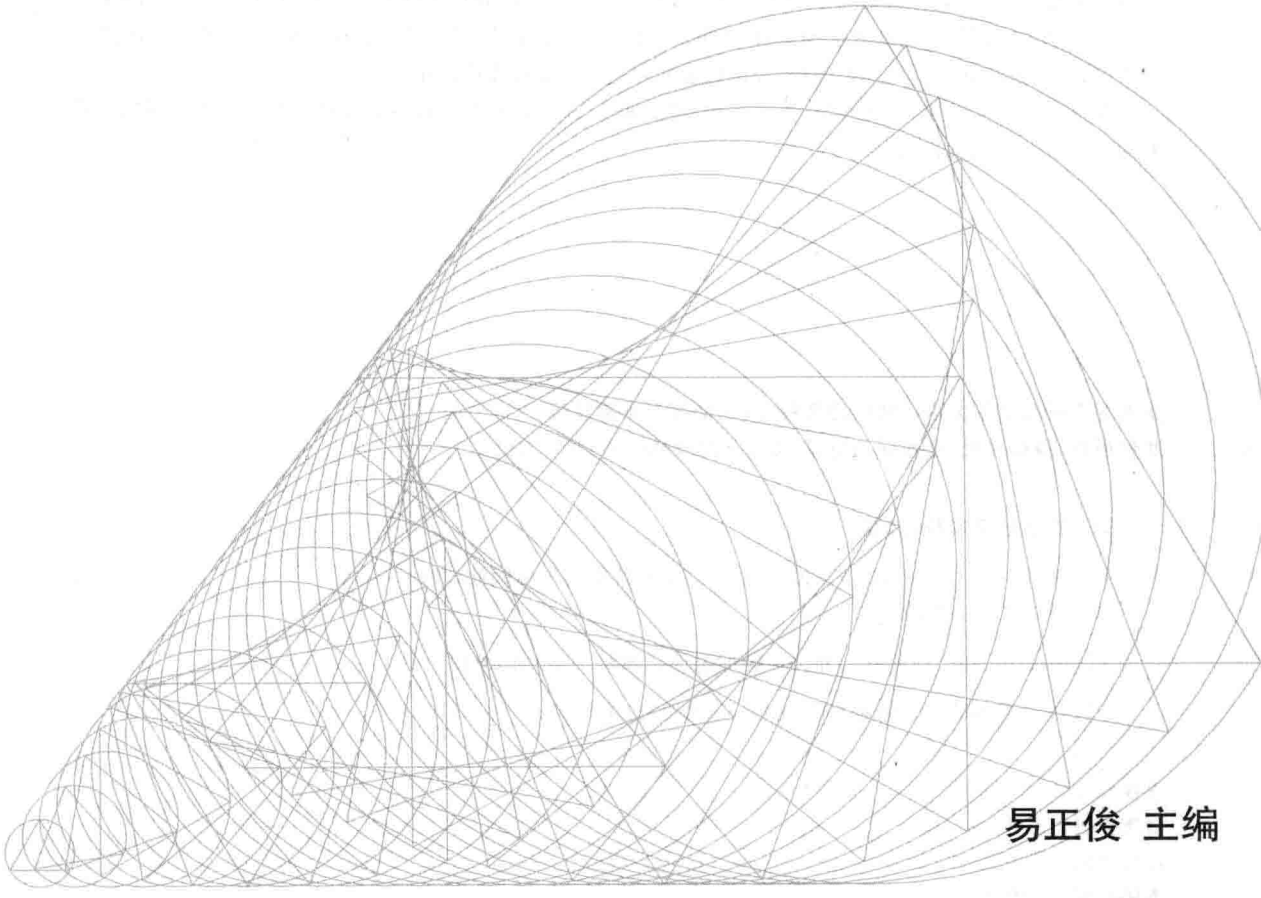
全国工程专业学位研究生教育国家级规划教材

易正俊　主编

数理统计及其工程应用

http://www.tup.com.cn

清华大学出版社



易正俊 主编

数理统计及其工程应用

清华大学出版社

内 容 简 介

本书是专为工程硕士和专业硕士学习数理统计及其工程应用而编写的教材.全书共8章,主要内容有:统计的基本概念及抽样分布、参数估计、假设检验、方差分析、正交试验设计、回归分析、系统聚类分析和主成分分析;每章配有R软件、SPSS软件或Excel软件等统计分析软件及相应的训练案例;习题的设置依据培养学生不同能力的要求分为A,B两组,A组主要是训练学生的应用能力,B组是提升学生的理论基础水平,书后附有概率基础知识回顾、分位数表和习题的部分答案或提示.

