南京航空航天大学
研究生系列精品教材

应用统计学

吴和成　编著

科学出版社

南京航空航天大学研究生系列精品教材

应用统计学

吴和成　编著

科学出版社

北京

内 容 简 介

本书介绍经济与管理学科中常用的统计分析理论与方法。全书分七章。第1章为统计推断的基本内容，主要包括参数估计与检验，方差分析；第2章较为系统地介绍非参数统计检验的基本方法和原理；第3章主要介绍线性回归分析的理论和方法；第4章简要介绍非线性回归分析的基本原理和方法；第5章介绍主成分分析；第6章介绍因子分析模型；第7章介绍马尔可夫链的基本内容。

本书可作为经济学和管理学研究生的应用统计学教材，也可作为从事相关专业教学和研究的教师参考用书。

图书在版编目（CIP）数据

应用统计学/吴和成编著. —北京：科学出版社，2015.8

南京航空航天大学研究生系列精品教材

ISBN 978-7-03-045540-6

Ⅰ.①应… Ⅱ.①吴… Ⅲ.①应用统计学-研究生-教材 Ⅳ.①C8

中国版本图书馆CIP数据核字（2015）第205227号

责任编辑：张 凯/责任校对：王 瑞

责任印制：徐晓晨/封面设计：蓝正设计

科学出版社出版

北京东黄城根北街16号

邮政编码：100717

http://www.sciencep.com

北京教图印刷有限公司印刷

科学出版社发行 各地新华书店经销

*

2015年9月第 一 版 开本：787×1092 1/16

2015年9月第一次印刷 印张：15

字数：356 000

定价：45.00元

（如有印装质量问题，我社负责调换）

前　言

统计是一门关于数据的科学。无论科学研究还是解决实际问题，如果需要从数据中获取信息并进行决策分析，统计方法总是一个有力工具。随着计算技术的迅速发展，运用统计方法解决现实问题变得更为容易和便捷。因而，统计理论和方法在科学研究和实践中发挥着越来越重要的作用。

编者从事研究生的应用统计学课程教学多年，始终认为教材建设是一个基本重要的问题，教学内容的选择一定要以培养目标为前提，兼顾教学对象的专业背景，以及科学的教学方法，方能保证教学质量和效果。

经管类学生必须具备运用统计方法解决实际问题的能力这一理念已成共识。但是，对于掌握分析工具的边界似乎不甚明确。

过于注重理论严密性的教材或纯粹以应用为主的教材都有其不足之处。对于前者，学生面对严密的理论推导和复杂的公式往往一片茫然；对于后者，学生可能仅仅知道方法或公式解决了一个问题，但对于为什么这样解决不得要领。工具的选择及其正确运用是应用者必须严谨对待的问题，这关系着决策的效果甚至成败。因此，正确掌握统计的基本原理和解决问题的思路，在此基础上借助统计分析软件进行科学研究或解决实际问题，是学生需要的基本素养。所以，基于这一理念来精选教学内容具有重要的理论和实际意义。

兼顾不同专业背景和不同专业的培养目标，科学选择既符合学位培养要求的教学内容，又保证整体的教学效果，同时也为不同专业的学生从事相应的研究和应用提供有力支撑。因此，本书从内容上作了精心选择。第 1 章较为系统地介绍统计推断的主要内容，主要有参数估计和参数假设检验，方差分析。其中，前者可以看成知识回顾。事实上，理工背景或经管背景的学生在大学阶段已学习过概率论与数理统计的基本内容。由于概率统计本身的抽象性及来自不同学校学生认识和理解上的差异性，进行系统回顾是必要的，而且，统计推断本身在科学研究和实践中也常用。方差分析可以帮助人们寻找引起某一结果的原因，其基本过程类似于假设检验，只是其解决了三个及以上的正态总体均值的假设检验问题。方差分析在质量管理中是一个重要的分析工具。第 2 章较为系统地介绍非参数统计检验问题。当人们面对不同类型数据或在总体信息十分有限的条件下进行决策时，参数推断方法会有明显的局限性，这体现在决策结果上。事实上，放宽假设更符合实际，但面临着很少的信息困难，非参数统计方法可以在此发挥重要作用。第 3 章介绍线性回归分析。线性回归分析应用广泛，解决问题的原理及方法丰富、深入，对其正确理解十分必要。本书主要介绍一元和多元线性回归模型及其相应的推断、回归诊断、含定性变量的回归模型等。第 4 章简要介绍非线性回归分析方法。其实非线性回归分析的内容是较为深入的，由于问题本身的复杂性，解决问题的原理和方法也显得抽象。因此，从应用的要求，本书仅介绍基本的非线性回归分析方法。第 5 章介绍主

