研究生公共数学系列教材

数理统计

◎ 孙祝岭　徐晓岭　编著

Mathematical Statistics

高等教育出版社

研究生公共数学系列教材

数理统计

孙祝岭　徐晓岭　编著

高等教育出版社

内容简介

本书系统地讲述了统计推断的基本理论和方法，包括参数估计和假设检验；详细地介绍了数理统计应用较广的方法：回归分析、方差分析和正交试验设计；简明扼要地介绍了数理统计三个分支——可靠性统计、多元统计分析和时间序列分析的众多有实用价值的方法；简单地介绍了著名统计软件 SAS 的使用方法。

本书是为非数学类专业研究生编写的教材，经适当取舍后也可作为非数理统计专业的其他本科专业，如数学、统计学等专业的教材，也可供大学师生、科研工作者、工程技术人员和需要进行数据处理的人员参考。

图书在版编目(CIP)数据

数理统计/孙祝岭，徐晓岭编著. —北京：高等教育出版社，2009. 1
ISBN 978 -7 -04 -024941 -5

Ⅰ. 数… Ⅱ. ①孙…②徐… Ⅲ. 数理统计 - 研究生 - 教材 Ⅳ. O212

中国版本图书馆 CIP 数据核字(2008)第 193768 号

策划编辑 张长虹 **责任编辑** 张耀明 **封面设计** 赵 阳
责任绘图 杜晓丹 **版式设计** 王 莹 **责任校对** 金 辉
责任印制 朱学忠

出版发行	高等教育出版社	购书热线	010 - 58581118
社　　址	北京市西城区德外大街 4 号	免费咨询	800 - 810 - 0598
邮政编码	100120	网　　址	http://www.hep.edu.cn
总　　机	010 - 58581000		http://www.hep.com.cn
		网上订购	http://www.landraco.com
经　　销	蓝色畅想图书发行有限公司		http://www.landraco.com.cn
印　　刷	北京明月印务有限责任公司	畅想教育	http://www.widedu.com
开　　本	787 × 960 1/16	版　　次	2009 年 1 月第 1 版
印　　张	21.5	印　　次	2009 年 1 月第 1 次印刷
字　　数	400 000	定　　价	33.50 元

本书如有缺页、倒页、脱页等质量问题，请到所购图书销售部门联系调换。

物料号 24941 - 00

前　言

数理统计是关于数据资料的收集、整理、分析和推断的一门学科,简单地说,它是处理数据的有效工具。在科学研究中,用统计方法从数据中获得信息和发现初步规律往往成为重大科学发现的先导。数据是有说服力的,一个观点有了数据支持会增强其合理性、一个结论用数据说明会提高其可信度。数理统计应用广泛,在工农业生产、国防建设、科学研究、经济等领域中都可以找到它富有成果的应用。数理统计已有多个分支,内容已非常丰富了。

现在国际上研究生教育发展有一种新趋向,有的专业学生需要在一些辅助专业甚至在其他系科选课,不再是只跟一位导师做研究,研究工作有向跨学科、多专业发展的倾向,很多研究项目由于综合性强,所以凭一个学科的知识难以攻关成功;另外一些交叉学科的出现,使过去毫不相关的领域有了联系;这些都对学生的知识面提出了新的要求。时代的发展,科学的进步,促使研究生教育发生着变化。为适合研究生教育培养发展的新需要,本书注重深度和广度的结合,既系统地阐明了数理统计的基本理论和方法,又有详有略地介绍了不少有实用价值的数理统计方法,其中有的方法简要说明它解决问题的思想或介绍它能解决什么问题,要应用时可查有关专著。这样写的目的是希望在有限的篇幅内,提供更多有益的信息,实用价值高的统计方法介绍得多是本书的特色。我们希望这本书在迈向信息化时代的道路上成为读者的助手。

本书是为非数学类专业的研究生编写的教材,其中大部分内容也适合非数理统计专业的本科生选用,也可作为科学研究人员、工程技术专家的参考书。为使教学有更大的灵活性,在部分章节、一些证明和有一定难度的习题前加上*号,可酌情选用。本书力求做到易教易学,前五章内容配有适量习题,书后还有一套增加难度的附加习题可供学有余力或对做统计习题感兴趣的读者选做,所有习题书后都附有答案。另附有一份曾使用过的试卷(含解答)供参考。书后还有5份阅读材料,其中有的介绍解题方法,有的叙述知识点的联系,有的是前述内容的补充以弥补不足,有的介绍统计方法或关注热点,供读者选读。

