普通高等教育统计与大数据专业“十三五”规划教材

多元统计分析

陈钰芬　陈骥　主编

清华大学出版社

北　京

内 容 简 介

本书系统地介绍了多元统计分析技术的基本思想和方法原理，以社会、经济、商务等领域的实际问题为案例，结合SAS软件，介绍各种方法的SAS操作、实现过程与结果解释，帮助读者理解并掌握多元统计分析的基本方法，熟练应用软件进行数据分析，提高对实际数据的分析挖掘能力。

本书内容重点突出、习题设置合理、教学资源丰富，适合作为普通高等学校统计学专业或大数据相关专业本科生、经济管理类或社科类专业研究生的教材，也可作为从事社会、经济、管理等研究和实践工作的人士进行量化研究的参考书。

图书在版编目(CIP)数据

多元统计分析 / 陈钰芬，陈骥 主编. —北京：清华大学出版社，2020.7
普通高等教育统计与大数据专业“十三五”规划教材
ISBN 978-7-302-54668-9

Ⅰ. ①多… Ⅱ. ①陈… ②陈… Ⅲ. ①多元分析－统计分析－高等学校－教材 Ⅳ. ①O212.4

中国版本图书馆 CIP 数据核字(2019)第 294840 号

责任编辑： 崔　伟
封面设计： 马筱琨
版式设计： 思创景点
责任校对： 成凤进
责任印制：

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： 北京鑫丰华彩印有限公司
经　　销： 全国新华书店
开　　本： 185mm×260mm　　**印　　张：** 16.5　　**字　　数：** 402 千字
版　　次： 2020 年 8 月第 1 版　　**印　　次：** 2020 年 8 月第 1 次印刷
定　　价： 49.00 元

产品编号：083806-01

前　言

随着大数据、云计算和人工智能时代的来临，多元统计分析这门集数学、统计学和计算机科学为一体的数据科学在全世界范围内迅速兴起。多元统计分析方法是处理多变量数据不可缺少的重要技术和方法，是大数据分析的重要工具。

多元统计分析是以概率统计为基础，应用线性代数的基本原理和方法，结合计算机对实际资料和信息进行分析挖掘的一种统计分析技术。它的应用性极强，在自然科学、社会科学、经济管理等各领域得到了越来越广泛的应用。

本书是浙江省一流学科、浙江省优势特色学科、浙江省一流专业的建设成果之一，由作者结合二十多年的教学和科研工作经验编写而成，着重突出以下特点。

(1) 注重统计基本思想。本书以深入浅出的方式简要阐述了多元统计分析的基本思想，有助于学生深刻理解多元分析的基本思想方法。

(2) 阐明统计基本原理。多元统计方法的数学原理较为抽象，为便于学生阅读并较好地理解，本书对各种方法的基本原理进行了详细的推导，在不失严谨的前提下，略过了一些复杂程度高但又不影响方法原理理解的数学推导，读者只需掌握初步的微积分、线性代数和概率统计知识，便能理解。

(3) 突出实际案例应用。本着深入浅出的宗旨，在系统介绍多元分析基本理论和方法的同时，结合社会、经济、商务等领域的研究实例，把多元分析的方法与实际应用结合起来，努力把我们在实践中应用多元分析的经验和体会融入其中。

(4) 结合 SAS 软件实现。多元统计分析的应用离不开计算机，本书案例主要运用 SAS 软件实现，在每种方法后结合实例介绍 SAS 软件的实现过程与结果解释。所有案例数据都是能获取的真实数据，这有利于将 SAS 软件更好地融入各章的内容中，使读者能深切地体会多元统计分析的意义，便于读者进入应用领域。

(5) 习题设置合理。为使读者掌握本书内容，又考虑到这门课程的应用性和实践性，每章都给出一些思考与练习题，这些习题安排侧重对基本概念的理解和知识点的实际应用，并不注重解题的数学技巧和难度。

(6) 教学资源丰富。为方便教学，本书提供教学课件、案例与习题数据以及习题答案。教师可扫描下页二维码，审核通过后，即可获取。除此之外，各章“SAS 实现与应用案例”部分还提供了微课视频，方便学生掌握 SAS 软件的操作方法，提高解决问题的能力。

本书旨在让学生理解并掌握多元统计分析的基本方法，熟练应用软件进行数据分析，适合作为统计学专业本科生和非统计学专业研究生的教材，也可作为大数据或其他专业学生学习多元统计分析的教材或教学参考书，还可作为从事社会、经济、管理等研究和实践的人士进行量化研究的参考书。

本书共分 10 章。第 1 章、第 2 章主要介绍一元统计推广到多元统计的内容，阐述了

多元正态分布的基本概念及其统计推断。第 3 章至第 10 章介绍了各种多元统计分析技术，这部分内容具有很强的实用性，特别是介绍了各种降维技术，将原始的多个指标化为少数几个综合指标，便于对数据进行分析挖掘。

本书由浙江工商大学陈钰芬教授和陈骥教授担任主编，具体编写分工为：李双博编写第 1～2 章，陈钰芬编写第 3～8 章，陈骥编写第 9～10 章。在本书的编写过程中，博士生陈思超、硕士生苏可和吴苏霞对数据处理和案例资料搜集进行了大量细致繁琐的工作。我们也参考和吸收了一些同类教材的成果，在此一并感谢！

由于编者水平有限，书中谬误之处在所难免，恳请读者批评指正。

教学资源

案例与习题数据

SAS 在线安装视频

陈钰芬

2020 年 5 月

目　录

第1章

多元正态分布

1.1 随机向量

在多元统计分析中，多元正态分布占有相当重要的地位。就理论而言，多元正态分布有相当优良的性质，因此多元统计分析的许多重要理论和方法或直接或间接地建立在正态分布的基础上，而围绕多元正态分布，已经建立了一套行之有效的统计推断方法。就实践而言，在实际中遇到的许多随机向量都服从或近似服从正态分布。

在研究许多实际问题时，往往会遇到多指标问题，即在一个问题中涉及多个随机变量。由于这些指标之间往往有某种联系，因此需要把这些指标作为一个总体来研究。多元统计分析研究的就是多指标的总体。

1.1.1 随机向量的定义

假定我们每次同时观测一个个体的 p 个指标，将这 p 个指标(即变量)放在一起得到一个 p 维随机向量 $\boldsymbol{X}=(X_1, X_2, \cdots, X_p)'$，表示同一次观测的 p 个变量，而由这 p 个需要观测的指标的个体所构成的总体，我们称为 p 元总体。每次观测得到一个样品，全体 n 个样品形成一个样本。

定义1.1 p 个随机变量 $X_1, X_2, \cdots, X_p$ 所组成的向量 $\boldsymbol{X}=(X_1, X_2, \cdots, X_p)'$ 称为随机向量。

