新编统计技术教程

XINBIAN TONGJI JISHU JIAOCHENG

王连生　编著

中国计量出版社
CHINA METROLOGY PUBLISHING HOUSE

新编统计技术教程

王连生　编著

中国计量出版社

图书在版编目（CIP）数据

新编统计技术教程/王连生编著.—北京：中国计量出版社，2008.10
ISBN 978-7-5026-2873-4

Ⅰ.新… Ⅱ.王… Ⅲ.统计学—教材 Ⅳ.C8

中国版本图书馆 CIP 数据核字（2008）第 110953 号

内 容 提 要

全书共分十一章。包括概率与数理统计基础，统计技术的新、旧工具，正交试验设计与方差分析，回归分析，过程能力与过程能力指数，抽样检验，参数估计与假设检验，FMEA 和 FTA 分析，SPC 与统计技术，六西格玛管理与统计技术等。并附有常用的统计分布表、正交表和抽样检验用表。

本书内容丰富，系统性强，通俗易懂，资料新颖，可作为高等学校相关专业的教学参考书，也可作为企业职工统计技术培训教材。

中国计量出版社出版
北京和平里西街甲 2 号
邮政编码　100013
电话（010）64275360
http：//www.zgjl.com.cn
北京市密东印刷有限公司印刷
新华书店北京发行所发行

*
787 mm×1092 mm　16 开本　印张 14　字数 337 千字
2008 年 9 月第 1 版　2008 年 9 月第 1 次印刷
*
印数 1— 2 000　**定价：28.00 元**

前　言

统计技术是以概率论为理论基础的应用数学的一个分支，是研究随机现象中确定的数字规律科学。它是指应用有关的统计方法，收集、整理、分析和解释统计数据并对其所反映的问题作出一定的结论(如产品质量变异的性质、程度和原因)的科学技术。

统计方法是统计技术的具体方法，如统计设计中七种旧工具：分层法、调查表法、因果图法、排列图法、直方图法、散布图法、控制图法；七种新工具：关联图法、KJ(卡片)法、系统图法、矩阵图法、矩阵数据分析法、PDPC法、箭头图法；高级统计技术：过程能力分析、正交试验设计与方差分析、回归分析、抽样检验、参数估计与假设检验、故障模式与影响分析。

统计技术在质量管理领域里，随着质量科学的发展而得到应用和发展。在20世纪40年代～50年代末统计质量控制(SQC)阶段，控制图得到应用；在我国推行全面质量管理TQM阶段，统计技术的七种旧工具得到一定的应用；ISO 9000标准的出现，依据ISO 9000标准建立的质量管理体系，统计技术作为一个要素写到该标准中。从20世纪80年代发展起来的六西格玛管理和质量源头管理，除了统计技术的七种旧工具应用外，七种新工具和高级统计技术又得到应用。

我国从1992年等同采用ISO 9000标准以来，许多作者围绕着ISO 9000标准中统计技术这一要素的贯彻实施编著了一些有关应用统计技术的书，但由于各个作者思路不同，对统计技术方面的内容选取和编排各不相同，书中统计技术自身体系和系统性被打乱了。因此，当前企业推行六西格玛管理，统计技术作为六西格玛管理有力工具，要大量应用统计技术，可是找不到一本系统地描述统计技术的书，给应用统计技术带来困难。在给企业员工进行统计技术系统知识培训时，还得翻阅许多有关统计技术的书。于是需要重新编写统计培训教程。另外，市面上已有的统计技术书中，对七种旧工具中的控制图判异准则，都还用已失效的准则，国家从2001年已发布了新的标准GB/T 4091—2001《常规控制图》判稳判异准则；抽样检验国家2003年已发布GB/T 2828.1—2003《计数抽样检验程序　第1部分：按接收质量限（AOL）检索的逐批检验抽样计划》新标准，而那些书中仍用已废止的GB/T 2828—1987标准，因此为了适应

在我国推行六西格玛管理——质量管理新时代的到来，基于上述两个方面的原因，急需编著一本系统讲述统计技术的新书，除了统计技术的七种旧工具内容外，还要有七种新工具内容及高级统计技术内容，同时将 GB/T 4091—2001、GB/T 2828.1—2003 两个标准内容编写进去。