由于我们水平有限,不当之处在所难免,恳请读者提出宝贵意见,我们将作进一步改进。

编　者

2008年3月于上海交通大学

目 录

第一章　统计推断准备

数理统计是关于数据资料的收集、整理、分析和推断的一门学科.数理统计的内容可以分为两部分:其一是如何合理有效地收集数据,数理统计的两个分支——试验设计和抽样调查研究的是这方面问题,本教材涉及这部分内容很少;其二是统计推断,应用考察对象的局部观测资料来推断考察对象的整体数量特征.当然,要了解整体的情况,最可靠的是采用全面调查的方法,但实际上,这往往是不必要或不可能的.比如:要大致了解一次全国性的有几十万人参加的外语考试的平均成绩.如果把每一考生的成绩相加再除以考生人数,那么可以计算得到真正的平均成绩,但由于考生很多,要得到结果计算量非常大,所以不必要这么做.实际上可以选部分考生的成绩就能对总平均成绩作出有效的估计.又比如:要检验某灯泡厂生产的灯泡的质量,检验的质量指标是使用寿命,但这种试验属于破坏性试验,测出一个灯泡的使用寿命后此灯泡就报废了,不可能对每个灯泡都做寿命试验.那么,怎样做才合理呢?数理统计的处理方法是:从该厂生产的所有灯泡中,抽取一小部分来进行试验,然后根据这一小部分灯泡的质量情况,来推断灯泡厂生产的灯泡的质量.

本章介绍统计推断的预备知识.

§1.1　一些基本概念

一、总体

定义 1.1.1　研究对象的全体组成的集合称为总体(或母体),而把组成总体的每个成员称为个体.

例 1.1.1　考察某大学学生的学习情况.

该大学的所有学生组成总体,即该大学学生总体,而其中每个学生是个体.

例 1.1.2　考察一批电灯泡的质量.

这批灯泡的全体组成总体,而其中每个灯泡是个体.

研究一个总体常常是研究总体的一项或多项指标,指标常常可以由个体的数量特征值计算得到.如对一个班级学生的平均体重感兴趣.该班全体学生组成总体,平均体重是一项指标.该班每个学生是个体,其体重是数量特征,班中每个学生的这一数量特征值之和除以班级学生人数就是平均体重这项指标.这时可

用个体的数量特征体重来表示总体.个体的数量特征是变量.故常记总体 X,或总体(X,Y),或总体$(X_1,X_2,\cdots,X_k)$等.

代表总体的变量可以看作是随机变量,当随机变量是正态变量时,此总体称为正态总体.以此类推有泊松总体,指数总体等.正态总体是最常见的总体.

二、样本

从总体中抽取个体的工作过程称为抽样.

定义 1.1.2 抽样得到的个体全体称为样本.样本中含有的个体数称为样本容量.

若研究总体是研究总体的一项指标,则可以用个体对应的数量特征值组成的数的集合来取代原来总体,这样便于统计处理.这时总体看作是一个数集.来自总体 X 的样本可记为 $X_1,X_2,\cdots,X_n$.类似的来自总体(X,Y)的样本可记为$(X_1,Y_1),(X_2,Y_2),\cdots,(X_n,Y_n)$等.

样本的取值称为样本观察值,记为 $x_1,x_2,\cdots,x_n$.

为解决问题的需要通常还要求样本满足下面两个条件:

1. 代表性好.抽样时应按等可能性原则选个体,即每个个体被选中是等可能的.这样组成样本的每一个变量有相同分布,其分布与总体分布相同.

2. 具有独立性.抽样时先后抽到的结果互不影响,即组成样本的变量间应相互独立.采用放回抽样时满足这一条件;采用不放回抽样时这一条件不满足,但总体规模很大时,可认为这一条件也近似满足.

满足上面两个条件的样本称为简单随机样本.即相互独立且服从相同分布的随机变量组成的样本.

注 后面不特别注明,提到的样本都是指简单随机样本.

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

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

若总体 X 为连续型随机变量,其密度函数为 $f(x)$,则样本 $X_1,X_2,\cdots,X_n$ 的联合密度函数为

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

若总体 X 为离散型随机变量,则样本 $X_1,X_2,\cdots,X_n$ 的联合分布列为

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

三、统计量

研究一个总体首先要获得关于这个总体的信息,通过抽样得到样本,样本中含有总体的有用信息,接下去的工作是对样本进行加工,把感兴趣的信息提炼出来,构造统计量的作用就是如此.