注：如无特殊说明，本书中所称向量均指列向量。

假定我们一共进行了 n 次观测，得到的数据放在一起排成一个 $n\times p$ 矩阵，称为样本数据阵(或样本资料阵)，记为

$$\boldsymbol{X}=\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

横看矩阵的第 i 行，$\boldsymbol{X}'_{(i)}=(x_{i1}, x_{i2}, \cdots, x_{ip})\ (i=1, \cdots, n)$ 表示第 i 个样品的观测值。在具体观测之前，它是一个 p 维的随机向量。

竖看矩阵的第 j 列，

$$\boldsymbol{X}_j=\begin{bmatrix}x_{1j}\\x_{2j}\\\vdots\\x_{nj}\end{bmatrix},\ (j=1,2,\cdots,p)$$

表示对第 j 个变量的 n 次观测。在具体观测之前，它是一个 n 维的随机向量。

利用这样的记号，我们可以将样本数据阵表示为

$$\boldsymbol{X}=\begin{bmatrix}\boldsymbol{X}'_{(1)}\\\boldsymbol{X}'_{(2)}\\\vdots\\\boldsymbol{X}'_{(n)}\end{bmatrix}=(\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots,\boldsymbol{X}_p)$$

在观测之前，它是一个随机阵。而一旦观测值取定，$\boldsymbol{X}$ 就是一个数据矩阵。

多元统计分析中所涉及的很多方法都是充分运用各种手段从样本资料阵中提取信息，因此本书中需要运用随机向量或是多个随机向量构成的随机阵的一些性质。需要注意的是，本章中的多元样本是指简单随机样本，即不同样品的观测值之间是相互独立的，但是对多元样本中的每个样品而言，p 个指标的观测值之间往往是有相依关系的。不同样品的观测值之间有相依关系的一般属于多元时间序列分析研究的范畴。

1.1.2 随机向量的分布

随机向量可以由它的分布函数来完全描述。

定义1.2 设 $\boldsymbol{X}=(X_1,X_2,\cdots,X_p)'$ 为 p 维随机向量，其联合分布函数为

$$F(x_1,\cdots,x_p)=P(X_1\leqslant x_1,\cdots,X_p\leqslant x_p)$$

记为 $\boldsymbol{X}\sim F$。

定义1.3 如果存在非负函数 $f(x_1,\cdots,x_p)$，使得对一切 $(x_1,\cdots,x_p)\in\mathbf{R}^p$，联合分布函数均可表示为

$$F(x_1,\cdots,x_p)=\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_p}f(t_1,\cdots,t_p)\,\mathrm{d}t_1\cdots\mathrm{d}t_p$$

则称 $\boldsymbol{X}$ 为连续型随机向量，称 $f(x_1,\cdots,x_p)$ 为 $\boldsymbol{X}$ 的联合概率密度函数，简称为密度函数或者分布密度。

密度函数有以下两条重要性质：

(1) $\forall\,(x_1,\cdots,x_p)\in\mathbf{R}^p$，$f(x_1,\cdots,x_p)\geqslant 0$；

(2) $\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}f(t_1,\cdots,t_p)\,\mathrm{d}t_1\cdots\mathrm{d}t_p=1$。

事实上，一个 p 维变量的函数 $f(x_1,\cdots,x_p)$ 能作为p中某个随机向量的分布密度当且仅当以上两条性质成立时。

对于随机向量，有时我们关注的是部分分量的分布信息，因此还需要定义边缘分布。

定义1.4 设 $\boldsymbol{X}=(X_1,X_2,\cdots,X_p)'$ 为 p 维随机向量，其联合分布函数为 $F(x_1,\cdots,x_p)$。$\boldsymbol{X}$ 的 q 个分量所组成的子向量 $(X_{i_1},\cdots,X_{i_q})'$ 的分布称为 $\boldsymbol{X}$ 的边缘(或边际)分布。

如果我们将 $\boldsymbol{X}$ 划分为 q 维子向量$\boldsymbol{X}^{(1)}$与 $p-q$ 维子向量$\boldsymbol{X}^{(2)}$，那么$\boldsymbol{X}^{(1)}$的边缘分布为

$$\begin{aligned}F^{(1)}(x_1,\cdots,x_q)&=P(X_1\leqslant x_1,\cdots,X_q\leqslant x_q)\\&=P(X_1\leqslant x_1,\cdots,X_q\leqslant x_q,X_{q+1}\leqslant\infty,\cdots,X_p\leqslant\infty)\\&=F(x_1,\cdots,x_q,\infty,\cdots,\infty)\end{aligned}$$

当 $\boldsymbol{X}$ 有分布密度时，$\boldsymbol{X}^{(1)}$ 也有分布密度，其边缘密度为

$$f^{(1)}(x_1,\cdots,x_q)=\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}f(x_1,\cdots,x_p)\,\mathrm{d}t_{q+1}\cdots\mathrm{d}t_p$$

在概率论中，我们学习过随机变量的条件分布与独立性等相关概念，随机向量中也有类似概念。

定义1.5　如果我们将 $\boldsymbol{X}$ 划分为 q 维子向量$\boldsymbol{X}^{(1)}$ 与 $p-q$ 维子向量$\boldsymbol{X}^{(2)}$，那么在给定 $\boldsymbol{X}^{(2)}$ 时，$\boldsymbol{X}^{(1)}$ 的分布称为条件分布。如果 $\boldsymbol{X}$ 有密度函数 $f(\boldsymbol{x}^{(1)},\boldsymbol{x}^{(2)})$，那么给定$\boldsymbol{X}^{(2)}$ 时，$\boldsymbol{X}^{(1)}$ 的密度函数为

$$f_1(\boldsymbol{x}^{(1)}\mid\boldsymbol{x}^{(2)})=f(\boldsymbol{x}^{(1)},\boldsymbol{x}^{(2)})/f_2(\boldsymbol{x}^{(2)})$$

其中，$f_2(\boldsymbol{x}^{(2)})$ 是$\boldsymbol{X}^{(2)}$ 的边缘密度。

定义1.6　若 p 个随机向量$\boldsymbol{X}_1,\cdots,\boldsymbol{X}_p$的联合分布等于各自边缘分布的乘积，则称$\boldsymbol{X}_1,\cdots,\boldsymbol{X}_p$是相互独立的。需要注意的是，如果$\boldsymbol{X}_1,\cdots,\boldsymbol{X}_p$相互独立，那么其中任意两个随机向量两两独立，但是反之不真。

1.1.3　随机向量的数字特征

设 $\boldsymbol{X}=(X_1,X_2,\cdots,X_p)'$，$\boldsymbol{Y}=(Y_1,Y_2,\cdots,Y_q)'$为两个随机向量。

若 $E(X_i)=\mu_i$存在，则称

$$E(\boldsymbol{X})=\begin{bmatrix}E(X_1)\\\vdots\\E(X_p)\end{bmatrix}=\begin{bmatrix}\mu_1\\\vdots\\\mu_p\end{bmatrix}$$