本书讲解简明扼要,图文并茂,注重应用,覆盖面广,也可以作为统计专业本科学生的教材及实际工作者的应用参考书和工具书.

图书在版编目(CIP)数据

数理统计及其工程应用/易正俊主编. —北京:清华大学出版社,2014
ISBN 978-7-302-36438-2

Ⅰ.①数… Ⅱ.①易… Ⅲ.①数理统计—教材 Ⅳ.①O212

中国版本图书馆CIP数据核字(2014)第095923号

责任编辑:刘 颖
封面设计:常雪影
责任校对:王淑云
责任印制:刘海龙

出版发行:清华大学出版社
网 址:http://www.tup.com.cn, http://www.wqbook.com
地 址:北京清华大学学研大厦A座 邮 编:100084
社 总 机:010-62770175 邮 购:010-62786544
投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn
质量反馈:010-62772015, zhiliang@tup.tsinghua.edu.cn
印 装 者:北京嘉实印刷有限公司
经 销:全国新华书店
开 本:185mm×260mm **印 张**:14.75 **字 数**:355千字
版 次:2014年7月第1版 **印 次**:2014年7月第1次印刷
印 数:1～3000
定 价:33.00元

产品编号:057930-01

PERFACE

前言

编者从事数理统计课程教学二十几年，主要是讲授数理统计的原理和方法，偏重于理论方面的教学；学生学习这门课程也主要是应付期终考试，虽然学生取得了相当满意的成绩，但在实际应用中无法用所学的数理统计知识来解决实际问题，学生在工作中遇到数理统计的应用问题只有回到母校找老师请教解决的方法.出现这种情况主要是由于我们的教材和对学生的考核标准有问题，在实际中有用的内容老师没有讲，教材没有写(即便是写了也是打了 * 号)，不作为考核内容.这给学生一种错误的感觉：数理统计课程只是考试有用，在实际中没有用.

全国工程专业学位研究生教育指导委员会根据社会实际工程领域对人才的需求，提出了工程硕士课程教学改革设想和指导性意见，旨在服务于行业创新发展的需求，提升职业能力，注重解决实际问题，提高在实践中发现问题、分析问题和解决问题的能力，通过整理和提炼实践工作中的问题，综合运用所学知识分析并解决问题，培养在实际工作中发现问题的敏感性、分析问题的科学性、处理问题的有效性.

“数理统计及其工程应用”是工程硕士培养的一门重要公共基础课程，在工程领域有广泛的应用.全国工程专业学位研究生教育指导委员会指派重庆大学易正俊教授组织编写《数理统计及其工程应用》教材，紧扣解决工程实际问题能力这一核心目标，为工程硕士、专业硕士的现有职业或未来职业提供专业支持，培养具有竞争优势的应用型创新创业人才.教材初稿完成后，由全国工程专业学位研究生教育指导委员会副主任陈子辰教授（浙江大学）和秘书长沈岩（清华大学）组织专家组成员齐欢教授（华中科技大学）、周杰教授（清华大学）、李大美教授（武汉大学）、韩中庚教授（解放军信息工程大学）对教材初稿进行讨论与修改.

本教材是专为工程硕士和专业硕士而编写的，具有以下特色：

1. 注重质量提升，突出职业需求导向，加强案例教学.案例的选取参考了国内外优秀教材和学术论文，博采众家之长，体现案例的实用性和趣味性，激发学生学习的积极性.

2. 训练学生借助统计分析软件（如 Excel 软件、R 软件和 SPSS 软件等）解决工程领域的实际问题，培养学生工程应用的意识和素质，提高解决工程实际问题的能力.

3. 基本概念和基本理论尽可能从学生熟悉的背景知识引入，采用几何图形等方法加强学生对基本理论和基本方法的理解，淡化比较复杂的理论推导，增强教材的可读性和可接受性.

4. 重视反例在学生理解、掌握基本概念和基本理论中的重要作用.

5. 习题的设置依据培养学生不同能力的要求分为 A，B 两组，A 组主要是训练学生的应用能力，B 组是提升学生的理论基础水平.