定义 1.1.3 样本的不含有未知参数的函数称为统计量.

设样本为 $X_1,X_2,\cdots,X_n$,统计量为 $g(X_1,X_2,\cdots,X_n)$.

当样本观察值取 $x_1,x_2,\cdots,x_n$ 时,函数值 $g(x_1,x_2,\cdots,x_n)$ 称为统计量的观察值.

在实际问题中,有时对总体的信息有所了解,这些信息可能使我们了解总体分布中的部分参数.因此在后面也经常讨论总体分布中的部分参数已知,部分参数未知的情形.

例 1.1.3 设总体 $X\sim N(\mu,\sigma^2)$,其中 μ 已知,但 σ^2 未知,来自总体 X 的样本为 $X_1,X_2,\cdots,X_n$,则

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

都是统计量,而 $\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})^2$ 不是统计量.

统计量是统计推断的基础.下面介绍两类常用统计量.

常用统计量之一:矩统计量

设样本为 $X_1,X_2,\cdots,X_n$,称 $\frac{1}{n}\sum_{i=1}^{n}X_i^k$ 为样本 k 阶原点矩,其中一阶原点矩称为样本均值,记为 $\bar{X}$,即 $\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$. 称 $\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^k$ 为样本 k 阶中心矩,特别 $k=2$ 时记为 S_n^2,即 $S_n^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$. 记 $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$,称 S^2 为样本方差;称 S 为样本标准差.

容易得 $S_n^2=\frac{1}{n}\sum_{i=1}^{n}X_i^2-\bar{X}^2\xlongequal{\text{记为}}\overline{X^2}-\bar{X}^2$,而 $S^2=\frac{n}{n-1}S_n^2$.

常用统计量之二:顺序统计量(或次序统计量)

设来自总体 X 的样本为 $X_1,X_2,\cdots,X_n$,其任意的观测值为 $x_1,x_2,\cdots,x_n$,把它们从小到大排列得 $x_{(1)}\leqslant x_{(2)}\leqslant\cdots\leqslant x_{(n)}$.

定义样本 $X_1,X_2,\cdots,X_n$ 的函数 $X_{(i)}$,当自变量取值为 $x_1,x_2,\cdots,x_n$,函数值取为 $x_{(i)}$,$i=1,2,\cdots,n$,称 $X_{(i)}$,$i=1,2,\cdots,n$ 为顺序统计量.分别称 $X_{(1)}$、$X_{(n)}$ 为最小、最大顺序统计量,$X_{(1)}$、$X_{(n)}$ 简单表达式为

$$X_{(1)} = \min(X_1, X_2, \cdots, X_n),$$
$$X_{(n)} = \max(X_1, X_2, \cdots, X_n).$$

称 $R = X_{(n)} - X_{(1)}$ 为样本极差，它反映样本取值的波动范围.

称

$$Me = \begin{cases} \dfrac{X_{(k)} + X_{(k+1)}}{2}, & n = 2k, \\ X_{(k+1)}, & n = 2k+1 \end{cases}$$

为样本中位数，它反映样本中间大小的取值，不受极端值的影响.

§1.2 抽样分布

统计推断基于统计量，掌握统计量的分布是有用的，有时甚至是必须的.

定义 1.2.1 统计量的分布称为抽样分布.

在常用的抽样分布中除正态分布外还有三种重要的分布：χ^2 分布，t 分布和 F 分布. 下面分别加以回顾和介绍.

一、正态分布

设 $X \sim N(\mu, \sigma^2)$，正态变量 X 的密度函数为

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

其中 μ, σ 为常数，$-\infty < \mu < +\infty$，$\sigma > 0$.

特例 当 $\mu = 0, \sigma = 1$ 时，$N(0,1)$ 为标准正态分布.

标准正态分布的密度函数和分布函数分别记为 $\varphi(x)$，$\Phi(x)$，

$$\Phi(x) = \int_{-\infty}^{x} \varphi(t)\,\mathrm{d}t.$$

性质 1 $\Phi(-x) = 1 - \Phi(x)$.

本书后面附有标准正态分布函数值表（见附表一）.