本书适用范围为：

(1) 有关综合性大学里经管学院质量专业教学作参考书；

(2) 有关计量学院(或质量专业学校)教学作参考书；

(3) 有关质量培训中心进行统计技术知识系统培训班教学作参考书；

(4) 有关大中企业自行(或外请)对员工(或 QC 小组成员)进行统计技术培训作参考书；

(5) 在全国大中企业推行六西格玛管理和推行“五环法”应用培训，在使用工具培训作参考书等。

因编者水平所限，书中可能存在缺点和错误，恳请读者批评指正。

编　者

2008 年 6 月

目　　录

第 1 章　概率与数理统计基础

1.1　统计的基本概念

1.1.1　统计的起源和应用

（1）统计学的起源和发展

人类为了清点劳动成果及实现原始的分配，就有了粗略的计数活动。“计”和“算”的活动之目的在于“管理”，二者是紧密结合的。17 世纪初，统计学成为一门科学，20 世纪初统计学被应用于管理工作，特别是用于产品质量管理中。在现代社会中，由于计算机的普遍应用，使得统计科学得到充分地发展，在预测、决策、控制等各种管理工作中发挥良好的作用。

（2）统计学的应用

统计学已在质量管理、市场预测、生产管理、气象、水文等科研和管理领域中得到充分地应用。在质量管理中统计学常用于以下方面：

① 市场分析，如可行性分析、市场调查和预测等；

② 产品设计，如可靠性分析、公差分析、故障模型分析等；

③ 过程控制，如控制图、过程能力分析等；

④ 检验和试验，如试验设计、统计抽样等；

⑤ 过程改进，如数据分析、性能评定、允差分析等。

1.1.2　统计的含义和分类

（1）统计的含义

统计是有目的地收集数据、整理数据并使用相应的方法制图、列表与分析数据的过程。

（2）统计分类

① 描述性或推理性统计：尽可能对样本或母体做出分析并得出其特征的结论。

② 归纳性统计：利用有限的样本决定其母体的特征。归纳性统计需应用概率的帮助才能得出较准确的结论。

1.1.3　统计的基本要素

（1）总体

——要调查或统计的某一对象的整体；

——由某一相同性质的许多个别单位组成的集合体(如一个批量的全部产品)；

——可以是客观存在的事物、人物，也可以是抽象的行为。

(2) 样本

——从总体中选出的若干数据的子集；

——受随机性影响的一组数据；

——一组独立（各数据出现互不影响）同分布（样本具有代表性）的随机变量。

(3) 推断

以样本所包含的信息为基础，对总体的某些特征做出决策、预计或统计。

(4) 推断的可靠性

——根据样本信息对总体做出推断；

——预测的优良程度即为推断的可靠性，其误差及可信度是可预知的。

1.1.4 统计的工作过程

(1) 设计：工作规划和安排。

——确定调查目的和对象；

——明确调查项目及制定调查表。

(2) 整理：将搜集的数据分组、加工、汇总。

——统计表；

——统计图；

——统计报告等。

(3) 分析：对加工后的数据进行对比研究，做出推断结论。

1.2 概率与数理统计用数学符号

1.2.1 $\sum$（总和）

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n$$

1.2.2 $\prod$（连乘积）

$$\prod_{i=1}^{n} x_i = x_1 \cdot x_2 \cdot \cdots \cdot x_n$$

1.2.3 $n!$（n 的阶乘）

$$n! = 1 \times 2 \times \cdots \times n$$