成分分析。第 6 章介绍因子分析。第 5，6 章主要介绍如何从数据中综合或提炼出可以帮助人们解决问题的信息。另外，也可以作为诸如回归分析变量选择的基础。第 7 章简要介绍随机过程的基本概念和常用的马尔可夫链，为有需要的学生提供一定的基础。

本书的理论介绍主要阐述其基本思想和解决问题的基本思路，以够用为度，但不失严谨性。例题的选择注重多样性和实用性，以便学生能够通过例题的求解来培养运用统计方法解决实际问题的能力。事实上，本书介绍的许多方法可以直接用于解决实际问题。本书为南京航空航天大学研究生教育优秀工程研究生教材建设项目。

本书引用国内外文献中的一些例题和习题，有些是根据文献中例题或习题重新设计，恕不一一指明出处，在此一并向相关人员表示感谢。

教材建设任重而道远。本书在这方面作了一些探索，限于编者学识，不妥之处在所难免，欢迎专家学者和读者批评指正。

编　者

2015 年 7 月

目　录

第1章　统计推断

房价问题是当前最热门的话题之一。一个城市房价的均价总是扑朔迷离。一个房价均价每平方米8千元的经济较为发达的省会城市，可能对于年轻人具有较大的吸引力。现实却是想要购买每平方米1万元房子的愿望，也可能只有在城郊结合部才能实现。事实上，需要弄清楚的是这个城市房子均价的变化区间、不同楼盘均价之间的差异程度、在某一价位以上的楼盘占比多少、不同区位楼盘均价之间的差异及其差异的变化趋势等。当不能获得全部楼盘销售均价的数据时(实际上难以得到真实的数据)，你如何来解决刚才提到的问题呢？

1.1　随机变量及其分布

随机试验的结果未必都是数量化的，如检验产品是合格品还是不合格品，调查居民对某一改革措施赞成还是反对等，这些实验的结果并不是一个数值。为了全面研究随机实验的结果，揭示随机现象的统计规律性，需要将随机实验的结果数量化，即需要引入随机变量概念。

为理解随机变量的涵义，从一个统计学文献中常用的一个例子，即抛掷硬币以观察正反面出现情况的这一试验开始。例如，将硬币连续抛掷三次(看成一次随机试验)，则所有可能结果的集合为 $\Omega=\{\mathrm{HHH},\mathrm{HHT},\mathrm{HTH},\mathrm{THH},\mathrm{HTT},\mathrm{THT},\mathrm{TTH},\mathrm{TTT}\}$，这里，用H表示正面，T表示反面。显然，当硬币均匀时，这8个结果的出现等可能。将试验所有可能结果组成的集合 Ω 称为样本空间。如果仅将注意力集中在正面出现的次数上，如以 X 表示这一试验中正面出现的次数，则 X 可能的取值为0,1,2,3。且易知，X 取这4个数的概率分别为1/8,3/8,3/8和1/8。事实上，这些概率值对应着试验结果出现的概率。例如，$X=1$ 对应着试验结果HTT,THT或TTH的出现，则 $X=1$ 的概率等于试验结果HTT,THT或TTH出现的概率之和。因此，X 是定义在样本空间上的一个实值函数。

