试验设计与数据分析

项可风　吴启光

上海科学技术出版社

试验设计与数据分析

项可风　吴启光

责任编辑　赵序明

上海科学技术出版社出版

（上海瑞金二路 450 号）

新华书店上海发行所发行　上海商务印刷厂印刷

开本 787×1092 1/32　印张 19.5　字数 427,000

1989 年 12 月第 1 版　1989年 12 月第 1 次印刷

印数：1—1,800

ISBN 7-5323-0368-3/O·26

定价：10.40 元

序　　言

试验设计与数据分析是研究如何正确地安排试验和分析数据的一门统计学科．它的一些基本方法是由著名统计学家R. A. Fisher在1935年著的《试验设计》一书中提出来的．当时这些方法主要用于农业、生物和医学试验．

近三十多年来，由于科学技术的发展，尤其是计算机的普及，促使这门学科在内容和方法上有着重大进展，目前已成为统计学中的一个重要应用分支．不仅在实验科学领域，而且在产品设计、管理科学和社会科学等领域中也日益广泛地应用试验设计中的方法．

在我国，对试验设计的应用，最先是由农业科学家从西方引进的，当时只限于农业田间试验．五十年代后期，在著名统计学家许宝騄教授创导下，数学工作者才深入这个领域．许先生在三、四十年代就在线性模型中作出过重大贡献，六十年代初他领导了一个试验设计讨论班，为国家培养了一批试验设计工作者．本书的一些章节收集了许先生的一些著名结果，兼作对许先生的纪念．二十多年来，由于我国数学工作者的努力，正交试验法在全国得到一定程度的普及，取得相当多的应用成果．在这方面也出版了一些普及性读物．但广大科技工作者不满足一般科普读物，还要求系统地学习试验设计知识，以便把试验设计中的更多方法用于解决实际问题．

近年来，作者曾在多种场合为工程技术人员、大专院校教师、统计专业的学生和研究生讲授试验设计课．本书就是在

这些讲稿的基础上经过整理并充实了近期文献中的一些新成果，汇集成册，以满足以上读者的要求.

就数学内容来说，试验设计不仅涉及统计分析，而且与组合数学有着密切的关系. 然而，目前国内正缺少把这两个方面结合起来从理论上系统地介绍这个学科的书，本书试图弥补这种不足. 根据我国读者的情况，本书力求从理论和应用的统一上去介绍这门学科的内容和方法. 因此，既要介绍问题的背景，又要给出数学的证明，使读者在学完这本书以后，除了能对这门学科在理论上有着系统地了解外，还能应用它的方法去解决实际中的问题.

全书共分十六章. 前四章是介绍试验设计中所必要的线性模型的知识. 对统计专门化的学生和研究生来说，这四章内容是进入这个分支，阅读有关专著和文献所必需的基础知识. 但对于应用试验设计方法的工程技术员，可跳越前四章. 第五、六两章重点介绍试验设计中的数据分析方法. 第七章主要通过两因素试验介绍多因素析因试验中的模型与设计问题. 弄通这章内容才能更好地理解第八、九章中所讨论的问题. 对于工程技术人员，要求掌握第五至七章的基本内容. 为了帮助读者学会基本方法，在这三章中附有一些实例. 第八、九两章要用到较多的代数知识，非数学专业的读者可跳越数学的证明，了解主要结果就可以了. 第十章介绍正交试验的数学原理，帮助读者理解一些普及读物中没有证明的内容. 第十一章证明正交试验的优良性. 对数学问题有兴趣的读者可读这章. 第十二至第十六章是实际工作中常用的一些方法. 非数学专业的读者可根据问题需要选读有关章节. 本书最后给出两个附录. 在附录 A 中，介绍阿达玛型和差集型正交表的构造，加上正文中已介绍的完备型表，掌握了

