



创优系列·工程硕士



本书含  
二维码



本书含  
精美PPT



# 管理统计学

STATISTICS FOR MANAGEMENT

吴和成 编 著



中国工信出版集团



电子工业出版社  
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY  
<http://www.phei.com.cn>



创优系列·工程硕士

# 管理统计学

STATISTICS FOR MANAGEMENT

吴和成 编 著

電子工業出版社·

Publishing House of Electronics Industry

北京·BEIJING

## 内 容 简 介

本书介绍了经济与管理学科中常用的统计分析理论与方法。全书分7章,第1章对统计推断进行了总结回顾;第2章详细介绍了方差分析的基本原理和方法;第3章较为系统地介绍了非参数统计方法的基本内容;第4章主要包含线性回归的主要理论和方法;第5章则简要介绍了非线性回归分析的基本原理和方法;第6章讨论了常用的统计决策方法;第7章介绍了常用的统计控制图。

本书可作为工程硕士和 MBA 培养的统计学教材,也可作为从事相关专业教学和研究的教师参考用书。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。  
版权所有,侵权必究。

## 图书在版编目(CIP)数据

管理统计学/吴和成编著. —北京:电子工业出版社,2015.4  
ISBN 978-7-121-25793-3

I. ①管… II. ①吴… III. ①经济统计学-高等学校-教材 VI. ①F222

中国版本图书馆 CIP 数据核字(2015)第 065947 号

策划编辑:王赫男

责任编辑:郝黎明

印 刷:北京中新伟业印刷有限公司

装 订:北京中新伟业印刷有限公司

出版发行:电子工业出版社

北京市海淀区万寿路 173 信箱 邮编:100036

开 本:787×1092 1/16 印张:15.5 字数:394 千字 插页:1

版 次:2015 年 4 月第 1 版

印 次:2015 年 4 月第 1 次印刷

定 价:37.00 元

凡所购买电子工业出版社图书有缺损问题,请向购买书店调换。若书店售缺,请与本社发行部联系,联系及邮购电话:(010)88254888。

质量投诉请发邮件至 zlt@phei.com.cn,盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线:(010)88258888。

# 前言

统计学是一门关于数据的科学。无论是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。特别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,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。

本书作者从事工程硕士和 MBA 统计学的教学和科研工作多年,认为教材建设始终是一个基本重要的问题。对于攻读专业学位的学生来说,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。对于前者,学员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,学员可能仅仅知道用这一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,是对学员的基本要求。因此,基于这一理念,来组织教学内容具有重要的实际意义。

专业学位教学对象的专业背景各异,要科学选择既符合专业学位培养要求的教学内容,又要保证整体的教学效果;既要介绍必需的统计知识,又要为不同专业的学生从事相应的研究提供有力支撑。因此,本书从内容上作了精心选择,第 1 章较为系统地回顾了统计推断的主要内容,也可看作预备知识。事实上,对于理工背景或经管背景的学员在大学阶段应接受过概率统计基本内容的学习。由于概率统计本身的抽象和学员工作年限较长,对于概率统计的基本知识也无多存量,更少有对这些知识的深入理解。另外,统计推断本身在科学研究和实践中也很常用,因而,系统回顾是必需的。第 2 章介绍了方差分析的基本内容。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其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 3 章较为系统地介绍了非参数统计检验问题。当人们面对不同类型的数据或在总体的信息十分有限的条件下进行决策时,参数推断方法就有明显的局限性,这体现在决策结果的效果上。事实上,放宽假设更为符合实际,但面临着很少信息的困难,非参数统计方法可以在此发挥其重要作用。第 4 章为线性回归分析。线性回归分析的广泛应用乃众所周知,其解决问题的原理和方法丰富且深入,对其正确理解十分必要。这一章主要介绍了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 5 章则简要介绍了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,这一章仅介绍了基本的非线性回归分析方法。第6章为统计决策。主要介绍了常用的统计决策方法。第7章为统计控制图。统计控制在生产实践中具有重要作用,这一章介绍了常用的统计控制图。

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。例题的选择注重多样性和实用性,以便学员能够通过例题的求解来培养其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。

