研究生教材

应用数理统计

施　雨　编著

西安交通大学出版社
XI'AN JIAOTONG UNIVERSITY PRESS

研究生教材

应用数理统计

施雨　编著

西安交通大学出版社

内容简介

本书比较系统地介绍了数理统计的基本概念、原理和方法。全书分为6章，内容包括数理统计的基本概念，参数估计，假设检验，方差分析与正交试验设计，回归分析以及统计决策理论与贝叶斯推断中的基本知识。各章均配有适量的习题。

本书适合作为高等院校工学、经管学科的研究生教材，也可作为理科高年级本科生教材。

图书在版编目(CIP)数据

应用数理统计/施雨编著．—西安：西安交通大学出版社，2005.9
（研究生系列教材）
ISBN 7-5605-2060-X

Ⅰ.应… Ⅱ.施… Ⅲ.数理统计—研究生—教材 Ⅳ.0212

中国版本图书馆CIP数据核字(2005)第069523号

书　　名	应用数理统计
编　　著	施雨
出版发行	西安交通大学出版社
地　　址	西安市兴庆南路25号(邮编：710049)
电　　话	(029)82668357，82667874(发行部) (029)82668315，82669096(总编办)
印　　刷	西安建筑科技大学印刷厂
字　　数	230千字
开　　本	850 mm×1 168 mm　1/32
插　　页	1
印　　张	9.125
版　　次	2005年9月第1版　2005年9月第1次印刷
印　　数	0001～3000
书　　号	ISBN 7-5605-2060-X/O·227
定　　价	14.00元

“研究生教材”总序

研究生教育是我国高等教育的最高层次，是为国家培养高层次的人才。他们必须在本门学科中掌握坚实的基础理论和系统的专门知识，以及从事科学研究工作或担负专门技术工作的能力。这些要求具体体现在研究生的学位课程和学位论文中。

认真建设好研究生学位课程是研究生培养中的重要环节。为此，我们组织出版这套“研究生教材”，以满足当前研究生教学，主要是公共课和一批新型的学位课程的教学需要。教材作者都是多年从事研究生教学工作，有着丰富教学和科学研究经验的教师。

这套教材首先着眼于研究生未来工作和高技术发展的需要，充分反映国内外的最新学术动态，使研究生学习之后，能迅速接近当代科技发展的前沿，以适应“四化”建设的要求；其次，也注意到研究生公共课程和学位课程应有它最稳定、最基本的内容，是研究生掌握坚实的基础理论和系统的专门知识所必要的，因此在研究生教材中仍应强调突出重点，突出基本原理和基本内容，以保持学位课程的相对稳定性和系统性，内容有足够的深度，而且对本门课程有较大的覆盖面。

这套“研究生教材”虽然从选题、大纲、组织编写到编辑出版，都经过了认真的调查论证和细致的定稿工作，但毕竟是第一次编辑这样的高层次教材系列，水平和经验都感不足，缺点与错误在所难免。希望通过反复的教学实践，广泛听取校内外专家学者和使用者的意见，使其不断改进和完善。

西安交通大学研究生院

西安交通大学出版社

前　言

与概率论一样，数理统计也是研究随机现象的统计规律性的一门数学学科。数理统计以概率论为基础，研究如何合理有效地收集受到随机性影响的数据，如何对所获得的数据进行整理和分析，从而为随机现象选择合适的数学模型并提供检验的方法，在此基础上对随机现象的性质、特点和统计规律做出推断和预测，直至为决策提供依据和建议。

由于随机现象是客观世界中普遍存在的一种现象，因而数理统计的应用十分广泛，在自然科学、社会科学、工程技术、军事科学、医药卫生以及工农业生产中都能用到数理统计的理论与方法。随着计算机的普及和软件技术的发展，多种使用便捷的统计软件的面世，使得各行各业中只要粗通统计知识的人，都可以方便地运用统计分析的各种工具，来为自己的研究课题服务。数理统计正在发挥着越来越大的作用，它的应用更加广泛深入。