1.2.4 C_m^n（组合数）

$$C_m^n = \frac{m!}{n!(m-n)!}$$

1.2.5 $[a_{ij}]_{m\times n}$：$m\times n$ 型矩阵

$$[a_{ij}]_{m\times n}=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

1.2.6 e（自然对数的底）

$$e=2.71828183\cdots$$

1.2.7 π（圆周率）

$$\pi=3.1415926\cdots$$

1.3 统计中数据的整理和分析

1.3.1 统计的对象——数据

(1) 数据的分析

① 计量值

计量值是通过测定得出的数据，它往往具有连续性的特点，如 1.10Ω，1.12Ω，1.14Ω，…。

② 计数值

计数值是通过个数的统计得出的数据，它往往具有间断性的特点，如 3 个、5 个、7 个、……。

(2) 计量值的特点

① 计量值往往由检测设备决定，其准确性决定于控制设备的精度和正确的操作；

② 计量值往往具有连续性的特点；

③ 计量值可用于判定计量对象的“合格”与“不合格”，但这种判定须依据事先确定的标准；

④ 计量值的有效数字的决定往往是重要的，这种有效数字决定的错误，可能会导致误判；

⑤ 计量值在使用统计方法时，可转化为计数值(如某区间数值出现的频数)。

(3) 计数值的特点

① 计数值是通过“点数”获得的数据，其准确性取决于统计的完整性；

② 计数值往往具有间断性的特点；

③ 计数值也可用于判定“合格”与“不合格”，但这种判定取决于直接观察案例数据及事先的假定。

1.3.2 有效数字与数字修约规则

(1) 有效数字

如果数据极限误差大于某一位的半个单位，该位就是有效数字的末位，且该位到左边的非零数字一共有 n 位，n 就是该数据的有效数字的位数；

举例：

某数据极限误差为 0.005，数据 0.234 的末位数应是 3，其有效数字为 2 位；

某数据极限误差为 0.5，数据 8700.33 的末位数应是 0，其有效数字为 4 位。

(2) 数字修约规则

① 对有效数字末位数字后一位数字实行四舍五入的规则：

如数字 0.6331，0.6333，0.6334，若取有效数字为 3 位时，应为 0.633。

② 当有效数字后一位数字为 5 时，视其前一位是单数时，末位数增加 1，如前一位数为双数时，舍去。

如:0.6335—0.634，0.6375—0.638，0.6345—0.634，0.6365—0.636（有效数字为 3 位）。

1.3.3 频数及频数分布表

(1) 频数（f）

将收集到的数据分为若干个组，属于每组数据的个数，称为频数。频数一般用 f 表示。

(2) 频数分布表

① 简单频数分布表

未经分组：直接表述各数据频数的统计表(见表 1－1)。

表 1－1　某宾馆一季度顾客投诉统计表

时间	1 月	2 月	3 月	总次数
频数	13	17	9	39

② 分组表

按规则将数据分组并计算出各组中数据频数的统计表(见表 1－2)。

表 1－2　某机械加工零件长度尺寸的统计表

长度/mm	组值	频数	累积频数分布	相对频数分布
14.00～14.50	14.25	2	2	0.09
14.50～15.00	14.75	5	7	0.32
15.00～15.50	15.25	9	16	0.73
15.50～16.00	15.75	4	20	0.91
16.00～16.50	16.25	2	22	1.00