为随机向量 $\boldsymbol{X}$ 的均值向量。

根据定义容易验证均值向量具有以下性质：

$$E(\boldsymbol{AX})=\boldsymbol{A}E(\boldsymbol{X})$$

$$E(\boldsymbol{AXB})=\boldsymbol{A}E(\boldsymbol{X})\boldsymbol{B}$$

其中 $\boldsymbol{A}$，$\boldsymbol{B}$ 为大小适合矩阵运算的常数矩阵。

若X_i与X_j的协方差存在$(i,j=1,\cdots,p)$，则称

$$\begin{aligned}D(\boldsymbol{X})&=E\left[\left(\boldsymbol{X}-E(\boldsymbol{X})\right)\left(\boldsymbol{X}-E(\boldsymbol{X})\right)'\right]\\&=\begin{bmatrix}D(X_1)&\operatorname{cov}(X_1,X_2)&\cdots&\operatorname{cov}(X_1,X_p)\\\operatorname{cov}(X_2,X_1)&D(X_2)&\cdots&\operatorname{cov}(X_2,X_p)\\\vdots&\vdots&\vdots&\vdots\\\operatorname{cov}(X_p,X_1)&\operatorname{cov}(X_p,X_2)&\cdots&D(X_p)\end{bmatrix}\end{aligned}$$

为随机向量 $\boldsymbol{X}$ 的协方差阵。

若X_i与Y_j的协方差存在$(i=1,\cdots,p;\ j=1,\cdots,q)$，则称

$$\operatorname{cov}(\boldsymbol{X},\boldsymbol{Y})=E\left[\left(\boldsymbol{X}-E(\boldsymbol{X})\right)\left(\boldsymbol{Y}-E(\boldsymbol{Y})\right)'\right]$$

$$=\begin{bmatrix}\operatorname{cov}(X_1,Y_1) & \operatorname{cov}(X_1,Y_2) & \cdots & \operatorname{cov}(X_1,Y_q)\\ \operatorname{cov}(X_2,Y_1) & \operatorname{cov}(X_2,Y_2) & \cdots & \operatorname{cov}(X_2,Y_q)\\ \vdots & \vdots & \vdots & \vdots\\ \operatorname{cov}(X_p,Y_1) & \operatorname{cov}(X_p,Y_2) & \cdots & \operatorname{cov}(X_p,Y_q)\end{bmatrix}$$

为随机向量 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 的协方差阵。当 $\boldsymbol{X}=\boldsymbol{Y}$ 时，$\operatorname{cov}(\boldsymbol{X},\boldsymbol{Y})$ 即为 $D(\boldsymbol{X})$。当 $\operatorname{cov}(\boldsymbol{X},\boldsymbol{Y})=0$ 时，称 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 不相关。如果 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 独立，则 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 不相关。反之不真。

若 X_i 与 X_j 的协方差存在 $(i,j=1,\cdots,p)$，则可以计算 X_i 与 X_j 的相关系数

$$r_{ij}=\frac{\operatorname{cov}(X_i,X_j)}{\sqrt{D(X_i)D(X_j)}}$$

将这 $p\times p$ 个相关系数排列成一个方阵 $\boldsymbol{R}=(r_{ij})_{p\times p}$，称为 $\boldsymbol{X}$ 的相关阵。

若记 X_i 的方差 $D(X_i)$ 为 σ_{ii}，则我们称 $\boldsymbol{V}^{1/2}=\operatorname{diag}\left(\sqrt{\sigma_{11}},\cdots,\sqrt{\sigma_{pp}}\right)$ 为标准差矩阵。协方差矩阵与相关阵有如下关系：

$$\boldsymbol{\Sigma}=\boldsymbol{V}^{1/2}\boldsymbol{R}\,\boldsymbol{V}^{1/2}\text{或}\ \boldsymbol{R}=(\boldsymbol{V}^{1/2})^{-1}\boldsymbol{\Sigma}\,(\boldsymbol{V}^{1/2})^{-1}。$$

根据协方差阵的定义，可以验证其具有以下性质：

(1) 随机向量 $\boldsymbol{X}$ 的协方差阵是对称非负定矩阵；

(2) $\operatorname{cov}(\boldsymbol{AX},\boldsymbol{BY})=\boldsymbol{A}\operatorname{cov}(\boldsymbol{X},\boldsymbol{Y})\boldsymbol{B}$，其中 $\boldsymbol{A}$，$\boldsymbol{B}$ 为大小适合矩阵运算的常数矩阵。

1.2 多元正态分布概述

1.2.1 多元正态分布的定义

在一元统计中，我们知道若 $X\sim N(0,1)$，则 X 的任意线性变换为 $Y=\sigma X+\boldsymbol{\mu}\sim N(\boldsymbol{\mu},\sigma^2)$。利用这一性质，我们可以由标准正态分布来定义一般正态分布。事实上，我们将这种方式推广到多元情况，可以得到多元正态分布的一种定义。

设 $X_1,\cdots,X_m$ 为 m 个相互独立标准正态变量，$\boldsymbol{X}=(X_1,\cdots,X_m)$ 为这 m 个随机变量构成的随机向量；设 $\boldsymbol{\mu}$ 为 p 维常数向量，$\boldsymbol{A}$ 为 $p\times m$ 维常数矩阵，则称 $\boldsymbol{Y}=\boldsymbol{AX}+\boldsymbol{\mu}$ 的分布为 p 元正态分布，或称 $\boldsymbol{Y}$ 为 p 维正态随机向量，记为 $\boldsymbol{Y}\sim N_p(\boldsymbol{\mu},\boldsymbol{A}\boldsymbol{A}')$。

大家知道一元正态随机变量的密度函数为

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{\frac{-(x-\boldsymbol{\mu})^2}{2\sigma^2}}\quad(\sigma>0,\ -\infty<x<\infty)\tag{1.1}$$

我们可以将式(1.1)改写为

$$f(x)=\frac{1}{(2\pi)^{1/2}\left|\sigma^2\right|^{1/2}}\exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'(\sigma^2)^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]$$

类似一元情况，我们给出 p 维正态分布的密度函数。

设 $\boldsymbol{X}\sim N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$，且 $\boldsymbol{\Sigma}$ 正定(为了保证 $\boldsymbol{\Sigma}^{-1}$ 存在)，那么 $\boldsymbol{X}$ 的联合密度函数为

$$f(\boldsymbol{x})=\frac{1}{(2\pi)^{p/2}\,|\boldsymbol{\Sigma}|^{1/2}}\exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'(\boldsymbol{\Sigma})^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]$$

例 1.1 设 $\boldsymbol{X}=\begin{bmatrix}X_1\\X_2\end{bmatrix}$ 服从二元正态分布，利用参数 $\mu_1=E(X_1)$，$\mu_2=E(X_2)$，$\sigma_1=\sqrt{D(X_1)}$，$\sigma_2=\sqrt{D(X_2)}$，$\rho=\dfrac{\operatorname{cov}(X_1,X_2)}{\sigma_1\sigma_2}$ 来表示 $\boldsymbol{X}$ 的联合密度。