本书的适用对象为工学、经管学科的研究生以及部分理科专业高年级本科生。编写本书的目的是希望对读者全面地提高统计修养提供一些帮助，故本书的取材比较广泛。

本书的第 1 章介绍数理统计的基本概念：总体，样本，统计量（包括样本矩、顺序统计量和经验分布函数等），一些重要分布（如，Γ 分布、β 分布、χ^2 分布、t 分布及 F 分布）以及正态总体下的抽样分布定理。其中，在顺序统计量概率密度的推导中采用了直观易懂的“概率微元法”。

第 2 章讨论统计推断的基本问题之一——参数估计，内容包括：点估计的概念与常用求法（矩法、极大似然法和顺序统计量法），估计量的评判标准（无偏性、有效性和相合性），充分统计量与完备统计量（有助于建立一致最小方差无偏估计理论），区间估计。

第 3 章讨论统计推断的另一基本问题——假设检验，内容有：

假设检验基本概念(假设的形式与检验规则,两类错误与功效函数,显著性检验的一般步骤等),正态总体参数的假设检验,非正态总体参数的假设检验(指数分布及大样本下任意分布的参数假设检验),单侧假设检验,非参数假设检验(分布假设检验、相同性检验和独立性检验等),假设检验的基本理论(检验的评价标准,一致最大功效检验及似然比检验等)。

第4章介绍方差分析(单因素,双因素方差分析)与正交试验设计(正交表,正交试验方案的设计以及试验结果的分析)。

第5章介绍回归分析,包括一元及多元线性回归中的参数估计、回归模型与回归系数的假设检验以及预测问题。

第6章简要介绍了统计决策理论与贝叶斯推断中的一些基本概念和常用方法。

书中配有丰富的例题和适量的习题,书末给出了习题的参考答案。

本书中的图、表以及公式是以章节作区分的,如图1.2.3表示第1章第2节的第3个图,其余的类似。

承蒙聂赞坎教授于百忙之中审阅本书,聂先生提出了不少宝贵的意见和建议。本书的出版得到西安交通大学出版社叶涛先生的热情支持和鼓励,编者在此向他们致以衷心的感谢。

对编者而言,编写本书的过程也是一次向国内外同行专家的学习过程。对于引用了其中的某些定理的证法、习题与例题的相关书籍,均已列入书末的参考文献中,谨此向这些书的作者一并致谢。

最后,还要感谢曾听过编者授课的那些研究生们。在与他们教学互动中,编者得到了许多的鼓励和反馈信息,其中有些同学还为编者提供了部分很好的习题素材。恕编者无法一一列举他们的姓名,借本书出版之际,向这些研究生们表达编者的谢忱。

编　者

2005.6

目 录

第1章　数理统计的基本概念

本章介绍数理统计中的基本概念和常用术语以及一些重要统计量的分布.

1.1　总体、样本与统计量

1.1.1　总体及其分布

在数理统计中,称所研究的对象的全体为**总体**(或**母体**),总体中的元素称为**个体**. 若总体中的个体数目为有限,则称之为**有限总体**;否则就称之为**无限总体**.

例 1.1.1　有一批灯管共 10 万支,在研究这批灯管的平均使用寿命时,该批灯管的全部使用寿命就组成一个总体,而其中每个灯管的使用寿命是个体.

例 1.1.2　在检查某本书的印刷质量时,假定该书有 500 页,则该书所有页码的印刷错误数组成一个总体,而每一页上的印刷错误数是个体.

例 1.1.3　考察硕士研究生班 150 位同学的年龄分布,假定分成三类:$x \leqslant 21$;$21 < x \leqslant 28$;$x > 28$. 这样 150 人的年龄类别组成一个总体,每一个人的年龄类别就是个体.

从上面的例子可以看出,数理统计所关心的并非每个个体的所有属性,而是它的某一项或若干项数量指标 X(亦即 X 可以是向量)和该数量指标 X 在总体中的分布情况. 在上述例子中,数量指标 X 分别是使用寿命、印刷错误数、年龄类别,相应的总体则是由数量指标 X 可能取值的全体组成的集合. 一方面,说到总体必对

应某数量指标 X 可能取值的集合. 另一方面,研究任意数量指标 X,其可能取值的全体即构成一个总体. 因而,以后就把两者等同起来. 所谓总体的分布就是指数量指标 X 的分布.