本书引用了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。

专业学位教育的统计学教材建设任重而道远。本书在这方面做了一些探索,尚有许多不成熟之处,诚恳欢迎专家学者和读者批评指正。

编 者

# 目 录

第1章 统计推断回顾	1	2.3.2 双因素无重复试验的方差分析	71
1.1 随机变量及其分布	3	2.4 应用举例	73
1.1.1 常用的随机变量及其分布	3	2.5 本章小结	75
1.1.2 随机变量的矩	8	问题与思考	76
1.1.3 分位点	10	第3章 非参数统计分析	77
1.2 抽样分布及其常用统计量的分布	11	3.1 符号检验	80
1.2.1 简单随机样本	11	3.1.1 两个总体分布是否相同的符号检验	80
1.2.2 抽样分布	12	3.1.2 总体中位数 $M_0$ 的检验	83
1.3 统计推断	19	3.1.3 数据序列的趋势存在性检验	85
1.3.1 参数估计	20	3.1.4 Wilcoxon 符号秩和检验	88
1.3.2 参数假设检验	31	3.2 秩和检验法	89
1.4 假设检验中的两个问题	47	3.3 多个样本的检验	92
1.4.1 置信区间与假设检验的关系	47	3.3.1 克鲁斯卡尔—沃利斯(Kruskal-Wallis)单向方差秩检验	92
1.4.2 假设检验中的 $P$ 值	48	3.3.2 费利德曼双向方差分析	95
1.5 本章小结	50	3.4 秩相关分析	98
问题与思考	50	3.4.1 Spearman 秩相关系数	98
第2章 方差分析	51	3.4.2 Kendall- $\tau$ 相关系数	100
2.1 单因素试验的方差分析	54	3.5 $\chi^2$ 检验法	102
2.1.1 问题的提出	54	3.5.1 拟合优度检验	102
2.1.2 解决问题的基本思路及过程	56	3.5.2 独立性检验(列联表分析)	106
2.2 多重比较	62	3.6 正态性的检验法	108
2.2.1 费歇尔(Fisher)的最小显著差法(LSD)	64	3.7 本章小结	109
2.2.2 基于庞弗洛尼(Bonferroni)调整的LSD法	64	问题与思考	109
2.2.3 图基(Tukey)多重比较法	65	第4章 线性回归分析	111
2.3 双因素试验的方差分析	66	4.1 一元线性回归模型	114
2.3.1 双因素等重复试验的方差分析	66	4.1.1 参数 $\beta_0, \beta_1$ 的估计	115
		4.1.2 误差项 $\varepsilon$ 方差 $\sigma^2$ 的估计	116

4.1.3	拟合回归线的性质/117	5.3.1	回归模型/166
4.1.4	正态误差回归模型/117	5.3.2	关于误差项问题/167
4.1.5	线性回归模型中自变量与因变量之间联系的描述测度/121	5.3.3	参数估计/167
4.1.6	一元线性回归建模流程/122	5.4	logistic 回归模型/170
4.2	多元线性回归模型/122	5.5	本章小结/173
4.2.1	多元线性回归模型的概念/122		问题与思考/173
4.2.2	回归系数的涵义/124	第6章	统计决策…………… 175
4.2.3	回归分析推断/124	6.1	统计决策概念/177
4.2.4	预测与控制/128	6.1.1	统计决策的定义/177
4.2.5	自变量与因变量线性相关程度的度量指标/128	6.1.2	统计决策的三个基本概念/177
4.2.6	多元线性回归模型中自变量的选择问题/132	6.1.3	决策的一般步骤/177
4.3	回归诊断/138	6.2	不确定型决策方法/178
4.3.1	残差及其性质/138	6.3	风险型决策方法/182
4.3.2	误差项的异方差/139	6.3.1	决策矩阵法/183
4.3.3	误差序列自相关性/141	6.3.2	决策树法/184
4.3.4	自变量的多重共线性/142	6.4	贝叶斯决策方法/187
4.3.5	异常点与强影响点/145	6.4.1	贝叶斯公式/187
4.4	含定性自变量的回归模型/147	6.4.2	贝叶斯决策方法/189
4.4.1	仅含定性自变量的回归模型/147	6.5	本章小结/193
4.4.2	含一个定量自变量和一个二值定性自变量的回归/148		问题与思考/194
4.4.3	含一个定量自变量和一个多值定性自变量的回归/152	第7章	统计过程控制图…………… 195
4.4.4	对于一个定量自变量和两个定性自变量的回归/153	7.1	过程波动/198
4.5	本章小结/154	7.2	控制图原理/198
	问题与思考/154	7.2.1	控制图的形式/198
第5章	非线性回归分析…………… 155	7.2.2	控制图的判断准则/200
5.1	可线性化的非线性回归模型/158	7.3	控制图/203
5.2	多项式回归模型/162	7.3.1	均值-极差 $\bar{X}-R$ 控制图/203
5.2.1	一元多项式回归模型/162	7.3.2	均值-标准差 $\bar{X}-S$ 控制图/207
5.2.2	二元多项式回归模型/164	7.3.3	单值 $\bar{X}-R_s$ 控制图/210
5.3	因变量为指示变量的回归模型/166	7.3.4	计数值控制图/212
		7.4	本章小结/219
			问题与思考/219
		附表	…………… 220
		参考文献	…………… 243