随机变量的严格定义如下：设 E 是一个随机试验，$S=\{e\}$ 为其样本空间，如果对于 S 中的每一个样本点 e，有一个实数 $X(e)$ 与之对应，则称这个定义在样本空间 S 上的实值函数 $X(e)$ 为随机变量。

随机变量 X 的分布函数定义如下：对于任意的实数 x，$-\infty<x<\infty$，称函数

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

为随机变量 X 的累积分布函数(简称分布函数)。实际上，$F(x)$ 是随机变量 X 取值不超过某一特定值的概率，故有累积之意。

容易看到，分布函数具有如下性质：

(1) $F(x)$ 是 x 的非减函数；

(2) $\lim\limits_{x\to+\infty}F(x)=1$;

(3) $\lim\limits_{x\to-\infty}F(x)=0$;

(4) $P\{a<x\leqslant b\}=F(b)-F(a)$,对一切 $a<b$。

1.1.1 常用的随机变量及其分布

1. 离散型随机变量及其分布

一个最多取可数个可能值的随机变量,称为离散型随机变量。对于一个离散型随机变量 X,记 $p_i=P\{X=x_i\}$,这里 x_i 为 X 的可能取值,则 $p_i>0$,且对于所有的 x_i,有 $\sum\limits_{i=1}^{+\infty}p_i=1$;$X$ 的分布函数 $F(x)=\sum\limits_{x_i\leqslant x}P\{X=x_i\}$。

下面介绍一些常用的离散型随机变量。

1) 0-1 分布

假定一个随机试验,其结果可以分为成功或失败,称这样的试验为伯努利试验。例如,试验的结果是成功,令 $X=1$,否则,令 $X=0$,则 X 的分布律为

$$P\{X=k\}=p^k(1-p)^{1-k},\quad k=0,1$$

这里,p 为试验结果是成功的概率,且 $0<p<1$。

随机变量 X 也称为伯努利随机变量,如果其分布律由上述公式给出,称 X 服从 0-1 分布,记为 $X\sim b(1,p)$。

在实践中,对产品进行质量检验,每抽出一件产品,只有两种结果,即要么是合格品,要么是不合格品,如记产品的合格率为 p,则产品的质量检验问题可以用 0-1 分布来描述。

2) 二项分布

若进行 n 次独立的伯努利试验,其中每次结果是成功的概率为 p,结果是失败的概率为 $1-p$。以 X 表示在 n 次独立的伯努利试验中成功出现的次数,则称 X 为具有参数 (n,p) 的二项随机变量,或称 X 服从参数 (n,p) 的二项分布,记为 $X\sim b(n,p)$。其分布律为

$$P\{X=k\}=C_n^kp^k(1-p)^{1-k},\quad k=0,1,2,\cdots,n$$

例 1.1 已知某生产线生产的产品是废品的概率为 0.1,且与任意的其他产品独立。现从生产线上随机抽取 3 件产品,则至多有一个废品的概率是多少?

解 以 X 表示这 3 件被抽产品中的废品数,则 X 为服从参数 $(3,0.1)$ 的二项随机变量。

$$P\{X\leqslant 1\}=P\{X=0\}+P\{X=1\}=C_3^0(0.1)^0(0.9)^3+C_3^1(0.1)^1(0.9)^2=0.972$$

例 1.2 某公司有 7 个顾问。假定每个顾问贡献正确意见的概率为 0.6,且设顾问之间是否贡献正确意见相互独立。先对某项目可行与否个别征求各顾问意见,并按多数顾问的意见作出决策。试求作出正确决策的概率。

解 以 X 表示 7 个顾问中贡献正确意见的人数,则 $X\sim b(7,0.6)$。从而作出正确决策的概率为

$$P\{X \geqslant 4\} = \sum_{k=4}^{7} P\{X = k\} = \sum_{k=4}^{7} C_7^k (0.6)^k (0.4)^{7-k} = 0.7102$$