本教材由易正俊教授担任主编，参加编写的作者还有颜军、刘朝林、荣腾中、彭智军、曹术存、袁玉兴和罗秀娟.

教材的编写得到教育部学位管理与研究生教育司、重庆大学研究生院、重庆大学数学与统计学院的资助；重庆大学数学与统计学院穆春来和西南大学原校长宋乃庆对教材的编写提出了宝贵的意见.编者在此深表感谢！

写好这样一本具有实际应用价值的教材，编者深感难度很大.由于编者学识有限，书中不妥之处真诚地欢迎读者批评指正.

编　者

2014 年 4 月

CONTENTS

目录

Chapter 1

第1章 统计的基本概念及抽样分布

1.1 统计的基本概念

1.1.1 总体、样本与统计量

总体(X)是指研究对象的全体；**个体**是组成总体的每个单元；**样本**是从总体中随机抽取的 n 个个体 $X_1, X_2, \cdots, X_n$，n 称为**样本容量**. 一旦抽取了一个样本 $X_1, X_2, \cdots, X_n$ 进行试验，抽取样本后得到一组数据$(x_1, x_2, \cdots, x_n)$，这组数据称为样本 $X_1, X_2, \cdots, X_n$ 的一组观察值，样本观察值是随着抽样的变化而变化的.

抽样的目的是用样本的特征去代表总体的特征. 在有些情况下，我们不可能对总体的每个个体进行逐一试验，特别是具有破坏性的试验. 如检验一批炮弹是否合格，检验办法是将抽取的炮弹进行试爆，这个检验就是破坏性的；对总体的量很大，个体比较小的检验，虽然检验可能不是破坏性的，但对总体的检验也不可能逐一进行. 如对豌豆种子的检验就是总体大、个体小. 抽样包括简单随机抽样、分群抽样和分层抽样. 本教材的抽样是指简单随机抽样，简单随机抽样所获得的样本 $X_1, X_2, \cdots, X_n$ 称为简单样本(simple sample). 它满足以下两点：

(1) 独立性：要求样本 $X_1, X_2, \cdots, X_n$ 为相互独立的随机变量；

(2) 代表性：要求每个样本 $X_i(i=1,2,\cdots,n)$ 与总体 X 具有相同的分布.

从简单样本的定义可以看出：简单样本是有放回地抽取得到的样本. 在实际工作中，我们的抽样都是无放回地抽样，从理论上说就不再是简单样本. 但总体中个体的数目很大，从中抽取一些个体对总体成分没有太大的影响，可近似地看成有放回的抽样，其样本仍可看成是独立同分布的.

简单随机抽样有很多的益处：抽样单元的随机选取排除了调查者的偏见，这种偏见可能调查者并没有意识到；与完全枚举相比，小样本减少很多成本，调查更省时；小样本的结论实际上可能比完全枚举更精确. 小样本的数据质量更容易监控，完全枚举需要更多的员工去实施；随机抽样技术使得抽样误差的估计变得可能；在抽样设计时，通常可以确定出满足预设误差水平的样本尺寸.

例 1.1 某灯泡厂进行技改并扩大了生产规模，要求生产的灯泡寿命在 1000h 以上才算合格品. 现从技改后生产的第一批灯泡中随机抽取 4 个，测得其使用寿命分别为 1200，1120，980，1350(单位：h)，试叙述总体、样本、样本容量、样本观察值.

解 总体是灯泡的使用寿命 X 的取值全体；抽取 4 个个体 X_1, X_2, X_3, X_4 就是一个样本，样本容量是 4；1200，1120，980，1350 是一组样本观察值.

1.1.2 样本的联合分布函数和联合分布密度函数

设总体 X 的分布函数为 $F(x)$，$X_1, X_2, \cdots, X_n$ 是来自总体 X 的样本，则该样本的联合分布函数为

$$F(x_1, x_2, \cdots, x_n) = P(X_1 \leqslant x_1, X_2 \leqslant x_2, \cdots, X_n \leqslant x_n) = \prod_{i=1}^{n} F(x_i).$$

当总体 X 是连续型随机变量且具有密度函数 $f(x)$ 时，则样本的联合密度函数为

$$f(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} f(x_i).$$

当总体 X 是离散型随机变量且具有分布律 $P(X=x_i)=p_i$ 时，则样本 $X_1, X_2, \cdots, X_n$ 的联合分布律为

$$P(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n) = \prod_{i=1}^{n} P(X = x_i).$$

例 1.2 假设总体 X 服从指数分布，密度函数为

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geqslant 0, \\ 0, & x < 0, \end{cases}$$

从总体中抽取容量为 n 的样本 $X_1, X_2, \cdots, X_n$，写出样本的联合分布密度函数.