解： 我们可以将协方差矩阵写作

$$\boldsymbol{\Sigma}=\begin{bmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{bmatrix}$$

从而其行列式为

$$|\boldsymbol{\Sigma}|=\sigma_1^2\sigma_2^2(1-\rho^2)$$

其逆矩阵为

$$\boldsymbol{\Sigma}^{-1}=\frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)}\begin{bmatrix}\sigma_2^2 & -\rho\sigma_1\sigma_2\\ -\rho\sigma_1\sigma_2 & \sigma_1^2\end{bmatrix}$$

将其带入密度公式中，可以得到 $\boldsymbol{X}$ 的联合密度为

$$f(x_1,x_2)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2-2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)+\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right\}$$

从密度函数的表达式可以看出，此密度函数的最高点坐标是 (μ_1,μ_2)。如果用与 xy 平面平行的平面去截二元正态密度函数曲面，所得截面为一个椭圆，称为概率密度等高线。

我们可以利用 SAS 系统绘制二维正态分布曲面的图形以及等高线图，具体如图 1-1～图 1-6 所示。

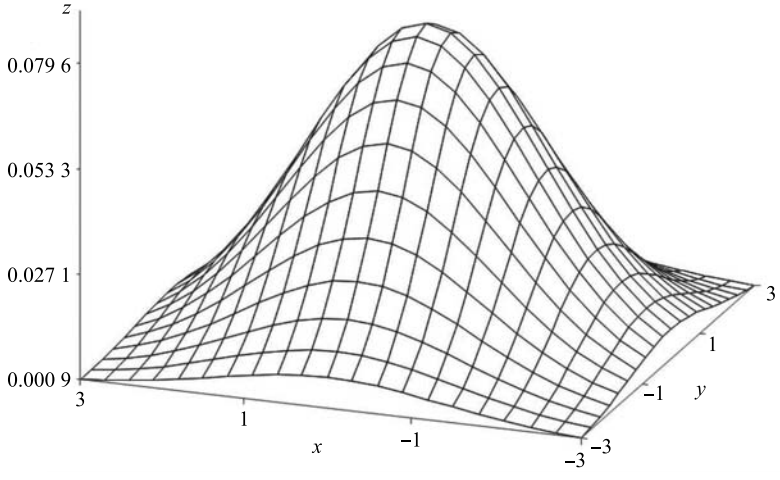

图 1-1　曲面图（$\sigma_{11}^2=\sigma_{22}^2=2$，$\rho_{12}=0$）

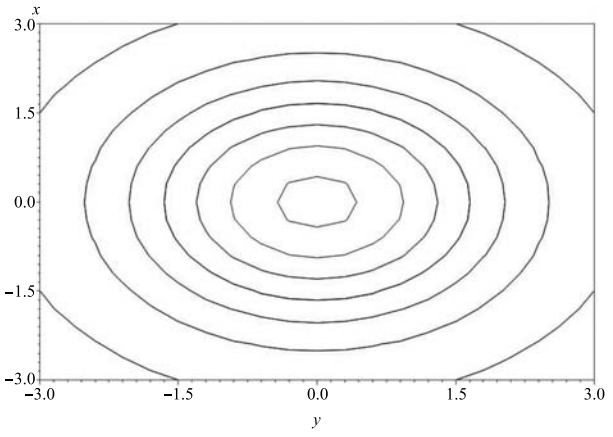

图 1-2　等高线图($\sigma_{11}^2=\sigma_{22}^2=2$, $\rho_{12}=0$)

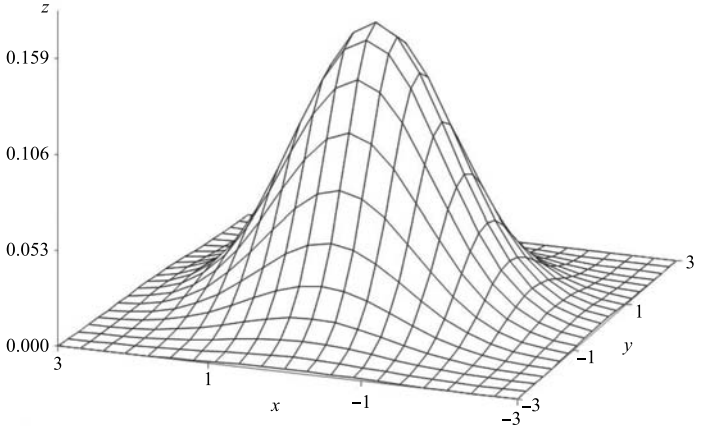

图 1-3　曲面图($\sigma_{11}^2=\sigma_{22}^2=1$, $\rho_{12}=0$)

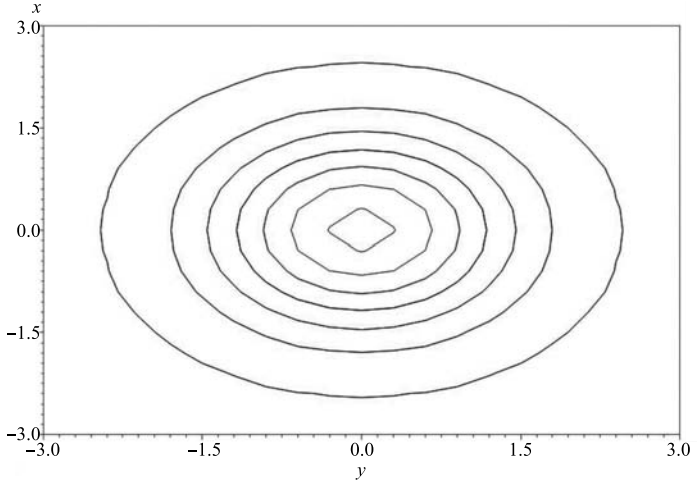

图 1-4　等高线图($\sigma_{11}^2=\sigma_{22}^2=1$, $\rho_{12}=0$)

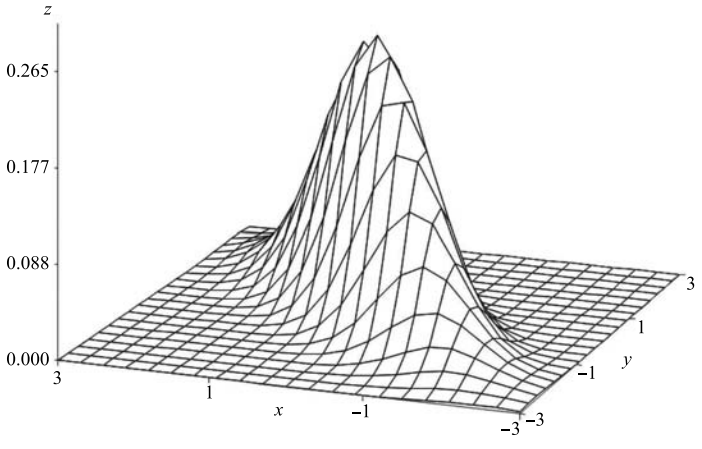

图 1-5　曲面图($\sigma_{11}^2=\sigma_{22}^2=1$，$\rho_{12}=0.8$)