例 1.3 某车间有 80 台机器，经过长时间的观察，得知每台机器发生故障的概率为 0.01。设机器发生故障与否相互独立，又设每个维修工在同一时间只能维修一台机器，则配备 3 个维修工共同维修 80 台机器，与配备 4 个维修工每人承担 20 台机器维修任务，哪个方案不能及时维修的概率较小？

解 (1) 按照第 1 种方案，以 X 表示 80 台机器中需要维修的机器数，可易见，$X \sim b(80, 0.01)$，则不能及时维修的概率为

$$P\{X \geqslant 4\} = \sum_{k=4}^{80} C_{80}^k (0.01)^k (0.99)^{80-k} = 0.0091$$

(2) 按照第 2 种方案，以 A_i 表示事件“第 $i(i=1,2,3,4)$ 个维修工承担的 20 台机器不能及时维修”，则所求的概率为

$$P(A_1 \cup A_2 \cup A_3 \cup A_4) \geqslant P(A_1) = \sum_{k=2}^{20} C_{20}^k (0.01)^k (0.99)^{20-k} = 0.0175$$

由此可见，第 1 种方案较好。

注 二项分布的概率计算可以调用 excel 中的函数 BINOMDIST。

3) 泊松分布

对于取值为 0,1,2,… 的随机变量 X，如对某个 $\lambda > 0$，有

$$P\{X = k\} = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0,1,2,\cdots$$

则称 X 为具有参数 λ 的泊松随机变量，或称 X 服从参数为 λ 的泊松分布，记为

$$X \sim \pi(\lambda)。$$

泊松分布的一个重要性质是可以用来近似二项分布。事实上，如果二项分布参数中的 n 较大，而 p 较小，对于二项分布的随机变量，取 $\lambda = np$，则

$$P\{X = k\} = C_n^k p^k (1-p)^{1-k} = \frac{n(n-1)\cdots(n-k-1)}{n^k} \cdot \frac{\lambda^k}{k!} \cdot \frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^k}$$

对于较大的 n 和较小的 p，有

$$\left(1-\frac{\lambda}{n}\right)^n \approx e^{-\lambda}, \quad \frac{n(n-1)\cdots(n-k-1)}{n^k} \approx 1, \quad \left(1-\frac{\lambda}{n}\right)^k \approx 1$$

从而，对于较大的 n 和较小的 p，有

$$P\{X = k\} = C_n^k p^k (1-p)^{1-k} \approx \frac{\lambda^k e^{-\lambda}}{k!}$$

例 1.4 假定某书一页上的印刷错误个数是一个具有参数 $\lambda = 1$ 的泊松随机变量，则在此页上至少有一个错误的概率为多少？

解 以 X 表示此页上的错误数，则 $X \sim \pi(1)$，从而

$$P\{X \geqslant 1\} = 1 - P\{X = 0\} = 1 - e^{-1} = 0.633$$

例 1.5 假定每天在高速公路上发生事故的数目是一个具有参数 $\lambda=3$ 的泊松随机变量，则今天没有发生事故的概率是多少？

解 以 X 表示今天在此条高速公路上发生的事故数，则

$$P\{X=0\}=e^{-3}=0.05$$

例 1.6(泊松分布在运营管理中的应用：排队) 在生活和工作中排队是常见现象，如在银行、超市、餐饮店等场所都会遇到排队的情况；再如，货车等待装货、生产线上的零件排队等待装配等。通过排队模型，可以帮助公司管理人员掌握排队的特征。

每小时到达某加油站要求加油的汽车数服从均值为 5 的泊松分布，则

(1) 接下来的 1 个小时内只有一辆车到达的概率是多少？

(2) 接下来的 3 个小时内有多于 20 辆汽车到达的概率是多少？

某 ATM 机使用人数服从泊松分布，每间隔 5 分钟平均有 1.5 个使用者，则

(1) 在接下来的 5 分钟内没有使用者的概率是多少？

(2) 接下来的 10 分钟内有 3 个或 3 个以上使用者的概率是多少？

作者可自行练习。

注 也可以调用 excel 中的函数 POISSON 进行计算。

4) 几何分布

设进行独立试验直到首次出现成功为止，其中每次试验成功的概率都是 p，以 X 表示直到首次成功所进行的试验次数，则称 X 为具有参数 p 的几何随机变量，或称 X 服从参数为 p 的几何分布，记为 $X\sim g(n,p)$。其分布律为