在试验中,抽取了若干个个体就观察到了 X 的这样或那样的数值,因而数量指标 X 是一个随机变量(或随机向量),于是总体的分布也就是随机变量 X 的概率分布. 以后,将总体记作 X. 在总体 X 服从某种分布时,为简便起见,有时就简称其为某总体. 例如,当总体 X 服从 $N(\mu,\sigma^2)$ 分布时,简称 X 为正态总体.

1.1.2 样本及其分布

从总体中取得一部分个体,这一部分个体称为**样本**(或**子样**). 样本中的每个个体称为**样品**. 样本中的个体数目称为**样本容量**(或**子样容量**).

取得样本的过程称为**抽样**. 抽样中采用的方法称为**抽样法**. 在数理统计中,一般采用**随机抽样法**,即从总体中随意地抽取若干个个体.

随机抽样得到的样本,所含样品是有一定顺序的,通常按它被抽到的先后顺序排列. 从总体 X 随机抽样得到的样本用 $X_1,\cdots,X_n$ 表示,或者用 n 维随机向量 $\boldsymbol{X}\triangleq(X_1,\cdots,X_n)^{\mathrm{T}}$ 表示. 样本 $(X_1,\cdots,X_n)^{\mathrm{T}}$ 可能取值的全体称为**样本空间**,记为 $\mathscr{X}$,它是 n 维欧氏空间 $\mathbb{R}^n$ 的子集.

设有样本 $X_1,\cdots,X_n$,若 $X_1,\cdots,X_n$ 是独立同分布的(以后简记为 i. i. d.)且 X_1 的分布与总体 X 的分布相同,则称它为**简单随机样本**.

简单随机样本具有较为简单的概率结构且易于数学处理. 若无特别申明,以后我们就只讨论简单随机样本,并简称其为样本.

对于简单随机样本 $X_1,\cdots,X_n$,若总体 X 的分布函数为 $F(x)$,则样本 $X_1,\cdots,X_n$ 的联合分布函数为

$$F_S(x_1,\cdots,x_n)=\prod_{i=1}^{n}F(x_i)$$

若总体 X 具有概率密度 $f(x)$，则样本 $X_1,\cdots,X_n$ 的联合概率密度为

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

若总体 X 具有分布律(概率函数) $p(x)$，其中 $p(a_i)=P\{X=a_i\}$，$i=1,2,\cdots$. 则样本 $X_1,\cdots,X_n$ 的联合分布律(联合概率函数)为

$$p_S(x_1,\cdots,x_n)=\prod_{i=1}^{n}p(x_i)$$

其中 $x_1,\cdots,x_n$ 在集合 $\{a_1,a_2,\cdots\}$ 中取值.

数理统计中的主要任务 —— 统计推断正是基于样本分布 ($F_S(x)$，$f_S(x)$ 或 $p_S(x)$，其中 $x=(x_1,\cdots,x_n)^{\mathrm{T}}$) 所提供的信息来完成的.

说样本 $(X_1,\cdots,X_n)^{\mathrm{T}}$ 是 n 维随机向量，这是针对进行一次抽样前而言，实施了一次抽样后，得到的是一个实向量 $(x_1,\cdots,x_n)^{\mathrm{T}}$，它是样本 $(X_1,\cdots,X_n)^{\mathrm{T}}$ 的一个观察值，称为**样本值**. 为了简便起见，有时把样本和样本值统称为样本.

1.1.3　统计量

1. 统计量概念

样本是推断总体特性的依据，但在获得样本之后，并不能由样本直接进行统计推断，需要先对样本进行加工和提炼，把样本中所含的总体的相关信息集中起来，即，针对不同的问题构造出样本的适当函数. 这种样本的函数就称为统计量.