说明：分组表中累积频数栏和相对频数栏可根据需要设立。

1.3.4 数据的中心趋势分布

(1) 平均数（算术平均数）$\overline{x}$

$$\overline{x}=\frac{\sum_{i=1}^{n}x_i}{n}=\frac{x_1+x_2+\cdots+x_n}{n}$$

式中：$\bar{x}$——算术平均数；

n——数据的个数；

x_i——第 i 个数据。

（2）中位数

当数据的个数为单数时，将数据依大小顺序排列，中间的那个数就是中位数；当数据的个数是偶数时，将数据依大小顺序排列，中间的两个数的平均数就为中位数。

例如：

1，3，4，5，7，8，9 的中位数是 5；

1，3，4，5，7，8，9，10 的中位数为(5+7)/2=6。

（3）众数

出现次数最多的数称为众数，如：

3，4，4，5，6，5，7，8，5，9，其众数为 5；

一些数列中可能无众数，如：

31，32，38，39，40，50

（4）平均数、中位数、众数之间的关系

当分布对称时，三值重合，如图 1—1 所示。

当分布不对称时，三值不重合，如图 1—2 所示。

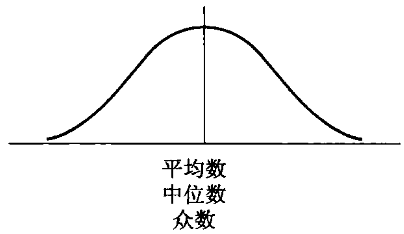


图 1—1　三值重合图

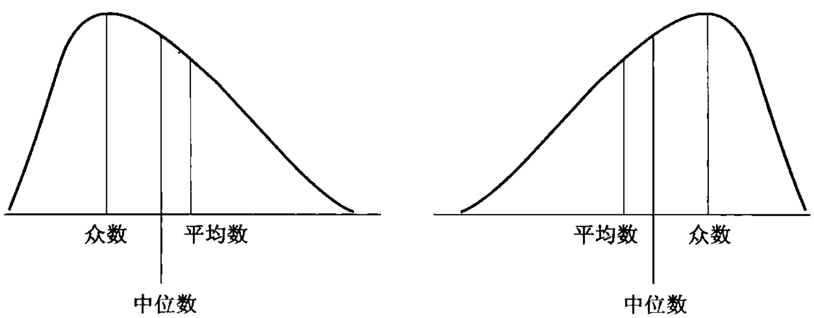


图 1—2　三值不重合图

1.3.5　离势的分析

（1）极差

计算公式为：

$$R=x_{\max}-x_{\min}$$

式中：R——极差；

$x_{\max}$——数列中最大数值；

$x_{\min}$——数列中最小数值。

极差法是一种离势测定的方法，适用于数据数量较少且难以计算时。

（2）标准差

计算公式为：

$$\sigma=\sqrt{\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n}}$$

简化计算方法，则

$$\sigma=\sqrt{\frac{\sum_{i=1}^{n}x_i^2}{n}-\left(\frac{\sum_{i=1}^{n}x_i}{n}\right)^2}$$

式中：σ——标准差；

x_i——第 i 个数据；

$\overline{x}$——数据的平均数；

n——数据的数目。

该公式适用于使用原数据计算。标准差是常见的离势测定值，其测定较准确。当数据列中存在极大或极小值时，应使用标准差作为测定值。

(3) 平均差（绝对平均差）

计算公式为：

$$MD=\frac{1}{n}\sum_{i=1}^{n}|x_i-\overline{x}|$$

式中：MD——平均差；

x_i——第 i 个数据；

$\overline{x}$——数据的平均数；

n——数据的数目。

平均差用绝对值表示；当数据分布较对称时，平均差相当于标准差的 80%左右。

1.4 概率的基础概念

1.4.1 概率基础概念

(1) 随机事件

随机事件是在一定条件下可能发生也可能不发生的事件；概率论研究的问题是随机事件在一定条件下发生可能性的大小。

(2) 概率

A 事件在 n 次试验出现了 m 次，$\frac{m}{n}$就称为 A 事件的概率。用公式表示为：

$$P(A)=\frac{m}{n}$$

$0\leqslant P(A)\leqslant 1$，是一个非负数。

(3) 条件概率与独立事件

① 条件概率

如果事件A、B是两个随机事件，且A为正概率事件(已发生)，则称在A事件发生的条件下B事件发生的概率为条件概率。

② 独立事件

若事件B发生的概率等于在事件A发生条件下事件B发生的条件概率，则称事件A与B为相互独立事件。相互独立的事件，其中一事件的发生与另一事件的发生无影响。

1.4.2 概率运算

(1) 概率加法定理

互不相容事件：不可能同时发生的事件。

完成某事件共有 m 类办法。第1类办法中有 n_1 种办法，第2类办法中有 n_2 种办法，……，第 m 类办法有 n_m 种办法。完成该事件共有 $M=n_1+n_2+\cdots+n_m$ 种方法。

加法公式为：

$$P(\mathrm{A}+\mathrm{B})=P(\mathrm{A})+P(\mathrm{B})-P(\mathrm{AB})$$

当A与B为互不相容事件时：

$$P(\mathrm{A}+\mathrm{B})=P(\mathrm{A})+P(\mathrm{B})$$

(2) 概率乘法定理

① 条件概率

事件A、B是两个随机事件，且A为正概率事件，即 $P(\mathrm{A})\neq 0$，则称在A事件发生的条件下事件B发生的概率为条件概率(也称为在事件A发生条件下的条件概率)。

② 乘法定理

完成某一事件有 m 个步骤(阶段)。完成第一步（阶段）有 n_1 种方法，完成第二步（阶段）有 n_2 种方法，……，完成第 m 步（阶段）有 n_m 种方法，必须完成这 m 个步骤（阶段）才能完成这一事件。完成这一事件共可能有 $M=n_1n_2\cdots n_m$ 种不同的方法。