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

例 1.7 对产品进行检验，直到检测到次品为止。设产品的合格率为 0.9，求直到第 11 个产品才检测到次品的概率。

解 以 X 表示首次检测到次品时所检测的产品数，则 $X\sim g(11,0.9)$，由此

$$P\{X=11\}=(1-0.1)^{10}0.1=9\times10^{-11}$$

2. 连续型随机变量及其分布

在某型号灯泡的寿命试验中，每一个被测试灯泡的寿命是一个非负实数，它可以取到某个区间中的任意一个数。同样该型号灯泡的寿命在某一范围内取值的概率也是客观存在的。将这样能取到一个区间中任意一个数的随机变量，称为连续型随机变量。

连续型随机变量的分布函数为

$$F(x)=P\{X\leqslant x\}=\int_{-\infty}^{x}f(x)\mathrm{d}x$$

这里，$f(x)$是连续型随机变量 X 的分布密度函数。

1) 均匀分布(记为 $X\sim U(a,b)$)

密度函数为

$$f(x)=\begin{cases}\dfrac{b-a}{2}, & a<x<b\\ 0, & \text{其他}\end{cases}$$

2) 指数分布(记为 $X\sim E(\lambda)$)

密度函数为

$$f(x)=\begin{cases}\lambda e^{-\lambda x}, & x>0\\ 0, & \text{其他}\end{cases}$$

例 1.8 已知某种轮胎的使用寿命 $X\sim E(0.1)$(单位:万公里)。现随机抽取这种轮胎5只,试求至少有两只轮胎的行驶距离不足30万公里的概率(1公里=1千米)。

解 以 X 表示任意一只这样的轮胎的使用寿命,则其寿命不足30万公里的概率为

$$P\{X<30\}=\int_{-\infty}^{30}0.1e^{-0.1x}dx=0.9502$$

于是5只轮胎中至少有两只轮胎的行驶距离不足30万公里的概率为

$$P=1-\sum_{k=0}^{2}C_5^k(0.9502)^k(0.0498)^{5-k}=0.99997$$

3) 正态分布(记为 $X\sim N(\mu,\sigma^2)$)

密度函数为

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},\quad -\infty<x<+\infty$$

当 $\mu=0,\sigma=1$ 时,$f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$,$-\infty<x<+\infty$称为标准正态分布的密度函数,对应的随机变量以 Z 表示,且记 $Z\sim N(0,1)$。$X\sim N(\mu,\sigma^2)$与 $Z\sim N(0,1)$之间的关系为 $Z=\frac{X-\mu}{\sigma}$,且

$$P\{a<X\leqslant b\}=\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)$$

这里,$\Phi(x)$为标准正态分布 Z 的分布函数。

由正态分布的密度函数图像(图1.1)可以看到,此曲线完全由均值 μ 和标准差 σ 决定,事实上,μ 决定了密度函数曲线的位置,也称位置参数;σ 决定了曲线的形状,也称尺度参数。