定义 1.1.1　设 $(X_1,\cdots,X_n)^{\mathrm{T}}$ 为总体 X 的一个样本，若 $g(x_1,\cdots,x_n)$ 为样本空间 $\mathcal{X}$ 到 $\mathbf{R}^k$ 的可测映射，且 g 中不含任何未知参数，则称 $t=g(X_1,\cdots,X_n)$ 为**统计量**.

注 (1) 当 $k=1$ 时，可测映射也称为可测函数. “可测函数”一词源于测度论，通俗地讲，$g(\cdot)$ 是可测函数就可以保证用 $g(\cdot)$ 来描述的各种事件都有概率可言. 自然科学与工程技术领域中的绝大多数函数，如连续函数、单调函数或者分段单调函数等都属于可测函数范畴.

(2) 当 $k>1$ 时，称 $t=(g_1(X_1,\cdots,X_n),\cdots,g_k(X_1,\cdots,X_n))^{\mathrm{T}}$ 为**向量值统计量**.

(3) 粗略地来说，统计量就是用作“统计推断的量”，因而它不能含未知参数.

例如，设 $(X_1,X_2,X_3)^{\mathrm{T}}$ 是来自正态总体 $N(\mu,\sigma^2)$ 的样本，其中 μ 为已知参数，σ^2 为未知参数，则 $X_1,\mu-(X_1+X_2+X_3)/3$，$\max(X_1,X_2,X_3)$ 是统计量，而 X_1/σ，$(X_1/\sigma)^2+(X_2/\sigma)^2+(X_3/\sigma)^2$ 不是统计量.

2. 样本矩

定义 1.1.2 设 $(X_1,\cdots,X_n)^{\mathrm{T}}$ 为总体 X 的一个样本，称统计量

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i \tag{1.1.1}$$

为**样本均值**；称统计量

$$S^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2=\frac{1}{n}\sum_{i=1}^{n}X_i^2-\overline{X}^2 \tag{1.1.2}$$

及

$$S^{*2}=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2 \tag{1.1.3}$$

分别为**样本方差**及**修正样本方差**，称样本方差的算术根 $S=\sqrt{S^2}$ 为**样本标准差**；称统计量

$$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k \tag{1.1.4}$$

及

$$B_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k \tag{1.1.5}$$

分别为**样本 k 阶原点矩**及**样本 k 阶中心矩**.

进行一次抽样后得到样本值 $x_1,\cdots,x_n$，把它代入定义 1.1.2 中各统计量的表达式，即可得到相应的统计量的观察值. 例如 $\bar{x} = \frac{1}{n}\sum_{i=1}^{n}x_i, s^{*2} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ 等等.

由大数定律可以证明，如果总体的 k 阶矩存在，那么样本的 k 阶矩依概率收敛于总体的 k 阶矩. 因此，当总体 X 的均值 μ，方差 σ^2 存在时，对任意的 $\varepsilon > 0$，有

$$\lim_{n\to\infty} P\{|\overline{X} - \mu| < \varepsilon\} = 1$$

$$\lim_{n\to\infty} P\{|S^2 - \sigma^2| < \varepsilon\} = 1$$

于是，在 n 很大时，可用一次抽样后所得的样本均值 $\bar{x}$ 和样本方差 s^2 分别作为总体 X 的均值 μ 和方差 σ^2 的近似值.