解 X_i 的密度函数为

$$f(x_i) = \begin{cases} \lambda e^{-\lambda x_i}, & x_i \geqslant 0, \\ 0, & x_i < 0, \end{cases}$$

$X_1, X_2, \cdots, X_n$ 是简单随机样本，且相互独立，所以 $X_1, X_2, \cdots, X_n$ 的联合密度函数等于边际密度函数的乘积，即

$$f(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} f(x_i) = \lambda^n e^{-\sum_{i=1}^{n} x_i}, \quad x_i \geqslant 0, i = 1, 2, \cdots, n.$$

例 1.3 设总体 $X \sim B(1, p)$，$X_1, X_2, \cdots, X_n$ 是来自总体的一个样本，写出样本的联合密度函数.

解 总体 X 的概率函数可以写成

$$f(x) = p^x (1-p)^{1-x}, \quad x = 0, 1.$$

X_i 的概率函数为

$$f(x_i) = p^{x_i} (1-p)^{1-x_i}, \quad x_i = 0, 1.$$

样本 $X_1, X_2, \cdots, X_n$ 是相互独立的，所以联合概率函数等于边际密度函数的乘积，为

$$f(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} f(x_i) = p^{\sum_{i=1}^{n} x_i} (1-p)^{n-\sum_{i=1}^{n} x_i}, \quad x_i = 0, 1.$$

1.1.3 统计量

1. 统计量的定义

统计量是不含参数的样本 $X_1, X_2, \cdots, X_n$ 的函数或即便含参数，但参数是已知的，记为 $T = T(X_1, X_2, \cdots, X_n)$，一旦获得样本的观察值，代入统计量得到一个数值. 例如 $X_1, X_2, \cdots, X_n$ 为来自总体 X 的一个样本，$T = \sum_{i=1}^{n} X_i$ 是一个统计量.

若 $X \sim N(\mu,\sigma^2)$，$X_1,X_2,\cdots,X_n$ 为 X 的一个样本，$\sum\limits_{i=1}^{n}\dfrac{X_i-\mu}{\sigma}$ 就不是统计量，因为含有未知的参数 μ 和 σ.

2. 常见统计量

(1) 样本均值：$\overline{X}=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i$.

(2) 样本方差：$S^2=\dfrac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2=\dfrac{1}{n-1}\left(\sum\limits_{i=1}^{n}X_i^2-n\overline{X}^2\right)$.

(3) 样本标准差：$S=\sqrt{\dfrac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2}$.

(4) 样本 k 阶原点矩：$M_k=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i^k$，$k=1,2,\cdots$

(5) 样本 k 阶中心矩：$M_k^*=\dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i-\overline{X})^k$，$k=1,2,\cdots$

3. 样本均值 $\overline{X}$ 的性质

(1) $\sum\limits_{i=1}^{n}(X_i-\overline{X})=0$.

(2) 若总体 X 的均值、方差存在，且 $EX=\mu$，$DX=\sigma^2$，则

$$E\overline{X}=\mu,\quad D\overline{X}=\frac{\sigma^2}{n}.$$

(3) 当 $n\to\infty$ 时，$\overline{X}\xrightarrow{P}\mu$.

证明 (1) $\sum\limits_{i=1}^{n}(X_i-\overline{X})=\sum\limits_{i=1}^{n}X_i-n\overline{X}=n\dfrac{\sum\limits_{i=1}^{n}X_i}{n}-n\overline{X}=n\overline{X}-n\overline{X}=0$.

(2) $E\overline{X}=E\left(\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i\right)=\dfrac{1}{n}\sum\limits_{i=1}^{n}EX_i=\dfrac{1}{n}\sum\limits_{i=1}^{n}EX=\mu$，

$D\overline{X}=D\left(\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i\right)=\dfrac{1}{n^2}\sum\limits_{i=1}^{n}DX_i=\dfrac{1}{n^2}\sum\limits_{i=1}^{n}DX=\dfrac{1}{n^2}n\sigma^2=\dfrac{\sigma^2}{n}$.