这三种方法，就可制作试验中所需要的各种正交表．在附录B中给出少量统计用表，为节省篇幅，凡在《常用数理统计表》中能查到的统计用表，均未列入附录B中．

本书的第十一、十四两章由吴启光同志执笔，本人对全书作了统一处理．

作者深切感谢中国科学技术大学研究生院、北京大学、清华大学、北京师范大学、浙江大学和山西大学等概率统计专业的师生，他们为作者提供了讲授这门课程的机会和对本书初稿提出了宝贵的意见．

在本书写作过程中，曾得到陈希孺、成平和刘璋温等同志的关心与鼓励．刘璋温同志为作者提供了组合设计的材料．在书稿写成后，华东师范大学的茆诗松和周纪芗等同志详细地审阅了底稿，提出许多宝贵的修改意见．作者在此向他们表示衷心感谢．

对书中不当之处，恳切希望专家和读者提出批评指正．

项可风

于中国科学院系统科学所

目　　录

第 一 章

正态随机变量及其有关统计量

在数据的统计分析中，正态随机变量扮演着重要角色。本章专门讨论正态变量及其有关统计量的分布问题，它是以后各章的主要基础理论。阅读这章所需要的概率统计及矩阵知识，都假定读者已经学过，没有学过的读者可在参考文献[10]与[54]中找到有关材料。

§1.1 多维正态变量及其线性函数的分布

众所周知，具有密度函数

$$f(y)=\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\} \tag{1.1.1}$$

的随机变量称为一维正态变量。其中 μ 称为均值，σ^2 称为方差。因为(1.1.1)式由 μ 与 σ^2 所完全确定，故简记为 $N(\mu,\sigma^2)$。特别是，当 $\mu=0$, $\sigma^2=1$ 时，称为标准正态变量。记分布为 $N(0, 1)$。

但是，往后我们主要讨论具有联合密度

$$f(\boldsymbol{y})=\frac{1}{\sqrt{(2\pi)^p|\boldsymbol{A}|}}\exp\left\{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\eta})'\boldsymbol{A}^{-1}(\boldsymbol{y}-\boldsymbol{\eta})\right\} \tag{1.1.2}$$

的 p 维正态变量。其中

$$\boldsymbol{y}=\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix}, \qquad \boldsymbol{\eta}=\begin{pmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_p \end{pmatrix},$$

$$\boldsymbol{A}=\begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix}.$$

$|\boldsymbol{A}|$是矩阵 $\boldsymbol{A}$ 的行列式, $\boldsymbol{A}'$ 表 $\boldsymbol{A}$ 的转置阵. 这里的 $\boldsymbol{A}$ 是对称矩阵, 即 $\boldsymbol{A}=\boldsymbol{A}'$. 往后我们均用大写黑体字母表示矩阵, 小写黑体字母表示向量.

$\boldsymbol{\eta}$ 是均值向量, $\boldsymbol{A}$ 为协方差矩阵, (1.1.2)由 $\boldsymbol{\eta}$ 和 $\boldsymbol{A}$ 所完全确定, 因此, 我们用 $N_p(\boldsymbol{\eta}, \boldsymbol{A})$ 表示 p 维正态分布.

下面推导多维正态变量的一些性质.

一、设 $\boldsymbol{y}$ 为 p 维正态变量, 具有分布 $N_p(\boldsymbol{\eta}, \boldsymbol{A})$, $\boldsymbol{B}$ 是 $p\times p$ 实满秩矩阵. 若令

$$\boldsymbol{z}=\boldsymbol{B}\boldsymbol{y} \tag{1.1.3}$$

经变换后, $\boldsymbol{z}$ 也是 p 维随机变量, 其变换的 Jacobian 行列式为$|\boldsymbol{B}^{-1}|$. 由(1.1.2), 可得到 $\boldsymbol{z}$ 的密度为

$$f(\boldsymbol{z})=\frac{1}{\sqrt{(2\pi)^p|\boldsymbol{BAB}'|}}\exp\{-\frac{1}{2}(\boldsymbol{z}-\boldsymbol{B\eta})'(\boldsymbol{BAB}')^{-1}(\boldsymbol{z}-\boldsymbol{B\eta})\}. \tag{1.1.4}$$

故 $\boldsymbol{z}$ 的分布为 $N_p(\boldsymbol{B\eta}, \boldsymbol{BAB}')$. 也就是说, 正态变量经过满秩变换后, 其正态性保持不变.

