后，成为求$P\ (31.5\leqslant d\leqslant 32.5)=P(1.37\leqslant z\leqslant 1.58)\ =0.0282$，即破损数恰好为32袋的可能性约为3%。

注：用二项分布计算$d=32$的概率为

$$b\ (32,\ 250,\ 0.10)=\frac{250!}{32!218!}\times 0.1^{32}\times 0.9^{218}$$
$$=0.0274$$

它同前面得到的0.0282很接近，但计算则难得多。

±0.5的连续性修正在 (a) 中由于n很大，故效果不明显。不加修正时$P\ (0\leqslant d\leqslant 25)\ =0.50$而不是0.5438，可是，当$n$减小以及当$P$接近0.时，是否进行修正，结果将明显不同。

例 1.4 某商品一年来的每周销售量数据整理后列于下面。现要求确定一个安全的库存量，做到既不脱销又不占压过多资金。为此，要回答 (a)周销售量是否同正态分布符合？ (b) 销售量的均值同中位数① 有何差别？ (c) 连续两周的周销售量超过60件的概率多大？(d)多大的安全库存量可做到95%不脱销？(e)设每件价格为五元，如要做到99%不脱销，占用资金要增

周销售量 件/周	达到该销售量的周数
10～20	2
20～30	1
30～40	6
40～50	8
50～60	17
60～70	12
70～80	5
80～90	0
90～100	1
	52

① 中位数指有序随机变量的中间项。