(3) 由概率论中的大数定律知，当 $n\to\infty$ 时，$\overline{X}\xrightarrow{P}\mu$.

4. 样本方差 S^2 的性质

(1) 如果 DX 存在，则 $ES^2=DX$，$EM_2^*=\dfrac{n-1}{n}DX$；

(2) 对任意实数 μ，有 $\sum\limits_{i=1}^{n}(X_i-\overline{X})^2\leqslant\sum\limits_{i=1}^{n}(X_i-\mu)^2$.

证明 (1) $ES^2=E\left[\dfrac{1}{n-1}\left(\sum\limits_{i=1}^{n}X_i^2-n\overline{X}^2\right)\right]=\dfrac{1}{n-1}\left(\sum\limits_{i=1}^{n}EX_i^2-nE\overline{X}^2\right)$

$$=\frac{n}{n-1}(EX^2-E\overline{X}^2)=\frac{n}{n-1}(DX+(EX)^2-D\overline{X}-(E\overline{X})^2)$$

$$=\frac{n}{n-1}\left(DX+(EX)^2-\frac{DX}{n}-(EX)^2\right)=DX,$$

$$M_2^*=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2=\frac{(n-1)S^2}{n},$$

$$EM_2^*=E\frac{(n-1)S^2}{n}=\frac{(n-1)ES^2}{n}=\frac{(n-1)DX}{n};$$

$$(2)\ \sum_{i=1}^{n}(X_i-\overline{X})^2=\sum_{i=1}^{n}((X_i-\mu)+(\mu-\overline{X}))^2$$

$$=\sum_{i=1}^{n}(X_i-\mu)^2+n(\mu-\overline{X})^2+2(\mu-\overline{X})\sum_{i=1}^{n}(X_i-\mu)$$

$$=\sum_{i=1}^{n}(X_i-\mu)^2+n(\mu-\overline{X})^2-2(\mu-\overline{X})(n\mu-\sum_{i=1}^{n}X_i)$$

$$=\sum_{i=1}^{n}(X_i-\mu)^2-n(\mu-\overline{X})^2\leqslant\sum_{i=1}^{n}(X_i-\mu)^2.$$

例 1.4 设总体 $X\sim U[0,\theta],\theta>0,X_1,X_2,\cdots,X_n$ 为样本，求 $E\overline{X},D\overline{X},EM_2^*$.

解 $E\overline{X}=EX=\frac{\theta}{2},D\overline{X}=\frac{1}{n}DX=\frac{1}{n}\frac{(\theta-0)^2}{12}=\frac{\theta^2}{12n}$,

$$EM_2^*=\frac{n-1}{n}DX=\frac{(n-1)\theta^2}{12n}.$$

1.2 顺序统计量、经验分布函数和直方图

1.2.1 顺序统计量

$X_1,X_2,\cdots,X_n$ 为总体 X 的样本，$x_1,x_2,\cdots,x_n$ 为样本观察值，将样本观察值 $x_1,x_2,\cdots,x_n$ 按从小到大的递增顺序进行排列：$x_{(1)}\leqslant x_{(2)}\leqslant\cdots\leqslant x_{(n)}$，$X_{(1)}$ 取最小观察值，$X_{(2)}$ 取次小值，$\cdots$，$X_{(n)}$ 取最大观察值，$X_{(1)},X_{(2)},\cdots,X_{(n)}$ 称为顺序统计量. $X_{(1)}$ 称为最小顺序统计量，$X_{(n)}$ 称为最大顺序统计量. $R=X_{(n)}-X_{(1)}$ 称为极差，极差在实际中用来衡量方差的大小，反映了随机变量 X 取值的分散程度.

$$\widetilde{X}=\begin{cases}X_{(\frac{n+1}{2})}, & n\text{ 为奇数},\\ \frac{1}{2}\left(X_{(\frac{n}{2})}+X_{(\frac{n}{2}+1)}\right), & n\text{ 为偶数}\end{cases}$$

为样本中位数. 样本中位数反映了随机变量 X 在实轴上分布的位置特征.