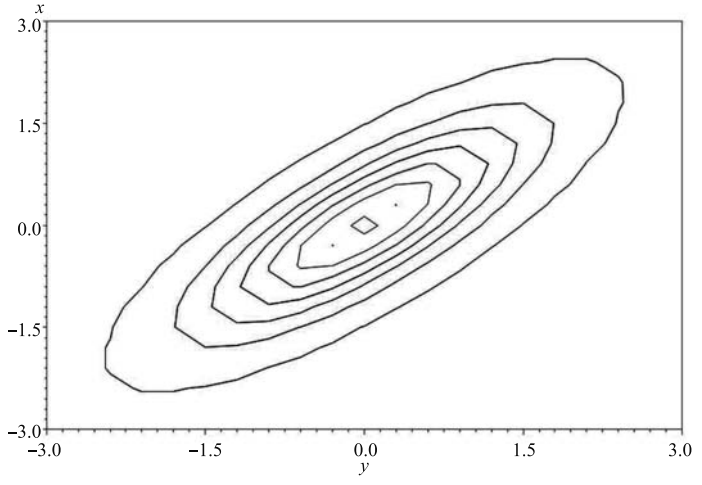

图 1-6　等高线图($\sigma_{11}^2=\sigma_{22}^2=1$，$\rho_{12}=0.8$)

图 1-1 与图 1-3 给出不同方差大小的正态图形，可以看出：方差较大时，密度函数曲面较为平缓；而方差较小时，(x_1, x_2)的取值更加集中在均值附近。这一点从图 1-2 和图 1-4 的相应等高线图也可以参照比对。图 1-5 给出当x_1与x_2有较强的相关性时，密度函数曲面较为陡立，而图 1-6 的等高线图可以看出强相关性的等高线比弱相关性的等高线的离心率要大。

1.2.2　多元正态变量的基本性质

多元正态分布在多元统计中占有十分重要的地位，许多重要理论与方法都建立在多元正态分布的性质之上。本节不加证明地给出多元正态变量的一些基本性质，以方便后文对正态分布及相关分布的处理。

设 $\boldsymbol{X}=(X_1, \cdots, X_p)' \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$。

(1) 若 $\boldsymbol{\Sigma}$ 为对角矩阵，则 $X_1, \cdots, X_p$ 相互独立。

(2) $\boldsymbol{X}$ 的任意边缘分布仍然为正态分布。特别的，如果将 $\boldsymbol{X}$，$\boldsymbol{\mu}$，$\boldsymbol{\Sigma}$ 作如下划分：

$$\boldsymbol{X}=\begin{bmatrix}\boldsymbol{X}^{(1)}\\ \boldsymbol{X}^{(2)}\end{bmatrix}\begin{matrix}q\\ p-q\end{matrix} \qquad \boldsymbol{\mu}=\begin{bmatrix}\boldsymbol{\mu}^{(1)}\\ \boldsymbol{\mu}^{(2)}\end{bmatrix} \qquad \boldsymbol{\Sigma}=\begin{bmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{bmatrix}$$

其中，$\boldsymbol{X}^{(1)}$ 与 $\boldsymbol{\mu}^{(1)}$ 为 p 维向量，$\boldsymbol{X}^{(2)}$ 与 $\boldsymbol{\mu}^{(2)}$ 为 $p-q$ 维向量，$\boldsymbol{\Sigma}_{11}$ 为 $q\times q$ 维矩阵，$\boldsymbol{\Sigma}_{12}$ 为 $q\times(p-q)$ 维矩阵，$\boldsymbol{\Sigma}_{21}$ 为 $(p-q)\times q$ 维矩阵，$\boldsymbol{\Sigma}_{22}$ 为 $(p-q)\times(p-q)$ 维矩阵，则 $\boldsymbol{X}^{(1)}\sim N_q(\boldsymbol{\mu}^{(1)}, \boldsymbol{\Sigma}_{11})$，$\boldsymbol{X}^{(2)}\sim N_{p-q}(\boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma}_{22})$。顺便指出，$\boldsymbol{X}^{(1)}$ 与 $\boldsymbol{X}^{(2)}$ 相互独立当且仅当 $\boldsymbol{\Sigma}_{12}$ 为零矩阵。

注意 如果一个随机向量的任意边缘分布都是正态分布，并不能推出它本身是多元正态分布。例如，考虑密度函数

$$f(x_1, x_2)=\frac{1}{2\pi}\mathrm{e}^{-\frac{1}{2}(x_1^2+x_2^2)}\left[1+x_1x_2\mathrm{e}^{-\frac{1}{2}(x_1^2+x_2^2)}\right]$$

所对应的随机向量 $(\boldsymbol{X}_1, \boldsymbol{X}_2)$。经过计算可以得出，$\boldsymbol{X}_1\sim N(0, 1)$，$\boldsymbol{X}_2\sim N(0, 1)$，但是它们的联合密度显然不是正态的。

(3) 设 $\boldsymbol{A}$ 是 $s\times p$ 阶常数矩阵，$\boldsymbol{d}$ 为 s 维常数向量，则 $\boldsymbol{AX}+\boldsymbol{d}$ 也服从正态分布，且 $\boldsymbol{AX}+\boldsymbol{d}\sim N_s(\boldsymbol{A\mu}+\boldsymbol{d}, \boldsymbol{A\Sigma A}')$。

(4) 若 $\boldsymbol{\Sigma}$ 为正定阵，则 $(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})\sim\chi^2(p)$。

1.3 多元正态分布的参数估计

在1.1.3节中我们给出了随机向量的数字特征。在实际应用中，均值向量和协方差矩阵等数字特征通常是未知的，需要利用样本来估计。本节考察多元正态总体的均值向量和协方差矩阵的估计，采用最常见的，也是具有较好性质的极大似然估计法给出其估计量，并给出极大似然估计法的性质。

1.3.1 多元样本的数字特征

考虑 p 元正态总体 $\boldsymbol{X}\sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$，设 $\boldsymbol{X}_{(1)}, \cdots, \boldsymbol{X}_{(n)}$ 为来自这个 p 元正态总体的简单随机样本，其中 $\boldsymbol{X}_{(i)}=(x_{i1}, \cdots, x_{ip})'(i=1, \cdots, p)$。