# 第 1 章 统计推断回顾

## 本章提要

主要回顾随机变量的概念、常用随机变量的分布；简单随机样本概念及常用统计量的分布；总体分布参数中均值、方差和比例的点估计；单个总体分布的均值、方差、比例的区间估计；两个总体分布的均值之差、方差之比和比例之差的区间估计；单个总体分布均值、方差、比例的假设检验；两个总体分布均值、方差、比例的假设检验；总体参数的区间估计与假设检验之间的关系；假设检验中的 P 值概念。





## 导入案例

房价问题似乎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。一个房子均价 8000 元的经济较为发达的省会城市，可能对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，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。当你不能获得全部楼盘销售均价的数据（实际上难以得到真实的数据）时，你如何解决刚才提到的这些问题呢？

## 1.1 随机变量及其分布

众

所周知,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,这些试验的结果并不是一个数值。为了全面研究随机试验的结果,揭示随机现象的统计规律性,需要将随机试验的结果数量化,即需要引入随机变量概念。

为理解随机变量的含义,我们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。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  $\Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$ , 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  $\Omega$  称为样本空间。如果我们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。如  $X = 1$  对应着试验结果 HTT, 或 THT 或 TTH 的出现,则  $X = 1$  的概率等于试验结果 HTT 或 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,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  $X(e)$  (简记为  $X$ ) 为随机变量。

定义随机变量  $X$  的分布函数如下:

对于任意的实数  $x, -\infty < x < \infty$ , 称函数

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

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)  $\lim_{x \rightarrow \infty} F(x) = 1$ ;
- (3)  $\lim_{x \rightarrow -\infty} F(x) = 0$ ;
- (4)  $P\{a < X \leq b\} = F(b) - F(a)$ , 对一切  $a < b$ 。

### 1.1.1 常用的随机变量及其分布

#### 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  $\sum_{i=1}^{\infty} p_i = 1$ ;  $X$  的分布函数  $F(x) = \sum_{x_i \leq x} P\{X = x_i\}$ 。

下面介绍一些常用的离散型随机变量。

### ◎ 0-1 分布(记为 $X \sim p(1, p)$ )

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

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

这里,  $p$  为试验结果为成功的概率,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。

在实践中,对产品进行质量检验,那么,每抽出一件产品,只有两种结果,即要么是合格品,要么为不合格品,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。

### ◎ 二项分布(记为 $X \sim b(n, p)$ )

如进行  $n$  次独立的伯努利试验,其中每次结果为成功的概率为  $p$ , 结果为失败的概率为  $1 - p$ 。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。其分布律:

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

【例 1.1】 已知某生产线生产的产品为废品的概率为 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, 则至多有一件废品的概率是多少?

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

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

【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。公司对某项目可行与否个别征求各顾问意见, 然后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

$$\begin{aligned} P\{X \geq 4\} &= \sum_{k=4}^7 P\{X = k\} \\ &= \sum_{k=4}^7 C_7^k (0.6)^k (0.4)^{7-k} \\ &= 0.7102 \end{aligned}$$

【例 1.3】 某车间有 89 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) 按照第一种方案, 以  $X$  表示 80 台机器中需要维修的机器数, 易见,  $X \sim b(80, 0.01)$ 。则不能及时维修的概率为

$$\begin{aligned}
 P\{X \geq 4\} &= \sum_{k=4}^{80} C_{80}^k (0.01)^k (0.99)^{80-k} \\
 &= 0.0091
 \end{aligned}$$