1.2.2 最大最小顺序统计量的分布

设 $F(x),\varphi(x)$ 分别为总体 X 的分布函数和分布密度函数，$X_1,X_2,\cdots,X_n$ 为 X 的样本，$F_{X_{(n)}}(x),\varphi_{X_{(n)}}(x)$ 分别为 $X_{(n)}$ 的分布函数和分布密度函数，$F_{X_{(1)}}(x),\varphi_{X_{(1)}}(x)$ 分别为 $X_{(1)}$ 的分布函数和分布密度函数，则对任意的实数 x，有

$$F_{X_{(n)}}(x)=P(X_{(n)}\leqslant x)=P(\max\{X_1,X_2,\cdots,X_n\}\leqslant x)$$

$$=P(X_1\leqslant x,X_2\leqslant x,\cdots,X_n\leqslant x)=\prod_{i=1}^{n}P(X_i\leqslant x)=F^n(x),$$

$$\varphi_{X_{(n)}}(x)=[F^n(x)]'=nF^{n-1}(x)\varphi(x),$$

$$\begin{aligned}F_{X_{(1)}}(x)&=P(X_{(1)}\leqslant x)=P(\min\{X_1,X_2,\cdots,X_n\}\leqslant x)\\&=1-P(\min\{X_1,X_2,\cdots,X_n\}>x)\\&=1-P(X_1>x,X_2>x,\cdots,X_n>x)\\&=1-P(X_1>x)P(X_2>x)\cdots P(X_n>x)\\&=1-[1-F(x)]^n,\end{aligned}$$

$$\varphi_{X_{(1)}}(x)=-n[1-F(x)]^{n-1}(-\varphi(x))=n\varphi(x)[1-F(x)]^{n-1}.$$

例 1.5 设总体 $X\sim U[0,\theta]$，$\theta>0$，$X_1,X_2,\cdots,X_5$ 为 X 的样本，分别求 $X_{(1)},X_{(5)}$ 的密度函数 $\varphi_{X_{(1)}}(x),\varphi_{X_{(5)}}(x)$.

解 因为 $X\sim U[0,\theta]$，$\theta>0$，所以 X 的密度函数与分布函数分别为

$$\varphi(x)=\begin{cases}\dfrac{1}{\theta}, & x\in[0,\theta],\\ 0, & x\notin[0,\theta],\end{cases}\qquad F(x)=\begin{cases}0, & x\leqslant 0,\\ \dfrac{x}{\theta}, & 0<x\leqslant\theta,\\ 1, & x>\theta,\end{cases}$$

$$\varphi_{X_{(5)}}(x)=5F^4(x)\varphi(x)=\begin{cases}\dfrac{5x^4}{\theta^5}, & x\in[0,\theta],\\ 0, & x\notin[0,\theta],\end{cases}$$

$$\varphi_{X_{(1)}}(x)=5\varphi(x)[1-F(x)]^4=\begin{cases}\dfrac{5}{\theta}\left(1-\dfrac{x}{\theta}\right)^4, & x\in[0,\theta],\\ 0, & x\notin[0,\theta].\end{cases}$$

1.2.3 经验分布函数与直方图

样本是总体的代表和反映，总体 X 的分布函数 $F(x)$ 称为理论分布，往往是未知的，在实际工作中一般可用经验分布函数去推断总体的分布，用直方图去描述（推断）总体 X（连续）的密度函数.

1. 经验分布函数

设 $x_1,x_2,\cdots,x_n$ 为来自总体 X 的样本观察值，将这些值从小到大排序：

$$x_{(1)}\leqslant x_{(2)}\leqslant\cdots\leqslant x_{(n)},$$

对任意实数 x，有

$$F_n(x)=\begin{cases}0, & x<x_{(1)},\\ \dfrac{k}{n}, & x_{(k)}\leqslant x<x_{(k+1)},\quad k=1,2,\cdots,n-1,\\ 1, & x\geqslant x_{(n)},\end{cases}$$

则称 $F_n(x)$ 为总体 X 的经验分布函数. 经验分布函数具有下列性质：

(1) $0\leqslant F_n(x)\leqslant 1$；

(2) $F_n(-\infty)=0$，$F_n(+\infty)=1$；

(3) $F_n(x+0)=F_n(x)$（右连续性）.

经验分布函数 $F_n(x)$ 同样满足总体分布函数 $F(x)$ 的三个基本性质，值得注意的是：对于样本的不同观察值 $x_1,x_2,\cdots,x_n$ 得到的经验分布函数 $F_n(x)$ 是不同的，在试验之前，对固

定的 x 值，$F_n(x)$是一个随机变量，当然也是一个统计量. 样本容量越大，用经验分布函数 $F_n(x)$作为分布函数$F(x)$的估计将会越精准. 图 1.1 给出了某正态总体 X 的理论分布函数 $F(x)$的曲线和经验分布函数 $F_n(x)$的曲线拟合情况，从此图形可以看出，经验分布函数可以近似代替总体的分布函数.