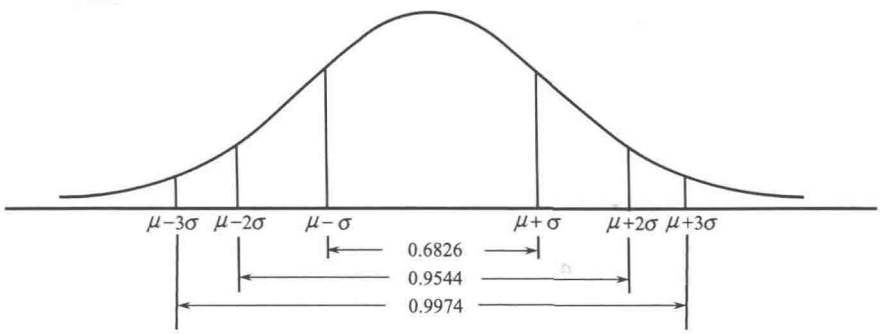

图 1.1 正态分布

正态随机变量的 3 个重要数据：若 $X\sim N(\mu,\sigma^2)$，则

$$P\{\mu-\sigma\leqslant X\leqslant\mu+\sigma\}=0.6826$$

$$P\{\mu-2\sigma\leqslant X\leqslant\mu+2\sigma\}=0.9544$$

$$P\{\mu-3\sigma\leqslant X\leqslant\mu+3\sigma\}=0.9974$$

我们可以看到，X 的取值几乎落在以均值为中心，3 倍的标准差为半径的对称区间中。此性质也称为 3σ 准则，其在产品的质量控制中有着重要应用。

例 1.9（招生录取线的确定） 某学校近年招生情况看好，申请者越来越多，因此，录取标准需要提高。经学校管理部门反复论证，制订出一个录取条件，即申请者的入学分数必须在前 1%以内。如果入学分数服从均值为 490，标准差为 61 的正态分布，则录取的最低分为多少？

解 以 X 表示申请者的入学分数，则 $X\sim N(490,61^2)$。记最低录取分数线为 $x_{0.01}$，则有

$$P\{X>x_{0.01}\}=0.01$$

即

$$P\left\{\frac{X-490}{61}>\frac{x_{0.01}-490}{61}\right\}=0.01,\quad \text{或 } P\left\{Z>\frac{x_{0.01}-490}{61}\right\}=0.01$$

这里，$Z\sim N(0,1)$，查附表 1 得 $\dfrac{x_{0.01}-490}{61}=2.3263$，即 $x_{0.01}=632$。

实际上，在上述常用分布的概率计算中，都可以运用 excel 统计计算中的相应函数，请读者思考。本例中，可以运用 excel 中的函数 NORMINV，立得 $x_{0.01}=632$。

1.1.2 随机变量的矩

若 $E(X^k)$存在，则称之为随机变量 X 的 k 阶原点矩，$k=1,2,\cdots$；

若 $E(X-E(X))^k$ 存在，则称之为随机变量的 k 阶中心矩，$k=2,3,\cdots$。

特别地，称随机变量 X 的一阶原点矩 $E(X)$为随机变量 X 的数学期望，也称为均值；称随机变量 X 的二阶中心矩 $E(X-E(X))^2$ 为随机变量 X 的方差，称$\sqrt{E(X-E(X))^2}$为随机变量 X 的标准差。

在实践中最常用的当属随机变量的数学期望与方差。

下面给出常用随机变量的数学期望与方差。

1. 离散情形

$$E(X)=\sum_{x_k}x_kp_k$$

$$D(X)=E[(X-E(X))^2]=E(X^2)-E^2(X)=\sum_{x_k}x_k^2p_k-\left(\sum_{x_k}x_kp_k\right)^2$$

这里，x_k 为离散型随机变量 X 的可能取值，且 $p_k=P\{X=x_k\}$。

(1) 若 $X\sim b(1,p)$，则

$$E(X)=p,\quad D(X)=p(1-p)$$

(2) 若 $X\sim b(n,p)$，则

$$E(X)=np,\quad D(X)=np(1-p)$$

(3) 若 $X\sim g(n,p)$,则

$$E(X)=\frac{1}{p},\quad D(X)=\frac{1-p}{p^2}$$

(4) 若 $X\sim\pi(\lambda)$,则

$$E(X)=\lambda,\quad D(X)=\lambda$$

例 1.10(项目管理) 设某项工程的合同规定:若 3 天完成,工程队可得 10000 元;若 4 天完成,则得 2500 元;若 5 天完成,工程队要赔偿 7000 元。以 X 表示工程队承接这项工程的收益,易见 X 是一个随机变量。根据以往资料知 X 的分布律如表 1.1 所示。