③ 乘法公式

$$P(\mathrm{AB})=P(\mathrm{B})P(\mathrm{A}|\mathrm{B})$$

当A、B为独立事件时：

$$P(\mathrm{AB})=P(\mathrm{A})P(\mathrm{B})$$

④ 全概率公式

如 A_1，A_2，…，A_n 互不相容，对任一事件B都有：

$$P(\mathrm{B})=\sum_{i=1}^{n}P(\mathrm{A}_i)P(\mathrm{B}\mid \mathrm{A}_i)$$

1.5 常用分布

1.5.1 常用的离散分布

这里将给出三个常用的离散分布，它们是二项分布、泊松分布与超几何分布。

1.5.1.1 二项分布

我们来考察由 n 将随机试验组成的随机现象，它满足如下条件：

（1）重复进行 n 将随机试验。例如，把一枚硬币连抛 n 次，检验 n 个产品的质量，对一个目标连续射击 n 次等。

（2）n 次试验间相互独立，即一次试验结果不对其他次试验结果产生影响。

（3）每次试验仅有两个可能结果，例如，正面与反面、合格与不合格、命中与不命中、具有某特性与不具有该特性，以下统称为“成功”与“失败”。

（4）每次试验成功的概率均为 P，失败的概率均为 $1-P$。

在上述四个条件下，设 X 表示 n 次独立重复试验中成功出现的次数，显然 X 可以取 0，1，…，n 等 $n+1$ 个值的离散随机变量，且它的概率函数为：

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

这个分布称为二项分布，记为 $b(n,\ p)$，其中 $\binom{n}{x}$ 是从 n 个不同元素中取出 x 个的组合数，它的计算公式为：

$$\binom{n}{x}=\frac{n!}{x!\,(n-x)!}$$

二项分布的均值、方差与标准差分别为：

$$E(X)=np$$
$$\mathrm{Var}(X)=np(1-p)$$
$$\sigma(X)=\sqrt{np(1-p)}$$

特例：$n=1$ 的二项分布称为二点分布。它的概率函数为：

$$P(X=x)=p^{x}(1-p)^{1-x},x=0,1$$

或列表如下：

X	0	1
P	$1-p$	p

它的均值、方差与标准差分别为

$$E(x)=p,\mathrm{Var}(x)=p(1-p),\sigma(x)=\sqrt{p-1(-p)}$$

【例 1－1】 在一个制造过程中，不合格品率为 0.1，如今从成品中随机取出 6 个，记 X 为 6 个成品中的不合格品数，则 X 服从二项分布 $b(6,\ 0.1)$，简记为 $X\sim b(6,\ 0.1)$。现研究如下几个问题：

（1）恰有 1 个不合格品的概率是多少？这里规定抽到不合格品为“成功”，则事件 $X=1$ 的概率为：

$$P(X=1)=\binom{6}{1}\times 0.1\times(1-0.1)^{6-1}$$

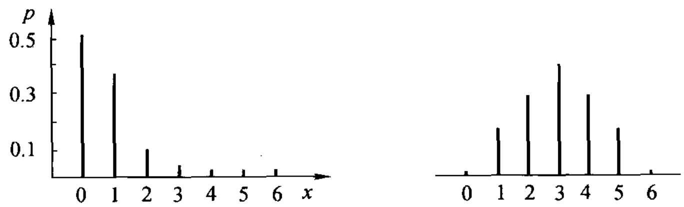
$$=6\times 0.1\times 0.9^5=0.3543$$

这表明，6个成品中恰有一个不合格品的概率为0.3543。类似可计算 $X=0$，$X=1$，…，$X=6$的概率，计算结果可列出一张分布列，具体如下：

X	0	1	2	3	4	5	6
P	0.5314	0.3543	0.0984	0.0146	0.0012	0.0001	0.0000

这里0.0000表示 $X=6$ 的概率取前4位小数的有效数字为零，实际它的概率为 $P(X=6)=0.000001$，并不严格为零。

还可以画出一张线条图［图1－3(a)］来表示这个分布（x 共有7个取值）。图上的横坐标为 X 的取值，纵轴为其相应概率。从此图上可以看出分布的形态，哪些 x 上的概率大，哪些 x 上的概率小。假如改变成功概率 p，其线条图亦会改变。譬如连抛六次硬币，其中正面出现次数 $X\sim b(6,\ 0.5)$。通过计算可画出其线条图［见图1－3(b)］，此图是对称的，如 $P(X=2)=P(X=4)=0.2343$。