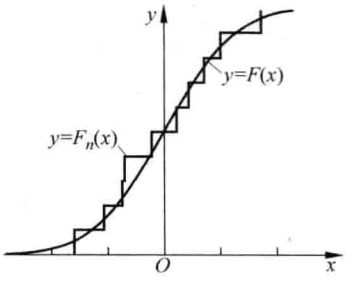


图 1.1 正态分布与经验分布拟合曲线

2. 直方图

直方图是用于近似连续型总体密度函数的曲线. 当样本容量 n 越大，且分组比较细时，近似程度也就越好.

假设 $x_1, x_2, \cdots, x_n$ 为连续型总体 X 的样本观察值. 构造直方图的步骤：

步骤 1 求出样本观察值 $x_1, x_2, \cdots, x_n$ 的极差 $x_{(n)} - x_{(1)}$.

步骤 2 确定组数与组距，将包含 $x_{(1)}, x_{(n)}$ 的区间$[a,b]$分成 m 个小区间：$[t_{i-1}, t_i)(i=1,2,\cdots,m)$，其中 a 略小于 $x_{(1)}$，b 略大于 $x_{(n)}$. 一般组数由经验公式 $m \approx 1.87(n-1)^{0.4}$ 确定，组距$=\dfrac{b-a}{m}$. $t_i = t_{i-1} + \dfrac{b-a}{m}, (i=1,2,\cdots,m, t_0 = a, t_m = b)$.

步骤 3 计算落入各区间样品个数，记落入区间$[t_{i-1}, t_i)$内的样品个数为 ν_i，称它为样本落入第 i 个区间的频数，称 $f_i = \dfrac{\nu_i}{n}$为样本落入区间$[t_{i-1}, t_i)$内的频率.

步骤 4 作图. 在 xOy 平面上，以 x 轴上第 i 个小区间$[t_{i-1}, t_i)$为底，以 $y_i = \dfrac{f_i}{t_i - t_{i-1}}$为高作第 i 个长方形，这样一排竖着的长方形所构成的图形就叫做直方图. 第 i 个长方形的面积为 f_i，所有长方形面积之和为 1. 沿直方图边缘的曲线就是连续型总体的密度函数曲线的近似曲线.

例 1.6 某轧钢厂生产一批同型号的钢材，为研究这批钢材的抗张力，从中随机抽取了 76 个样品做张力实验，测出数据见表 1.1.

表 1.1 钢材抗张力数据表 kg/cm²

41.0	37.0	33.0	44.2	30.5	27.0	45.0	28.5	31.2	33.5	38.5	41.5
42.0	45.5	42.5	39.0	38.8	35.5	32.5	29.6	32.6	34.5	37.5	39.5
42.8	45.1	42.8	45.8	39.8	37.2	33.8	31.2	29.0	35.2	37.8	41.2
43.8	48.0	43.6	41.8	36.6	34.8	31.0	32.0	33.5	37.4	40.8	44.7
40.2	41.3	38.8	34.1	31.8	34.6	38.3	41.3	30.0	35.2	37.5	40.5
38.1	37.3	37.1	41.5	29.5	29.1	27.5	34.8	36.5	44.2	40.0	44.5
40.6	36.2	35.8	31.5								

根据表中的数据，作出直方图.

解 根据作直方图的步骤，计算结果如表 1.2，其图形如图 1.2 所示.

表 1.2 直方图计算表

分组区间	频数 ν_i	频率 f_i	纵坐标值 y_i
[27,30)	8	0.105	0.035
[30,33)	10	0.132	0.044
[33,36)	12	0.158	0.053
[36,39)	17	0.224	0.074
[39,42)	14	0.184	0.061
[42,45)	11	0.145	0.048
[45,48)	4	0.053	0.018

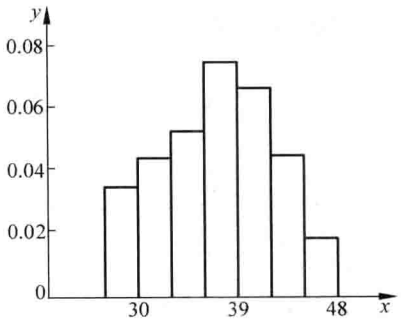


图 1.2 直方图

1.3 抽样分布及分位数

统计量是对总体分布和总体所含参数进行推断的基础，由于统计量是一个随机变量，称统计量的分布为抽样分布(sampling distribution). 一般确定一个统计量的分布是十分复杂的，要用到许多概率知识. 本节将讨论正态总体下一些常用的抽样分布.