设 $X \sim N(0,1)$，有计算公式

$$P(a < X \leqslant b) = \Phi(b) - \Phi(a).$$

设 $X \sim N(\mu, \sigma^2)$，有计算公式

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

性质 2 设 $X \sim N(\mu, \sigma^2)$，则 $EX = \mu, DX = \sigma^2$.

正态变量有很多好的性质，如正态向量的线性变换仍是正态向量；正态向量的分量独立和不相关等价；正态向量的分量的线性组合仍是正态变量；独立正态

变量的线性组合仍是正态变量等.

二、χ^2 分布

定义 1.2.2 设 $X_1, X_2, \cdots, X_n$ 相互独立且同服从 $N(0,1)$，则称 $\chi^2 = X_1^2 + X_2^2 + \cdots + X_n^2$的分布为自由度是 n 的χ^2 分布，记为$\chi^2 \sim \chi^2(n)$. 其密度函数为

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

其中伽玛函数 $\Gamma(t)$ 是一种特殊函数，定义为

$$\Gamma(t) = \int_0^{+\infty} x^{t-1} \mathrm{e}^{x}\, \mathrm{d}x, t > 0,$$

$\Gamma(1) = 1, \Gamma\left(\dfrac{1}{2}\right) = \sqrt{\pi}$是常用函数值.

函数具有性质：$\Gamma(t+1) = t\Gamma(t)$（递推公式）. 例如，

$$\Gamma\left(\frac{5}{2}\right) = \Gamma\left(\frac{3}{2}+1\right) = \frac{3}{2}\Gamma\left(\frac{3}{2}\right) = \frac{3}{2}\Gamma\left(\frac{1}{2}+1\right) = \frac{3}{2}\times\frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{3}{4}\sqrt{\pi}.$$

当 n 为自然数时，$\Gamma(n) = (n-1)!$.

当 $t \in (0,1)$ 时，可查表确定其函数值（一些数学手册上有表可查）；

当 $t \in (1, +\infty)$ 时，可应用递推公式结合查表确定其函数值.

χ^2 分布密度函数图像如图 1.1：

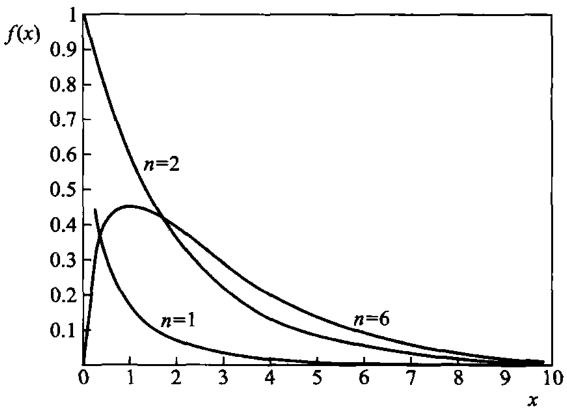

图 1.1 χ^2 分布密度曲线

性质

(1) 设 $X \sim \chi^2(n)$，则 $EX = n, DX = 2n$.

(2) 设 $X \sim \chi^2(n)$，$Y \sim \chi^2(m)$，且 X 与 Y 相互独立，则 $X + Y \sim \chi^2(n+m)$.

注 性质(1)表明χ^2 变量的数学期望是分布的自由度，而方差是分布的自

由度的 2 倍.

性质(1)的证明方法一:记

$$EX = \int_0^{+\infty} x \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} \mathrm{e}^{-x/2} \mathrm{d}x = \frac{1}{2^{n/2}\Gamma(n/2)} \int_0^{+\infty} x^{n/2} \mathrm{e}^{-x/2} \mathrm{d}x$$

$$= \frac{2^{(n+2)/2}\Gamma((n+2)/2)}{2^{n/2}\Gamma(n/2)} \int_0^{+\infty} \frac{1}{2^{(n+2)/2}\Gamma((n+2)/2)} x^{(n+2)/2-1} \mathrm{e}^{-x/2} \mathrm{d}x,$$

由密度函数的性质 $\int_{-\infty}^{+\infty} f(x)\mathrm{d}x = 1$, 得

$$上式 = \frac{2^{(n+2)/2}\Gamma((n+2)/2)}{2^{n/2}\Gamma(n/2)} \times 1 = 2 \times \frac{n}{2} = n.$$

用类似方法可证明 $EX^2 = n^2 + 2n$,从而 $DX = EX^2 - (EX)^2 = 2n$.