3. 顺序统计量及其分布

定义 1.1.3　设 $(X_1,\cdots,X_n)^{\mathrm{T}}$ 为总体 X 的一个样本，其观察值为 $(x_1,\cdots,x_n)^{\mathrm{T}}$，将 $x_1,\cdots,x_n$ 由小到大进行排列，依次记为 $x_{(1)},\cdots,x_{(n)}$，即 $x_{(1)} \leqslant \cdots \leqslant x_{(n)}$. 按下述方法定义统计量 $X_{(k)}$：当样本 $(X_1,\cdots,X_n)^{\mathrm{T}}$ 取值为 $(x_1,\cdots,x_n)^{\mathrm{T}}$ 时，规定 $X_{(k)}$ 取值为 $x_{(k)}$，由此得到的 $(X_{(1)},\cdots,X_{(n)})^{\mathrm{T}}$ 称为样本 $(X_1,\cdots,X_n)^{\mathrm{T}}$ 的**顺序统计量**或**次序统计量**，$X_{(k)}$ 称为样本 $(X_1,\cdots,X_n)^{\mathrm{T}}$ 的第 k 个顺序统计量，$X_{(1)}$ 称为样本 $(X_1,\cdots,X_n)^{\mathrm{T}}$ 的最小顺序统计量，$X_{(n)}$ 称为样本 $(X_1,\cdots,X_n)^{\mathrm{T}}$ 的最大顺序统计量.

例如，若样本 $(X_1,X_2,X_3,X_4)^{\mathrm{T}}$ 的两个观察值分别是 $(1.3,-0.5,0.7,2)^{\mathrm{T}}$，$(0.4,3.1,-0.7,0)^{\mathrm{T}}$，则顺序统计量 $(X_{(1)},X_{(2)},X_{(3)},X_{(4)})^{\mathrm{T}}$ 对应的观察值分别是 $(-0.5,0.7,1.3,2)^{\mathrm{T}}$，$(-0.7,0,0.4,3.1)^{\mathrm{T}}$.

对顺序统计量 $(X_{(1)},\cdots,X_{(n)})^{\mathrm{T}}$，显然有 $X_{(1)} \leqslant X_{(2)} \leqslant \cdots$

$\leqslant X_{(n)}$.

下面,我们来讨论顺序统计量的概率分布.

若总体 X 的概率密度为 $f(x)$,则 X 落入一个很小区间$(x-\mathrm{d}x,x)$ 的概率为

$$P\{x-\mathrm{d}x\leqslant X\leqslant x\}=f(x)\mathrm{d}x+o(\mathrm{d}x)$$

这里,$o(\mathrm{d}x)$ 表示 $\mathrm{d}x$ 的高阶无穷小,$f(x)\mathrm{d}x$ 称为 X 的概率微元.反之,若存在一个非负函数 $f(x)$ 使得上式成立,则 $f(x)$ 就是 X 的概率密度. 这种寻求 X 的概率密度的方法称为"概率微元法". 下面,我们就用概率微元法来推导各种顺序统计量的概率分布.

设总体 X 的分布函数为 $F(x)$,概率密度为 $f(x)$,$(X_1,\cdots,X_n)^{\mathrm{T}}$ 为总体 X 的一个样本,该样本的顺序统计量为 $(X_{(1)},\cdots,X_{(n)})^{\mathrm{T}}$.

(1)$X_{(k)}$ 的概率密度 $f_{(k)}(x)(1\leqslant k\leqslant n)$

把实数轴分为:$(-\infty,x-\mathrm{d}x]$,$(x-\mathrm{d}x,x]$,$(x,+\infty)$ 三个区间,其中 $\mathrm{d}x$ 取得足够小,使得样本观察值仅有一个落入区间$(x-\mathrm{d}x,x]$,而有两个或两个以上观察值落入该区间的概率为零或为 $o(\mathrm{d}x)$,这样一来,要使 $X_{(k)}$ 的观察值落入区间$(x-\mathrm{d}x,x]$,当且仅当样本观察值中有 $k-1$ 个落入$(-\infty,x-\mathrm{d}x]$,一个落入$(x-\mathrm{d}x,x]$,$n-k$ 个落入$(x,+\infty)$. 此时,

$$\begin{aligned}&P\{x-\mathrm{d}x<X_{(k)}\leqslant x\}\\&=\frac{n!}{(k-1)!1!(n-k)!}[F(x-\mathrm{d}x)]^{k-1}[f(x)\mathrm{d}x+o(\mathrm{d}x)]\\&\quad\cdot[1-F(x)]^{n-k}\end{aligned}$$