(2) 按照第二种方案, 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

$$\begin{aligned}
 P(A_1 \cup A_2 \cup A_3 \cup A_4) &\geq P(A_1) \\
 &= \sum_{k=2}^{20} C_{20}^k (0.01)^k (0.99)^{20-k} \\
 &= 0.0175
 \end{aligned}$$

由此可见, 第一种方案较好。

注: 二项分布的概率计算可以调用 Excel 中的函数 BINOMDIST。

### ◎ 泊松分布 (记为 $X \sim \pi(\lambda)$ )

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

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

则称  $X$  为具有参数  $\lambda$  的泊松随机变量, 或称  $X$  服从参数为  $\lambda$  的泊松分布。

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

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

对于大的  $n$  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, 对于大的  $n$  和较小的  $p$ , 有

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

**【例 1.4】** 假定在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

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

**【例 1.5】** 假定每天在某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, 则此高速公路上今天没有发生事故的概率?

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一种常见现象，如在银行、超市和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为多少？

(2) 接下来的 3 个小时内有多于 20 辆汽车到达的概率为多少？

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为多少？

(2) 接下来的 10 分钟内有 3 个或 3 个以上使用者的概率？

读者可自行练习。

注：也可以调用 Excel 中的函数 POISSON 进行计算。

### ◎ 几何分布(记为 $X \sim g(n, p)$ )

设进行独立试验直到首次出现成功为止，其中每次试验出现成功的概率都是  $p$ ，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。其分布律为

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

**【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 0.90，那么直到第 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  $X \sim g(11, 0.1)$ ，由此

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

## 连续型随机变量及其分布

在某型号灯泡的寿命试验中，每一个被测试的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。我们将这样一个能取到一个区间中任意一个数的随机变量，称之为连续型随机变量。

连续型随机变量的分布函数：

$$F(x) = P\{X \leq x\} = \int_{-\infty}^x f(x) dx$$

这里， $f(x)$  是连续型随机变量  $X$  的分布密度函数。

### ◎ 均匀分布(记为 $X \sim U(a, b)$ )

$$\text{密度函数: } f(x) = \begin{cases} \frac{b-a}{2}, & a < x < b \\ 0, & \text{其他} \end{cases}$$

◎ 指数分布(记为  $X \sim E(\lambda)$ )

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

【例 1.8】 已知某种轮胎的使用寿命  $X \sim E(0.1)$  (单位: 万千米)。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 万千米的概率。

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 万千米的概率为

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

于是 5 只轮胎中至少有两只轮胎的行驶距离不足 30 万千米的概率

$$\begin{aligned} P &= 1 - \sum_{k=0}^1 C_5^k (0.9502)^k (0.0498)^{5-k} \\ &= 0.99996 \end{aligned}$$

◎ 正态分布(记为  $X \sim N(\mu, \sigma^2)$ )

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

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  $Z \sim N(0, 1)$  之间有

关系,  $Z = \frac{X - \mu}{\sigma}$ , 且

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

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的图像(见图 1.1) 可以看到, 此曲线完全由均值  $\mu$  和标准差  $\sigma$  决定, 事实上,  $\mu$  决定了密度函数曲线的位置, 也称为置参数;  $\sigma$  决定了曲线的形状, 也称为尺度参数。

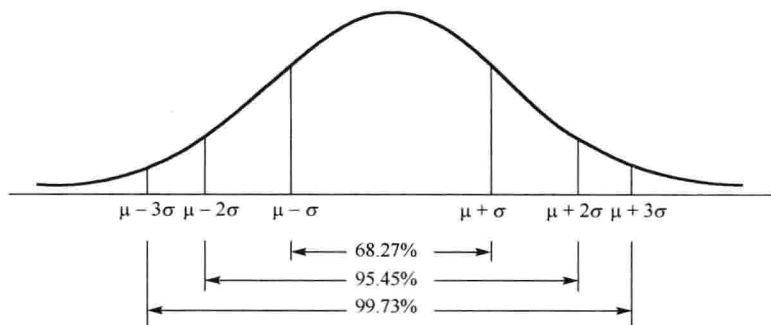


图 1.1 正态分布

正态随机变量的 3 个重要数据。

若  $X \sim N(\mu, \sigma^2)$ , 则

$$P\{\mu - \sigma \leq X \leq \mu + \sigma\} = 0.6827$$

$$P\{\mu - 2\sigma \leq X \leq \mu + 2\sigma\} = 0.9545$$

$$P\{\mu - 3\sigma \leq X \leq \mu + 3\sigma\} = 0.9973$$

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  $3\sigma$  准则, 其在产品的质量控制在有着重要应用。