性质(1)的证明方法二(应用 Γ 函数计算):

$$EX = \int_0^{+\infty} x \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} \mathrm{e}^{-x/2} \mathrm{d}x = \frac{1}{2^{n/2}\Gamma(n/2)} \int_0^{+\infty} x^{n/2} \mathrm{e}^{-x/2} \mathrm{d}x$$

$$\overset{令 x/2=t}{=} \frac{1}{2^{n/2}\Gamma(n/2)} \int_0^{+\infty} (2t)^{n/2} \mathrm{e}^{-t} \times 2\mathrm{d}t$$

$$= \frac{2}{\Gamma(n/2)} \int_0^{+\infty} t^{(n+2)/2-1} \mathrm{e}^{-t} \mathrm{d}t = \frac{2}{\Gamma(n/2)} \times \Gamma(n+2/2)$$

$$= \frac{2}{\Gamma(n/2)} \times \frac{n}{2} \times \Gamma(n/2) = n.$$

性质(2)表明独立 χ^2 变量具有可加性,这一性质可以应用求和的分布的卷积公式推导得到,证明略.

下面不加证明的给出一个后面要用到的结果.

定理 1.2.1(柯赫伦定理) 设 $X_1, X_2, \cdots, X_n$ 相互独立且同服从 $N(0,1)$,

$$\sum_{i=1}^{n} X_i^2 = Q_1 + Q_2 + \cdots + Q_k,$$

其中 $Q_i, i=1,2,\cdots,k$ 是 $X_1, X_2, \cdots, X_n$ 的秩为 n_i 的二次型. 则 $Q_i, i=1,2,\cdots,k$ 相互独立,且依次服从自由度为 n_i 的 χ^2 分布(即 $Q_i \sim \chi^2(n_i), i=1,2,\cdots,k$)的充要条件是

$$n_1 + n_2 + \cdots + n_k = n.$$

三、t 分布

定义 1.2.3 设 $X \sim N(0,1), Y \sim \chi^2(n)$ 且 X 与 Y 相互独立,则称 $T = \frac{X}{\sqrt{Y/n}}$ 的分布为自由度是 n 的 t 分布,记为 $T \sim t(n)$.

分布 $t(n)$ 的密度函数为

$$f(x)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}},\quad -\infty<x<+\infty,$$

其图像如图 1.2：

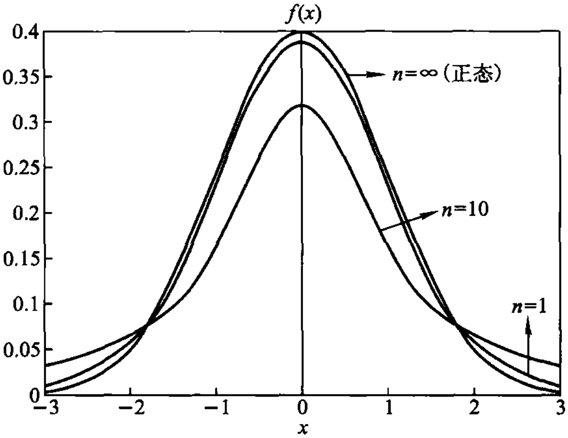

图 1.2 t 分布密度曲线

四、F 分布

定义 1.2.4 设 $X\sim\chi^2(n)$，$Y\sim\chi^2(m)$，且 X 与 Y 相互独立，则称 $F=\dfrac{X/m}{Y/m}$ 的分布为自由度是 (n,m) 的 F 分布，记为 $F\sim F(n,m)$.

分布 $F(n,m)$ 的密度函数为

$$f(x)=\begin{cases}\dfrac{\Gamma\left(\frac{n+m}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)}\left(\dfrac{n}{m}\right)^{\frac{n}{2}}x^{\frac{n}{2}-1}\left(1+\dfrac{nx}{m}\right)^{-\frac{n+m}{2}}, & x\geqslant 0,\\ 0, & x<0,\end{cases}$$

其图像如图 1.3：

性质 若 $F\sim F(n,m)$，则 $\dfrac{1}{F}\sim F(m,n)$.

应用 F 分布的定义即可得到结果.

下面给出在统计推断中几个常用的结果.