(a) 二项分布$b(6,0.1)$的线条图　(b) 二项分布$b(6,0.5)$的线条图

图1－3　二项分布 $b(n,\ p)$ 的线条图

(2) 不超过1个不合格品的概率为：

$$P(X\leqslant 1)=P(X=0)+P(X=1)=0.5314+0.3543=0.8857$$

这表明，6个成品中不超过1个不合格品的概率为0.8857。

在实际中经常要求形如“$X\leqslant x$”的概率，在概率论中把事件“$X\leqslant x$”的概率称为 X 的分布函数，也称为累积分布函数，记为 $F(x)$，即：

$$F(x)=P(X\leqslant x)$$

二项分布的分布函数已编制了数表，详见附表1－1，此表可帮助我们计算二项概率，例如从附表1－1中可查得：

$$P(X\leqslant 1)=0.8857,\ P(X\leqslant 4)=0.9999$$

于是可算得：

$$P(1\leqslant X\leqslant 4)=P(X\leqslant 4)-P(X\leqslant 1)=0.9999-0.8857=0.1142$$

(3) 二项分布 $b(6,\ 0.1)$ 的均值、方差与标准差分别为：

$$E(X)=np=6\times 0.1=0.6$$
$$\mathrm{Var}(X)=np(1-p)=6\times 0.1\times 0.9=0.54$$

$$\sigma(X)=\sqrt{np(1-p)}=\sqrt{0.54}=0.73$$

1.5.1.2 泊松分布

泊松分布可用来描述不少随机变量的概率分布。例如：

(1) 在一定时间内，电话总站接错电话的次数；

(2) 在一定时间内，某操作系统发生的故障数；

(3) 一个铸件上的缺陷数；

(4) 一平方米玻璃上气泡的个数；

(5) 一件产品因擦伤留下的痕迹个数；

(6) 一页书上的错字个数。

从这些例子可以看出，泊松分布总与计点过程相关联，并且计点是在一定时间内、或一定区域内、或一特定单位内的前提下进行的，若 λ 表示某特定单位内的平均点数（$\lambda>0$），又令 X 表示某特定单位内出现的点数，则 X 取 x 值的概率为：

$$P(X=x)=\frac{\lambda^x}{x!}e^{-\lambda},x=0.1,2,\cdots \tag{1-2}$$

这个分布就称为泊松分布，记为 $P(\lambda)$，其中 e 为自然对数的底，即 2.71828…

泊松分布的均值与方差（在数量上）是相等的、均为 λ，即：

$$E(X)=\lambda,\ \mathrm{Var}(X)=\lambda,\ \sigma(X)=\sqrt{\lambda} \tag{1-3}$$

【例 1－2】 某大公司一个月内发生重大事故数 X 是服从泊松分布的随机变量，根据过去事故的记录，该大公司在一个月内平均发生 1.2 起重大事故，这表明：X 服从 $\lambda=1.2$ 的泊松分布，现考察如下事件的概率：

(1) 在一个月内发生 1 起重大事故的概率为：

$$P(X=1)=\frac{1.2}{1!}e^{-1.2}=0.362$$

类似地也可计算 X 取其他值的概率，现罗列于如下分布列中：

X	0	1	2	3	4	5	6	7	…
P	0.301	0.362	0.216	0.087	0.026	0.006	0.002	0.000	…

此例中，X 理论上也可以取 8，9，…等值。由于取这些值的概率的前三位小数皆为零，甚至更小，已无多大实际意义，故可不列出，当作不可能事件处理。也可把此 8 个概率画一张线条图，如图 1－4 所示。

(2) 在一个月内发生重大事故超过 2 起的概率为：

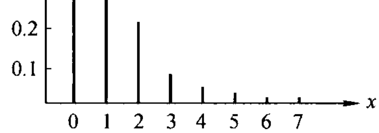


图 1－4 泊松分布 $P(1.2)$ 的线条图

$$\begin{aligned}P(X>2)&=P(X=3)+P(X=4)+P(X=5)+P(X=6)+P(X=7)+\cdots\\&=1-[P(X=0)+P(X=1)+P(X=2)]\\&=1-(0.301+0.362+0.216)=0.121\end{aligned}$$

这表明，该公司在一个月内发生重大事故超过 2 起的