由于 $\boldsymbol{A}$ 是正定矩阵, 故存在满秩矩阵 $\boldsymbol{C}$, 使 $\boldsymbol{A}=\boldsymbol{CC}'$. 因此, 若在(1.1.3)中, 取 $\boldsymbol{B}=\boldsymbol{C}^{-1}$, 则 $\boldsymbol{z}$ 的分布为 $N_p(\boldsymbol{C}^{-1}\boldsymbol{\eta}, \boldsymbol{I})$. 也就是说, 对任何多维正态变量, 总可找到一个线性变换, 经

变换后，使其各分量为相互独立方差为 1 的多维正态.

二、现在求 $\boldsymbol{y}$ 的特征函数. 当 y 为一维 $N(0, 1)$时，不难得到其特征函数为

$$\varphi(t)=\exp\left\{-\frac{t^2}{2}\right\}. \tag{1.1.5}$$

因此，由(1.1.5)$N(\mu, \sigma^2)$的特征函数，即为

$$\varphi(t)=\exp\left\{i\mu t-\frac{1}{2}\sigma^2 t^2\right\}. \tag{1.1.6}$$

由(1.1.6)，当 $\boldsymbol{y}$ 为 $N_p(\boldsymbol{C}^{-1}\boldsymbol{\eta}, \boldsymbol{I})$时，其特征函数为

$$\varphi(\boldsymbol{t})=\exp\left\{i(\boldsymbol{C}^{-1}\boldsymbol{\eta})'\boldsymbol{t}-\frac{1}{2}\boldsymbol{t}'\boldsymbol{t}\right\}. \tag{1.1.7}$$

当 $\boldsymbol{y}$ 为 $N_p(\boldsymbol{\eta}, \boldsymbol{A})$时，由(1.1.7)作一变换，就得其特征函数

$$\varphi(\boldsymbol{t})=\exp\left\{i\boldsymbol{\eta}'\boldsymbol{t}-\frac{1}{2}\boldsymbol{t}'\boldsymbol{A}\boldsymbol{t}\right\}. \tag{1.1.8}$$

反之，对一切非负对称矩阵 $\boldsymbol{A}$，(1.1.8)也可作为多维正态变量的一个定义，可以不涉及密度函数. 当$|\boldsymbol{A}|=0$ 时，称为退化(降维)正态变量. 其密度函数不能表成 (1.1.2) 的形式.

三、设 $\boldsymbol{a}$ 是常数向量，$L(\boldsymbol{y})\triangleq\boldsymbol{a}'\boldsymbol{y}$ 是正态变量的一个线性函数，由(1.1.8)，其特征函数为

$$\begin{aligned}\varphi(t)&=E\exp\{it\boldsymbol{a}'\boldsymbol{y}\}\\&=\exp\left\{it\boldsymbol{a}'\boldsymbol{\eta}-\frac{1}{2}t^2\boldsymbol{a}'\boldsymbol{A}\boldsymbol{a}\right\}.\end{aligned} \tag{1.1.9}$$

由(1.1.6)，$L(\boldsymbol{y})$的分布为 $N(\boldsymbol{a}'\boldsymbol{\eta}, \boldsymbol{a}'\boldsymbol{A}\boldsymbol{a})$. 也就是说，正态变量的线性函数 $L(\boldsymbol{y})$依然是正态变量，其均值

$$E(L(\boldsymbol{y}))=\boldsymbol{a}'E(\boldsymbol{y})=\boldsymbol{a}'\boldsymbol{\eta},$$

而方差

$$V_{ar}(L(\boldsymbol{y}))=\boldsymbol{a}'V_{ar}(\boldsymbol{y})\boldsymbol{a}=\boldsymbol{a}'\boldsymbol{A}\boldsymbol{a}.$$