**【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, 制定出一个录取条件, 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  $X$  表示为请者的入学分数, 则  $X \sim N(490, 61^2)$ 。记最低录取分数线为  $x_{0.01}$ , 则有

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

0.01, 这里,  $Z \sim N(0, 1)$ , 查表得  $\frac{x_{0.01} - 490}{61} = 2.3263$ , 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 中统计计算中的相应函数方便进行, 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, 则称其为随机变量  $X$  的  $k$  阶原点矩,  $k = 1, 2, \dots$ ; 若  $E(X - E(X))^k$  存在, 则称其为随机变量的  $k$  阶中心矩,  $k = 2, 3, \dots$ 。特别地, 称随机变量  $X$  的阶原点矩  $E(X)$  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,  $\sqrt{E(X - E(X))^2}$  称为随机变量  $X$  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。

#### 离散情形

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

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

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

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

(2) 若  $X \sim b(n, p)$ , 则

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

(3) 若  $X \sim g(n, p)$ , 则

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

(4) 若  $X \sim \pi(\lambda)$ , 则

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

**【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, 工程队可得 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  $X$  的分布律为

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:

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

### 连续情形

设  $X$  为连续型随机变量, 则  $X$  的数学期望定义为:

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

若积分  $\int_{-\infty}^{+\infty} |xf(x)| dx$  收敛。

(1) 若  $X \sim U(a, b)$ , 则

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

(2) 若  $X \sim E(\lambda)$ , 则

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

(3) 若  $X \sim N(\mu, \sigma^2)$ , 则

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

**【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下:

表 1.2 四个环节相关资料

环 节	所需时间的期望(单位: 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:这里涉及四个正态分布,即四个环节完工时间  $X_1, X_2, X_3, X_4$  分别服从正态分布  $N(18, 8), N(12, 5), N(27, 6), N(8, 2)$ 。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,

$$\begin{aligned} P\{X > 60\} &= 1 - P\{X \leq 60\} = 1 - P\left\{\frac{X - 65}{\sqrt{21}} \leq \frac{60 - 65}{\sqrt{21}}\right\} \\ &= 1 - \Phi\left(-\frac{5}{\sqrt{21}}\right) = 0.8621 \end{aligned}$$

### 1.1.3 分位点

设随机变量  $X$  的分布函数为  $F(x)$ , 称满足下列方程的实数  $f_\alpha$  为随机变量  $X$  的分布的上  $\alpha$  分位点。

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

对于标准正态分布,将其上  $\alpha$  分位点记为  $z_\alpha$ , 即  $z_\alpha$  满足方程:

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

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。我们将常用的随机变量的分布及其数字特征归纳如下,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 $n, p, 0 < p < 1$	$P\{X = k\} = C_n^k p^k (1-p)^{n-k},$ $k = 0, 1, 2, \dots, n$	$np$	$np(1-p)$
泊松分布, 参数为 $\lambda, \lambda > 0$	$P\{X = k\} = \frac{\lambda^k}{k!} e^{-\lambda},$ $k = 0, 1, 2, \dots$	$\lambda$	$\lambda$
几何分布, 参数为 $p, 0 < p < 1$	$P\{X = n\} = p(1-p)^{n-1},$ $n = 1, 2, \dots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$

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	$f(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{其他} \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
指数分布, 参数 $\lambda, \lambda > 0$	$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
伽玛分布, 参数 $(n, \lambda), \lambda > 0$	$f(x) = \begin{cases} \frac{\lambda e^{-\lambda x} (\lambda x)^{n-1}}{(n-1)!}, & x \geq 0 \\ 0, & x < 0 \end{cases}$	$\frac{n}{\lambda}$	$\frac{n}{\lambda^2}$
正态分布, 参数 $(\mu, \sigma^2)$	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$ $-\infty < x < \infty$	$\mu$	$\sigma^2$