1.3.1 正态分布的导出分布

1. χ^2(卡方)分布

(1) χ^2 分布的定义

设 $X_1, X_2, \cdots, X_n$ 相互独立，且 $X_i \sim N(0,1)(i=1,2,\cdots,n)$，则称 $\sum_{i=1}^{n} X_i^2$ 为服从自由度为 n 的卡方分布. 记为 $\chi^2 = \sum_{i=1}^{n} X_i^2 \sim \chi^2(n)$.

(2) χ^2 分布的密度函数及其图像

$\chi^2(n)$分布的密度函数为

$$f(x,n)=\begin{cases}\dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\dfrac{n}{2}\right)}x^{\frac{n}{2}-1}\mathrm{e}^{-\frac{x}{2}}, & x>0,\\ 0, & x\leqslant 0.\end{cases}$$

密度函数图像在第一象限内是非负的，如图 1.3 所示.

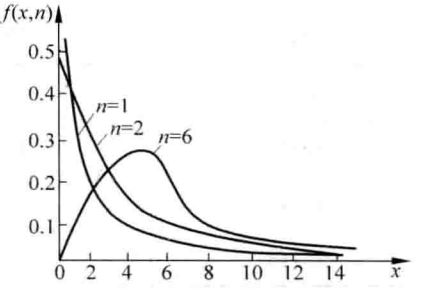


图 1.3 χ^2 分布的密度函数图像

(3) χ^2 分布的性质

① 若 $X\sim\chi^2(n)$，则 $EX=n, DX=2n$;

②(可加性)若 $X\sim\chi^2(n_1), Y\sim\chi^2(n_2)$，且 X,Y 相互独立，则 $X+Y\sim\chi^2(n_1+n_2)$.

证明 ① 因为 $X=\sum_{i=1}^{n}X_i^2$, $X_i\sim N(0,1)(i=1,2,\cdots,n)$，故

$$EX_i=0, DX_i=1.$$

又因为

$$DX_i=EX_i^2-(EX_i)^2,$$

故

$$EX_i^2=DX_i+(EX_i)^2=1,$$

所以

$$EX=\sum_{i=1}^{n}EX_i^2=n.$$

又因

$$\begin{aligned}EX_i^4&=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}x^4\mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x=\frac{-1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}x^3(-x\mathrm{e}^{-\frac{x^2}{2}})\mathrm{d}x\\&=\frac{-1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}x^3\mathrm{d}\mathrm{e}^{-\frac{x^2}{2}}=\frac{-1}{\sqrt{2\pi}}\left(x^3\mathrm{e}^{-\frac{x^2}{2}}\Big|_{-\infty}^{+\infty}-3\int_{-\infty}^{+\infty}x^2\mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x\right)\\&=3\int_{-\infty}^{+\infty}x^2\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x=3EX_i^2=3,\end{aligned}$$

所以

$$DX_i^2=EX_i^4-(EX_i^2)^2=3-1=2,\quad 故\quad DX=\sum_{i=1}^{n}DX_i^2=\sum_{i=1}^{n}2=2n.$$

② 由于 X 是 n_1 个独立的标准正态分布的平方和，Y 是 n_2 个独立的标准正态分布的平方和，又因 X,Y 是相互独立的，所以 $X+Y$ 是 n_1+n_2 个独立的标准正态分布的平方和，根据定义有 $X+Y\sim\chi^2(n_1+n_2)$.

例 1.7 设 $X_1,X_2,\cdots,X_n$ 独立同分布于 $N(\mu,\sigma^2)$，求 $E\sum_{i=1}^{n}(X_i-\mu)^2, D\sum_{i=1}^{n}(X_i-\mu)^2$.

解 因为 $\sum_{i=1}^{n}\left(\dfrac{X_i-\mu}{\sigma}\right)^2=\dfrac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2}\sim\chi^2(n)$，故

$$E\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2}=n, D\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2}=2n,$$

于是有