定理 1.2.2（Fisher 定理） 设 $X_1,X_2,\cdots,X_n$ 是来自正态总体 $N(\mu,\sigma^2)$ 的样本，则

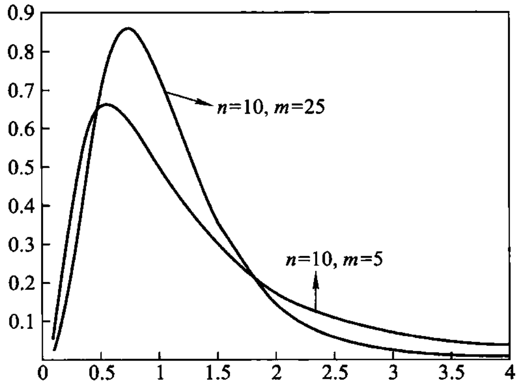

图 1.3　F 分布密度曲线

(1) $\bar{X} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$,

(2) $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$,

(3) $\bar{X}$ 与 S^2 相互独立.

*证明　设 $\boldsymbol{A} = (a_{ij})_{n\times n}$ 是一个 n 阶正交矩阵，它的第一行为 $\left(\dfrac{1}{\sqrt{n}}, \dfrac{1}{\sqrt{n}}, \cdots, \dfrac{1}{\sqrt{n}}\right)$，其他行元素只要满足正交矩阵的条件就行，则

$$\sum_{j=1}^{n} a_{ij} = 0, \quad \sum_{j=1}^{n} a_{ij}^2 = 1, \quad i = 2, \cdots, n$$

作正交变换：

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} \dfrac{1}{\sqrt{n}} & \dfrac{1}{\sqrt{n}} & \cdots & \dfrac{1}{\sqrt{n}} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix},$$

则有

$$Y_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i = \sqrt{n}\bar{X},$$

$$(Y_1, Y_2, \cdots, Y_n) \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = (X_1, X_2, \cdots, X_n)\boldsymbol{A}'\boldsymbol{A} \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} = (X_1, X_2, \cdots, X_n) \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix},$$

即

$$Y_1^2+Y_2^2+\cdots+Y_n^2=X_1^2+X_2^2+\cdots+X_n^2=\sum_{i=1}^{n}(X_i-\bar{X})^2+n\bar{X}^2,$$

所以

$$Y_2^2+Y_3^2+\cdots+Y_n^2=\sum_{i=1}^{n}(X_i-\bar{X})^2=(n-1)S^2,$$

下面再推导 Y_i 的分布，$i=1,2,\cdots,n$，由正态向量的线性变换仍是正态向量，得 $\begin{pmatrix}Y_1\\Y_2\\\vdots\\Y_n\end{pmatrix}$ 仍为正态向量.

设

$$\begin{pmatrix}Y_1\\Y_2\\\vdots\\Y_n\end{pmatrix}\sim N_n\left(\begin{pmatrix}\mu_1\\\mu_2\\\vdots\\\mu_n\end{pmatrix},\boldsymbol{\Sigma}\right),$$

其中 $\boldsymbol{\Sigma}=(\sigma_{ij})_{n\times n}$，$\mu_1=EY_1=E\sqrt{n}\bar{X}=\sqrt{n}\mu$，当 $i\geqslant 2$ 时，

$$\mu_i=EY_i=\sum_{j=1}^{n}a_{ij}EX_j=\Big(\sum_{j=1}^{n}a_{ij}\Big)\mu=0,$$

$$\begin{aligned}\sigma_{ij}&=\mathrm{COV}(Y_i,Y_j)=\mathrm{COV}\Big(\sum_{k=1}^{n}a_{ik}X_k,\sum_{l=1}^{n}a_{jl}X_l\Big)\\&=\sum_{k=1}^{n}\sum_{l=1}^{n}a_{ik}a_{jl}\mathrm{COV}(X_k,X_l)=\begin{cases}\sigma^2, i=j,\\0\ \ , i\neq j,\end{cases}\end{aligned}$$

所以 $Y_1,Y_2,\cdots,Y_n$ 相互独立，且由正态向量的各分量仍是正态变量，易得 $Y_1\sim N(\sqrt{n}\mu,\sigma^2)$，且 $Y_i\sim N(0,\sigma^2)$，$i=2,\cdots,n$，所以

$$\bar{X}=\frac{Y_1}{\sqrt{n}}\sim N\left(\mu,\frac{\sigma^2}{n}\right),\ \frac{(n-1)S^2}{\sigma^2}=\frac{\sum\limits_{i=2}^{n}Y_i^2}{\sigma^2}\sim\chi^2(n-1),$$

且 $\bar{X}$ 与 S^2 相互独立. 证毕

推论 条件同上定理，有

(1) $\dfrac{\bar{X}-\mu}{\sigma}\sqrt{n}\sim N(0,1)$；

(2) $\dfrac{\bar{X}-\mu}{S}\sqrt{n}\sim t(n-1)$.