样本均值向量 $\bar{\boldsymbol{X}}$ 的定义为

$$\bar{\boldsymbol{X}}=\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_{(i)}=(\bar{x}_1, \cdots, \bar{x}_p)'=\frac{1}{n}\boldsymbol{X}'\boldsymbol{1}_n \tag{1.2}$$

在式(1.2)中，$\bar{x}_i=\frac{1}{n}\sum_{b=1}^{n}x_{bi}(i=1, \cdots, p)$，$\boldsymbol{1}_n$ 是一个 n 维的分量全为1的向量。

样本离差阵的定义为

$$\begin{aligned}\boldsymbol{A}&=\sum_{b=1}^{n}(\boldsymbol{X}_{(b)}-\bar{\boldsymbol{X}})(\boldsymbol{X}_{(b)}-\bar{\boldsymbol{X}})'\\ &=\boldsymbol{X}'\boldsymbol{X}-n\bar{\boldsymbol{X}}\bar{\boldsymbol{X}}'\end{aligned}$$

$$=\boldsymbol{X}'\left[\boldsymbol{I}_n-\frac{1}{n}\mathbf{1}_n\mathbf{1}_n'\right]\boldsymbol{X}$$

$$=(a_{ij})_{p\times p} \tag{1.3}$$

在式(1.3)中，$a_{ij}=\sum_{b=1}^{n}(x_{bi}-\bar{x}_i)(x_{bj}-\bar{x}_j)(i,j=1,\cdots,p)$。

样本协方差阵的定义为

$$\boldsymbol{S}=\frac{1}{n-1}\boldsymbol{A}=(s_{ij})_{p\times p}(\text{或者}\boldsymbol{S}^*=\frac{1}{n}\boldsymbol{A})$$

此时，$s_{ij}=\frac{1}{n-1}\sum_{b=1}^{n}(x_{bi}-\bar{x}_i)(x_{bj}-\bar{x}_j)(i,j=1,\cdots,p)$。

样本相关阵的定义为

$$\boldsymbol{R}=(r_{ij})_{p\times p} \tag{1.4}$$

在式(1.4)中，$r_{ij}=\frac{s_{ij}}{\sqrt{s_{ii}}\sqrt{s_{jj}}}=\frac{a_{ij}}{\sqrt{a_{ii}}\sqrt{a_{jj}}}$ $(i,j=1,\cdots,p)$。

1.3.2　均值向量和协方差矩阵的极大似然估计

设$\boldsymbol{X}_{(1)},\cdots,\boldsymbol{X}_{(n)}$ 为来自 p 元正态总体$N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$的简单随机样本，利用极大似然法可以求出 $\boldsymbol{\mu}$ 和$\boldsymbol{\Sigma}$ 的参数估计分别为$\hat{\boldsymbol{\mu}}=\bar{\boldsymbol{X}}$，$\hat{\boldsymbol{\Sigma}}=\boldsymbol{S}^*$。

$\hat{\boldsymbol{\mu}}$ 和$\hat{\boldsymbol{\Sigma}}$具有如下基本性质：

$E\bar{\boldsymbol{X}}=\boldsymbol{\mu}$，即$\bar{\boldsymbol{X}}$ 是$\boldsymbol{\mu}$ 的无偏估计。但是$E\boldsymbol{S}^*=\frac{n-1}{n}\boldsymbol{\Sigma}$，因此$\boldsymbol{\Sigma}$的极大似然估计不是无偏估计。在上文的定义中，我们将$\boldsymbol{S}=\frac{1}{n-1}\boldsymbol{A}$ 定义为样本协方差阵，就是因为$\boldsymbol{S}$ 是$\boldsymbol{\Sigma}$ 的无偏估计。

可以证明 $\bar{\boldsymbol{X}}$，$\boldsymbol{S}$ 是$\boldsymbol{\mu}$，$\boldsymbol{\Sigma}$ 的具有最小方差的无偏估计，也即 $\bar{\boldsymbol{X}}$，$\boldsymbol{S}$ 是$\boldsymbol{\mu}$，$\boldsymbol{\Sigma}$ 的有效估计。此外，$\bar{\boldsymbol{X}}$，$\boldsymbol{S}$ 还是$\boldsymbol{\mu}$，$\boldsymbol{\Sigma}$ 的相合估计以及充分统计量。关于如何利用极大似然法求得 $\boldsymbol{\mu}$ 和$\boldsymbol{\Sigma}$ 的参数估计以及$\bar{\boldsymbol{X}}$，$\boldsymbol{S}$ 的统计性质的证明，有兴趣的读者可以阅读相关文献了解[①]。

样本均值向量和样本离差阵在正态总体下还有一些重要性质。

定理1.1　设 $\bar{\boldsymbol{X}}$ 和$\boldsymbol{A}$ 分别为 p 元正态总体$N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$的样本均值向量和样本离差阵，则

(1) $\bar{\boldsymbol{X}}\sim N_p\left(\boldsymbol{\mu},\frac{1}{n}\boldsymbol{\Sigma}\right)$；

(2) 若设$\boldsymbol{Z}_1,\cdots,\boldsymbol{Z}_{n-1}$ 独立同$N_p(\boldsymbol{0},\boldsymbol{\Sigma})$ 分布，则 $\boldsymbol{A}$ 与$\sum_{t=1}^{n-1}\boldsymbol{Z}_t\boldsymbol{Z}_t'$同分布；

(3) $\bar{\boldsymbol{X}}$ 与$\boldsymbol{A}$ 相互独立；

(4) $\boldsymbol{A}$ 为正定阵的充要条件是$n>p$。

注意　这时 $\boldsymbol{A}$ 是随机矩阵，因此“$\boldsymbol{A}$ 为正定阵”这句话的含义实际上是“$\boldsymbol{A}$ 为正定阵”这个事件的概率为 1。

---

① 高惠璇. 应用多元统计分析[M]. 北京：北京大学出版社，2005.

1.4 常用分布与抽样分布

在数理统计中我们学习过，为了了解总体，我们对总体抽样得到样本，然后对样本进行加工，得到一个不包含未知量的样本的函数，这个样本函数我们一般称为统计量。在多元统计中也有类似的概念，比如我们前面介绍的样本均值向量 $\bar{\boldsymbol{X}}$ 和样本离差阵 $\boldsymbol{A}$ 等都是不含未知量的样本的函数，因此它们都是统计量。统计量的分布称为抽样分布。

在一元正态总体中，用于检验参数 μ，σ^2 的抽样分布有 χ^2 分布、t 分布以及 F 分布。这些抽样分布推广到多元正态总体中，与之对应的分布为 Wishart 分布、Hotelling T^2 分布以及 Wilks 分布。

1.4.1 Wishart 分布

如果从一元正态总体 $N(\mu, \sigma^2)$ 中抽取 n 个简单随机样本 $X_1, \cdots, X_n$，我们用样本方差

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

来估计 σ^2，此时 $\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X}) \sim \chi^2(n-1)$。因此，可以得到 $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$。那么对 p 元正态总体，样本协方差阵 $\boldsymbol{S} = \frac{1}{n-1}\boldsymbol{A}$ 又有怎样的分布呢？