注意到

$$F(x-\mathrm{d}x)=F(x)-[f(x)\mathrm{d}x+o(\mathrm{d}x)]$$

以及小 o 的运算法则,可得

$$\begin{aligned}&P\{x-\mathrm{d}x<X_{(k)}\leqslant x\}\\&=\frac{n!}{(k-1)!(n-k)!}[F(x)]^{k-1}[1-F(x)]^{n-k}f(x)\mathrm{d}x+o(\mathrm{d}x)\end{aligned}$$

由此可知，$X_{(k)}$ 的概率密度为

$$f_{(k)}(x)=\frac{n!}{(k-1)!(n-k)!}[F(x)]^{k-1}[1-F(x)]^{n-k}f(x) \tag{1.1.6}$$

特别，$X_{(1)}$，$X_{(n)}$ 的概率密度分别为

$$f_{(1)}(x)=n[1-F(x)]^{n-1}f(x) \tag{1.1.7}$$

$$f_{(n)}(x)=n[F(x)]^{n-1}f(x) \tag{1.1.8}$$

(2) $X_{(k)}$ 与 $X_{(j)}$ 的联合概率密度 $f_{(k)(j)}(x,y)(1\leqslant k<j\leqslant n)$

设 $x<y$，把实数轴分为：$(-\infty,x-\mathrm{d}x]$，$(x-\mathrm{d}x,x]$，$(x,y-\mathrm{d}y]$，$(y-\mathrm{d}y,y]$，$(y,+\infty)$ 等五个区间，其中 $\mathrm{d}x,\mathrm{d}y$ 取得充分小，这样一来，要使 $X_{(k)}$ 的观察值落入区间 $(x-\mathrm{d}x,x]$ 且 $X_{(j)}$ 的观察值落入区间 $(y-\mathrm{d}y,y]$，当且仅当样本观察值中有 $k-1$ 个落入区间 $(-\infty,x-\mathrm{d}x]$，$j-1-k$ 个落入区间 $(x,y-\mathrm{d}y]$，$n-j$ 个落入区间 $(y,+\infty)$. 此时，

$$\begin{aligned}
&P\{x-\mathrm{d}x<X_{(k)}\leqslant x,y-\mathrm{d}y<X_{(j)}\leqslant y\}\\
&=\frac{n!}{(k-1)!1!(j-1-k)!1!(n-j)!}[F(x-\mathrm{d}x)]^{k-1}\\
&\quad\cdot[f(x)\mathrm{d}x+o(\mathrm{d}x)][F(y-\mathrm{d}y)-F(x)]^{j-1-k}\\
&\quad\cdot[f(y)\mathrm{d}y+o(\mathrm{d}y)][1-F(y)]^{n-j}\\
&=\frac{n!}{(k-1)!(j-1-k)!(n-j)!}[F(x)]^{k-1}\\
&\quad\cdot[F(y)-F(x)]^{j-1-k}[1-F(y)]^{n-j}f(x)f(y)\mathrm{d}x\mathrm{d}y\\
&\quad+\mathrm{d}x\cdot o(\mathrm{d}y)+\mathrm{d}y\cdot o(\mathrm{d}x)+o(\mathrm{d}x)\cdot o(\mathrm{d}y)
\end{aligned}$$

由此可知，$X_{(k)}$ 与 $X_{(j)}$ 的联合概率密度为

$$\begin{aligned}
f_{(k)(j)}(x,y)&=\frac{n!}{(k-1)!(j-1-k)!(n-j)!}\\
&\quad\cdot[F(x)]^{k-1}[F(y)-F(x)]^{j-1-k}[1-F(y)]^{n-j}f(x)f(y)
\end{aligned} \tag{1.1.9}$$

上述等式对一切 $x<y$ 都成立，而在其他场合 $f_{(k)(j)}(x,y)=0$.