证明 (1) 因为 $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$,所以标准化后,得$\frac{\bar{X}-\mu}{\sigma}\sqrt{n} \sim N(0,1)$.

(2) 因为

$$\frac{\bar{X}-\mu}{\sigma}\sqrt{n} \sim N(0,1), \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1),$$

且两变量相互独立.所以

$$\frac{\frac{\bar{X}-\mu}{\sigma}\sqrt{n}}{\sqrt{\frac{(n-1)S^2}{\sigma^2(n-1)}}} \sim t(n-1),$$

即

$$\frac{\bar{X}-\mu}{S}\sqrt{n} \sim t(n-1).$$ 证毕

定理 1.2.3 设 $X_1, X_2, \cdots, X_n$ 是来自 $N(\mu_1, \sigma_1^2)$, $Y_1, Y_2, \cdots, Y_m$ 是来自 $N(\mu_2, \sigma_2^2)$的样本,两样本相互独立,

记

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2,$$

$$\bar{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i, S_Y^2 = \frac{1}{m-1}\sum_{i=1}^{m}(Y_i - \bar{Y})^2,$$

则 (1)

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}} \sim N(0,1);$$

(2)

$$\frac{\sigma_2^2 S_X^2}{\sigma_1^2 S_Y^2} \sim F(n-1, m-1).$$

证明 (1) 因为 $\bar{X} \sim N\left(\mu_1, \frac{\sigma_1^2}{n}\right)$, $\bar{Y} \sim N\left(\mu_2, \frac{\sigma_2^2}{m}\right)$,且两变量相互独立.

所以

$$\bar{X}-\bar{Y} \sim N\left(\mu_1-\mu_2, \frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}\right),$$

所以

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}} \sim N(0,1).$$

（2）因为

$$\frac{(n-1)S_X^2}{\sigma_1^2} \sim \chi^2(n-1),\quad \frac{(m-1)S_Y^2}{\sigma_2^2} \sim \chi^2(m-1),$$

且两变量相互独立.

所以

$$\frac{\dfrac{(n-1)S_X^2}{\sigma_1^2(n-1)}}{\dfrac{(m-1)S_Y^2}{\sigma_2^2(m-1)}} = \frac{\sigma_2^2 S_X^2}{\sigma_1^2 S_Y^2} \sim F(n-1,m-1).$$

证毕

定理 1.2.4 设 $X_1,X_2,\cdots,X_n$ 是来自总体 $N(\mu_1,\sigma^2)$ 的样本，$Y_1,Y_2,\cdots,Y_n$ 是来自总体 $N(\mu_2,\sigma^2)$ 的样本，两样本相互独立，记

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i,\quad S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2,$$

$$\bar{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i,\quad S_Y^2 = \frac{1}{m-1}\sum_{i=1}^{m}(Y_i-\bar{Y})^2,$$

则

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_1)}{\sqrt{(n-1)S_X^2+(m-1)S_Y^2}}\sqrt{\frac{nm(n+m-2)}{n+m}} \sim t(n+m-2).$$

证明 因为 $\bar{X} \sim N\left(\mu_1,\dfrac{\sigma^2}{n}\right)$，$\bar{Y} \sim N\left(\mu_2,\dfrac{\sigma^2}{m}\right)$，且两变量相互独立，

所以

$$\bar{X}-\bar{Y} \sim N\left(\mu_1-\mu_2,\left(\frac{1}{n}+\frac{1}{m}\right)\sigma^2\right),$$

所以

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sigma\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}} \sim N(0,1),$$

因为

$$\frac{(n-1)S_X^2}{\sigma^2} \sim \chi^2(n-1),\ \frac{(m-1)S_Y^2}{\sigma^2} \sim \chi^2(m-1),$$

且两变量相互独立，所以

$$\frac{(n-1)S_X^2+(m-1)S_Y^2}{\sigma^2} \sim \chi^2(n+m-2),$$

且此变量与上面标准正态变量独立，所以