这里先简要介绍一下如何定义随机矩阵的分布。设随机矩阵

$$\boldsymbol{X} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix}$$

将该矩阵的列向量（或者行向量）一个接一个地连接起来，组成一个长向量，这种操作一般称为将矩阵拉直为向量。这个拉直向量的分布就定义为随机矩阵 $\boldsymbol{X}$ 的分布。随机矩阵的分布还有其他不同的定义，本书中所指的随机矩阵的分布都是以拉直向量的分布来定义的。当 $\boldsymbol{X}$ 为对称阵时，只需要考虑下三角部分组成的长向量的分布，即 $(X_{11}, X_{21}, \cdots, X_{n1}, X_{22}, \cdots, X_{n2}, \cdots, X_{nn})$ 的分布。

定义1.7 设 $\boldsymbol{X}_{(b)} \sim N_p(\boldsymbol{\mu}_b, \boldsymbol{\Sigma})(b=1, \cdots, n)$ 是相互独立的 n 个 p 维正态变量，记 $\boldsymbol{X} = (\boldsymbol{X}_{(1)}, \cdots, \boldsymbol{X}_{(n)})'$ 为一个 $n \times p$ 矩阵，则称随机阵 $\boldsymbol{W} = \sum_{b=1}^{n}\boldsymbol{X}_{(b)}\boldsymbol{X}'_{(b)} = \boldsymbol{X}'\boldsymbol{X}$ 的分布为自由度为 n 的 p 维非中心 Wishart 分布，记为 $\boldsymbol{W} \sim W_p(n, \boldsymbol{\Sigma}, \Delta)$。其中，$\Delta$ 一般称为非中心参数，$\Delta = \sum_{b=1}^{n}\boldsymbol{\mu}_b\boldsymbol{\mu}'_b$。当 $\boldsymbol{\mu}_b = 0$ 时，我们一般称为中心 Wishart 分布，记为 $\boldsymbol{W} \sim W_p(n, \boldsymbol{\Sigma})$。

当 $p=1$，$\boldsymbol{\mu}_b = 0$ 时，$\boldsymbol{X}_{(b)} \sim N(0, \sigma^2)$，此时 $\boldsymbol{W} = W_1(n, \sigma^2) = \sum_{b=1}^{n}X^2_{(b)} \sim \sigma^2\chi^2(n)$。也

就是说，$W_1(n,1)$ 就是$\chi^2(n)$。因此，Wishart 分布是χ^2 分布在多元正态情形下的推广。

下面我们不加证明地给出 Wishart 分布的几条性质。

(1) 设$\boldsymbol{X}_{(b)}\sim N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})(b=1,\cdots,n)$相互独立，则样本离差阵 $\boldsymbol{A}$ 服从 Wishart 分布，即

$$\boldsymbol{A}=\sum_{b=1}^{n}(\boldsymbol{X}_{(b)}-\bar{\boldsymbol{X}})(\boldsymbol{X}_{(b)}-\bar{\boldsymbol{X}})'\sim W_p(n-1,\boldsymbol{\Sigma})$$

(2) 设$\boldsymbol{W}_i\sim W_p(n_i,\boldsymbol{\Sigma})$ $(i=1,\cdots,k)$ 相互独立，若令 $n=n_1+\cdots+n_k$，则有

$$\sum_{i=1}^{k}\boldsymbol{W}_i\sim W_p(n,\boldsymbol{\Sigma})$$

这个性质一般称为 Wishart 分布关于自由度 n 具有可加性，这点与χ^2分布类似。

(3) 设 p 阶随机阵 $\boldsymbol{W}\sim W_p(n,\boldsymbol{\Sigma})$，$\boldsymbol{C}_{m\times p}$为常数矩阵，则

$$\boldsymbol{C}\boldsymbol{W}\boldsymbol{C}'\sim W_m(n,\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}')$$

特别地，如果取 $\boldsymbol{C}$ 为向量$\boldsymbol{l}=(l_1,\cdots,l_p)'$，则有$\boldsymbol{l}'\boldsymbol{W}\boldsymbol{l}\sim W_1(n,\boldsymbol{l}'\boldsymbol{\Sigma}l)$，也即$\dfrac{\boldsymbol{l}'\boldsymbol{W}\boldsymbol{l}}{\boldsymbol{l}'\boldsymbol{\Sigma}\boldsymbol{l}}\sim\chi^2(n)$。

1.4.2　Hotelling T^2分布

在一元统计中我们学过，若 $X\sim N(0,1)$，$Y\sim\chi^2(n)$，且 X 与 Y 独立，则随机变量 $t=\dfrac{X}{\sqrt{Y/n}}$服从自由度为n 的t 分布，也称为学生分布。我们还学过，如果将 t 平方，就得到

$$t^2=\frac{nX^2}{Y}\sim F(1,n)$$

即$t^2(n)$服从第一自由度为 1、第二自由度为 n 的中心 F 分布。下面仿照一元情形将t^2的分布推广到 p 元总体的情形。

定义1.8　设 $\boldsymbol{W}\sim W_p(n,\boldsymbol{\Sigma})$，$\boldsymbol{X}\sim N_p(0,\boldsymbol{\Sigma})$，$n\geqslant p$，$\boldsymbol{\Sigma}>0$，且$\boldsymbol{W}$ 与$\boldsymbol{X}$ 相互独立，则称随机变量 $T^2=n\boldsymbol{X}'\boldsymbol{W}^{-1}\boldsymbol{X}$ 所服从的分布称为第一自由度为 p，第二自由度为 n 的 Hotelling T^2分布，记为

$$T^2\sim T^2(p,n)$$

注意　我们可以证明T^2分布只与 n，p 有关，与 $\boldsymbol{\Sigma}$ 无关，因此在表示T^2分布的记号中没有 $\boldsymbol{\Sigma}$。

T^2分布与 F 分布也有一定的关系。在一元统计中，如果 $t=\dfrac{X}{\sqrt{Y/n}}\sim t(n)$，则有 $t^2=\dfrac{X^2}{Y/n}\sim F(1,n)$。推广到 p 元情形，这个关系是$\dfrac{n-p+1}{pn}T^2(p,n)=F(p,n-p+1)$。这一点的证明以及更多相关性质的介绍可以参看相关文献[①]。下面我们不加证明给出T^2分布的两条重要性质。这两条性质在多元正态总体的假设检验中将会用到。

(1) 设$\boldsymbol{X}_{(b)}(b=1,\cdots,n)$是从 p 维正态总体$N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$中抽取的 n 个随机样本，$\bar{\boldsymbol{X}}$ 为样本均值向量，$\boldsymbol{A}$ 为样本离差阵，则统计量

---

① 高惠璇. 应用多元统计分析[M]. 北京：北京大学出版社，2005.