表 1.1 分布律

X	$x_1=10000$	$x_2=2500$	$x_3=-7000$
p_k	$\frac{1}{8}$	$\frac{5}{8}$	$\frac{2}{8}$

试计算工程队的平均收益。

解 $E(X)=10000\times\frac{1}{8}+2500\times\frac{5}{8}-7000\times\frac{2}{8}=1062.50$。

2. 连续情形

设 X 为连续型随机变量,若积分 $\int_{-\infty}^{+\infty}|xf(x)|\,\mathrm{d}x$ 收敛,则 X 的数学期望定义为

$$E(X)=\int_{-\infty}^{+\infty}xf(x)\mathrm{d}x$$

(1) 若 $X\sim U(a,b)$,则

$$E(X)=\frac{a+b}{2},\quad D(X)=\frac{(b-a)^2}{12}$$

(2)若 $X\sim E(\lambda)$,则

$$E(X)=\frac{1}{\lambda},\quad D(X)=\frac{1}{\lambda^2}$$

(3) 若 $X\sim N(\mu,\sigma^2)$,则

$$E(X)=\mu,\quad D(X)=\sigma^2$$

例 1.11(项目管理) 某工程项目有四个关键性环节,完成每个环节所需要的时间的期望和方差如表 1.2 所示。

表 1.2 四个环节相关资料

环节	所需时间的期望/天	方差
1	18	8
2	12	5
3	27	6
4	8	2

在项目管理中，可以假设完工时间近似服从正态分布。那么，此工程项目的四个关键性环节完成时间在 60 天以上的概率为多少？

解 这里涉及 4 个正态分布，即 4 个环节完工时间 X_1, X_2, X_3, X_4 分别服从正态分布 $N(18,8), N(12,5), N(27,6)$ 和 $N(8,2)$。记 X 为完成这四个关键环节所需的时间，则 $X \sim N(65,21)$。由此，

$$P\{X>60\}=1-P\{X\leqslant 60\}=1-P\left\{\frac{X-65}{\sqrt{21}}\leqslant\frac{60-65}{\sqrt{21}}\right\}$$

$$=1-\Phi\left(-\frac{5}{\sqrt{21}}\right)=0.8621$$

1.1.3 分位点

设随机变量 X 的分布函数为 $F(x)$，称满足下列方程

$$F(f_\alpha)=1-\alpha, \quad \text{或 } 1-F(f_\alpha)=\alpha$$

的实数 f_α 为随机变量 X 的分布的上 α 分位点。

例如，对于标准正态分布，将其上 α 分位点记为 z_α，即 z_α 满足方程

$$\int_{z_\alpha}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=\alpha$$

在应用中，常用的有 $z_{0.025}=1.96$，$z_{0.05}=1.645$。这些可以在标准正态分布表中查到。

将常用随机变量的分布及其数字特征归纳为表 1.3 和表 1.4，供使用时查阅。

表 1.3 离散型随机变量的分布律及其数字特征

随机变量的名称	分布律	均值	方差
二项分布，参数为 $n,p,0\leqslant p\leqslant 1$	$P\{X=k\}=C_n^k p^k(1-p)^{n-k}$ $k=0,1,2,\cdots,n$	np	$np(1-p)$
泊松分布，参数为 $\lambda,\lambda>0$	$P\{X=k\}=\frac{\lambda^k}{k!}e^{-\lambda}$ $k=0,1,2,\cdots$	λ	λ
几何分布，参数为 $p,0\leqslant p\leqslant 1$	$P\{X=n\}=p(1-p)^{n-1}$ $n=1,2,\cdots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$

表 1.4 连续型随机变量的分布律及其数字特征

随机变量的名称	分布密度函数	均值	方差
(a,b) 上的均匀分布	$f(x)=\begin{cases}\frac{1}{b-a}, & a<x<b\\ 0, & \text{其他}\end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$