特别，$X_{(1)}$ 与 $X_{(n)}$ 的联合概率密度为

$$f_{(1)(n)}(x,y)=\begin{cases}n(n-1)[F(y)-F(x)]^{n-2}f(x)f(y), & x<y\\ 0, & x\geqslant y\end{cases}$$

(1.1.10)

用类似的方法还可求得前 r 个顺序统计量 $X_{(1)},\cdots,X_{(r)}$ ($r\leqslant n$) 的联合概率密度

$$f_{(1)\cdots(r)}(x_1,\cdots,x_r)=\begin{cases}\dfrac{n!}{(n-r)!}[1-F(x_r)]^{n-r}f(x_1)\cdots f(x_r), & x_1<\cdots<x_r\\ 0, & \text{其他}\end{cases}$$

(1.1.11)

特别,顺序统计量 $(X_{(1)},\cdots,X_{(n)})^{\mathrm{T}}$ 的联合概率密度为

$$f_{(1)\cdots(n)}(x_1,\cdots,x_n)=\begin{cases}n!f(x_1)\cdots f(x_n), & x_1<x_2<\cdots<x_n\\ 0, & \text{其他}\end{cases}$$

(1.1.12)

4. 样本中位数与样本极差

设 $(X_1,\cdots,X_n)^{\mathrm{T}}$ 为总体 X 的一个样本,其顺序统计量为 $(X_{(1)},\cdots,X_{(n)})^{\mathrm{T}}$,由 $(X_{(1)},\cdots,X_{(n)})^{\mathrm{T}}$ 可定义在应用上有重要意义的样本中位数与样本极差.

称统计量

$$Me=\begin{cases}X_{((n+1)/2)}, & n\text{ 为奇数}\\ \dfrac{1}{2}(X_{(n/2)}+X_{(n/2+1)}), & n\text{ 为偶数}\end{cases}$$

(1.1.13)

为**样本中位数**(Median). 其观察值记为

$$me=\begin{cases}x_{((n+1)/2)}, & n\text{ 为奇数}\\ \dfrac{1}{2}(x_{(n/2)}+x_{(n/2+1)}), & n\text{ 为偶数}\end{cases}$$

样本中位数是反映样本值位置特征的一个量,它可以用于推断总体分布的中位数及总体的对称中心. 样本中位数具有计算方便且不受样本值中的异常值(outlier)影响的特点,因而,有时它比

样本均值更具有代表性.

称统计量

$$R = X_{(n)} - X_{(1)} \tag{1.1.14}$$

为**样本极差**(Range). 其观察值记为

$$r = x_{(n)} - x_{(1)}$$

样本极差是反映样本值分散程度的量,在某些场合,它可以用于推断总体的标准差.

5. 经验分布函数

定义 1.1.4　设$(X_1,\cdots,X_n)^{\mathrm{T}}$为总体 X 的一个样本,$(X_{(1)},\cdots,X_{(n)})^{\mathrm{T}}$为样本的顺序统计量. 当样本的观察值为$(x_1,\cdots,x_n)^{\mathrm{T}}$时,顺序统计量的观察值为$(x_{(1)},\cdots,x_{(n)})^{\mathrm{T}}$,对任意实数 x,记

$$F_n(x) = \begin{cases} 0, & x < x_{(1)} \\ \dfrac{k}{n}, & x_{(k)} \leqslant x < x_{(k+1)}, k = 1,2,\cdots,n-1 \\ 1, & x_{(n)} \leqslant x \end{cases} \tag{1.1.15}$$

则称 $F_n(x)$ 是**经验分布函数**.

若记

$$u(x) = \begin{cases} 1, & x \geqslant 0 \\ 0, & x < 0 \end{cases}$$

则(1.1.15)式可改写为

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} u(x - x_i) \tag{1.1.16}$$

其中$\sum_{i=1}^{n} u(x-x_i)$表示 n 个样品 $x_1,\cdots,x_n$ 中不超过 x 的个数.

对于经验分布函数 $F_n(x)$,我们可从两方面来分析.

一方面,对样本$(X_1,\cdots,X_n)^{\mathrm{T}}$的任一观察值$(x_1,\cdots,x_n)^{\mathrm{T}}$,$F_n(x)$是 x 的实函数,并且容易验证它具有如下性质:

(1) $F_n(x)$ 是 x 的单调非降函数;