有趣的是，若 $\boldsymbol{y}$ 为 p 维随机变量，具有有限

$$E(\boldsymbol{y})=\boldsymbol{\eta},\ V_{ar}(\boldsymbol{y})=\boldsymbol{A}.$$

若对任意 $\boldsymbol{a}$, $\boldsymbol{a}'\boldsymbol{y}$ 均为一维正态变量,则 $\boldsymbol{y}$ 一定是 $N_p(\boldsymbol{\eta},\ \boldsymbol{A})$. 因为只要在(1.1.9)中,令 $t=1$,即得

$$\varphi(\boldsymbol{a})=\exp\left\{i\boldsymbol{a}'\boldsymbol{\eta}-\frac{1}{2}\boldsymbol{a}'\boldsymbol{A}\boldsymbol{a}\right\},$$

它相当于在(1.1.8)中取 $\boldsymbol{a}=\boldsymbol{t}$. 于是得到下面的定理:

定理1.1.1 $\boldsymbol{y}$ 为 $N_p(\boldsymbol{\eta},\ \boldsymbol{A})$ 的充要条件是: 对任意 $\boldsymbol{a}$, $\boldsymbol{a}'\boldsymbol{y}$ 为 $N(\boldsymbol{a}'\boldsymbol{\eta},\ \boldsymbol{a}'\boldsymbol{A}\boldsymbol{a})$.

在一些书中,直接把定理1.1.1作为多维正态的定义.如同用特征函数(1.1.8)定义多维正态一样, 也不涉及密度函数,当 $\boldsymbol{A}$ 不满秩时, 为退化分布. 若不特别申明,下面的讨论包括退化正态分布.

四、设 $\boldsymbol{B}$ 是 $m\times p$ 阶矩阵, $\boldsymbol{y}$ 为 $N_p(\boldsymbol{\eta},\ \boldsymbol{A})$, 现可推广(1.1.3)变换为: $\boldsymbol{z}=\boldsymbol{B}\boldsymbol{y}$. 这时 $\boldsymbol{z}$ 为 m 维随机变量,由(1.1.8), $\boldsymbol{z}$ 的特征函数为

$$\varphi(\boldsymbol{t})=E\exp(i\boldsymbol{t}'\boldsymbol{B}\boldsymbol{y})=\exp\left(i\boldsymbol{t}'\boldsymbol{B}\boldsymbol{\eta}-\frac{1}{2}\boldsymbol{t}'\boldsymbol{B}\boldsymbol{A}\boldsymbol{B}'\boldsymbol{t}\right),$$

故 $\boldsymbol{z}$ 为 $N_m(\boldsymbol{B}\boldsymbol{\eta},\ \boldsymbol{B}\boldsymbol{A}\boldsymbol{B}')$,于是得到以下定理:

定理1.1.2 正态变量经过任意的线性变换后依然是正态变量.

这里需要注意的是: 当 $rk(\boldsymbol{B})<m$ 时, $\boldsymbol{z}$ 是退化分布. 然而有下面的定理:

定理1.1.3 若 $\boldsymbol{y}$ 为 $N_p(\boldsymbol{\eta},\ \boldsymbol{A})$, $rk(\boldsymbol{A})=m<p$, 则存在一个变换(1.1.10),使 $\boldsymbol{z}$ 为 $N_m(\boldsymbol{B}\boldsymbol{\eta},\ \boldsymbol{I})$, 即 $\boldsymbol{z}$ 的各分量相互独立、方差为1的分布.

证 因 $\boldsymbol{A}$ 是对称非负的,由矩阵秩的分解定理: $\boldsymbol{A}=\boldsymbol{C}\boldsymbol{C}'$, 其中 $\boldsymbol{C}$ 为 $p\times m$, $rk(\boldsymbol{C})=m$. 故 $\boldsymbol{C}'\boldsymbol{C}$ 是满秩方阵. 令 $\boldsymbol{z}=$