$$T^2=(n-1)\left[\sqrt{n}\,(\bar{\boldsymbol{X}}-\boldsymbol{\mu})\right]'\boldsymbol{A}^{-1}\left[\sqrt{n}\,(\bar{\boldsymbol{X}}-\boldsymbol{\mu})\right]$$
$$=n(n-1)(\bar{\boldsymbol{X}}-\boldsymbol{\mu})'\boldsymbol{A}^{-1}(\bar{\boldsymbol{X}}-\boldsymbol{\mu})\sim T^2(p,n-1)$$

(2) 设有两个 p 维正态总体$N_p(\boldsymbol{\mu}_1,\boldsymbol{\Sigma})$，$N_p(\boldsymbol{\mu}_2,\boldsymbol{\Sigma})$，从这两个总体中抽出容量分别为$n_1$和$n_2$的两个样本。记$\bar{\boldsymbol{X}}_1$，$\bar{\boldsymbol{X}}_2$为两样本的均值向量，$\boldsymbol{S}_1$，$\boldsymbol{S}_2$为两样本协方差阵，并记

$$\boldsymbol{S}_p=\frac{n_1\boldsymbol{S}_1+n_2\boldsymbol{S}_2}{n_1+n_2-2}$$

若$\boldsymbol{\mu}_1=\boldsymbol{\mu}_2$，则

$$\frac{n_1n_2}{n_1+n_2}(\bar{\boldsymbol{X}}_1-\bar{\boldsymbol{X}}_2)'\boldsymbol{S}_p^{-1}(\bar{\boldsymbol{X}}_1-\bar{\boldsymbol{X}}_2)\sim T^2(p,n_1+n_2-2)$$

1.4.3 Wilks Λ 分布

我们在数理统计中学过，若 $X\sim\chi^2(m)$，$Y\sim\chi^2(n)$，且 X 与 Y 相互独立，则

$$F=\frac{X/m}{Y/n}\sim F(m,n)$$

在一元统计中 F 分布主要用来做方差齐性检验，两个总体的样本方差的比在原假设下是服从 F 分布的。在多元总体中，样本的协方差阵是一个矩阵，不能再简单相除得到统计量了。因此，我们考虑用与协方差阵有关的一个量来描述总体的离散程度(或者称为变异度)。这样的参数我们一般称为广义方差。有哪些数量指标来描述广义方差呢？一般而言，多用矩阵的行列式、迹或者特征值来描述。目前最常用的是利用行列式定义的。有了广义方差的定义，再仿照 F 分布的定义，称两个广义方差之比的统计量为 Wilks Λ 统计量。

定义1.9 设 $\boldsymbol{X}\sim N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$，则称协方差阵的行列式$|\boldsymbol{\Sigma}|$为 $\boldsymbol{X}$ 的广义方差。再设$\boldsymbol{A}_1\sim W_p(n_1,\boldsymbol{\Sigma})$，$\boldsymbol{A}_2\sim W_p(n_2,\boldsymbol{\Sigma})$ $(\boldsymbol{\Sigma}>0,\ n_1\geqslant p)$，且$\boldsymbol{A}_1$与$\boldsymbol{A}_2$独立，则称

$$\Lambda=\frac{|\boldsymbol{A}_1|}{|\boldsymbol{A}_1+\boldsymbol{A}_2|}$$

为 Wilks 统计量或 Λ 统计量，其所遵从的分布称为 Wilks 分布，记为 $\Lambda\sim\Lambda(p,n_1,n_2)$。

Wilks 分布比较复杂，在不同的情形下许多学者对其精确分布及近似分布都进行了深入的研究。当 p 或者n_2比较小而$n_1>p$ 时，可以通过 F 分布得到 Λ 统计量的精确分布，具体情况如表 1-1 所示。

表 1-1 $\Lambda\sim\Lambda(p,n_1,n_2)$与 F 分布的关系$(n_1>p)$

p	n_2	统计量 F	F 的自由度
任意	1	$\frac{(n_1-p+1)}{p}\cdot\frac{(1-\Lambda)}{\Lambda}$	$p,\ n_1-p+1$
任意	2	$\frac{(n_1-p)}{p}\cdot\frac{(1-\sqrt{\Lambda})}{\sqrt{\Lambda}}$	$2p,\ 2(n_1-p)$
1	任意	$\frac{(1-\Lambda)}{\Lambda}\cdot\frac{n_1}{n_2}$	$n_2,\ n_1$
2	任意	$\frac{(1-\sqrt{\Lambda})}{\sqrt{\Lambda}}\cdot\frac{(n_1-1)}{n_2}$	$2n_2,\ 2(n_1-1)$

当$n_2>2$，$p>2$ 时，我们有这样的近似分布：

$$当 n_1\to\infty，-\left[n_1-\frac{1}{2}(p-n_2+1)\right]\ln\Lambda\sim\chi^2(p\ n_2)。$$

此外，类似于 F 分布中 $F(n，m)$ 与 $\dfrac{1}{F(m，n)}$ 同分布，Λ 分布也有一个类似的性质：若 $n_2<p$，则 $\Lambda(p，n_1，n_2)=\Lambda(n_2，p，n_1+n_2-p)$。

【课后练习】

1. 设 $(X_1，X_2，X_3)$ 的联合密度为

$$f(x_1，x_2，x_3)=\begin{cases}\dfrac{1-\sin x\sin y\sin z}{8\pi^2} & 0\leqslant x\leqslant 2\pi，0\leqslant y\leqslant 2\pi，0\leqslant z\leqslant 2\pi\\ 0 & 其他\end{cases}$$

（1）求 X_1 的边缘密度。

（2）求 $(X_1，X_2)$ 的边缘密度。

（3）试证明 X_1，X_2，X_3 两两独立但不互相独立。

2. 设

$$\boldsymbol{A}=\begin{bmatrix}1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3}\\ 1/\sqrt{2} & -1/\sqrt{2} & 0\\ 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6}\end{bmatrix}$$

（1）试证明 $\boldsymbol{A}$ 是一个正交矩阵（即 $\boldsymbol{A}\boldsymbol{A}'=\boldsymbol{I}_3$）。

（2）已知 $X\sim N_3(\mu\boldsymbol{I}_3，\sigma^2\boldsymbol{I}_3)$，设 $\boldsymbol{Y}=(Y_1，Y_2，Y_3)'=\boldsymbol{A}\boldsymbol{X}$，试证明

① $Y_2^2+Y_3^2+Y_4^2=\sum\limits_{i=1}^{3}(X_i-\overline{X})^2$，其中 $\overline{X}=\dfrac{1}{3}(X_1+X_2+X_3)$；

② $Y_1\sim N(\sqrt{3}\mu，\sigma^2)$，$Y_2，Y_3\sim N(0，\sigma^2)$；

③ $Y_1，Y_2，Y_3$ 相互独立。