

林业试验设计 技术与方法

洪伟 编著

北京科学技术出版社

Q52117

林业试验设计 技术与方法

洪 伟 编著



北林图 A00111191



北京科学技术出版社

425449

(京)新登字 207 号

林业试验设计技术与方法

洪 伟 编著

*

北京科学技术出版社出版

(北京西直门南大街 16 号)

邮政编码 100035

新华书店总店科技发行所发行·各地新华书店经销

朝阳第一教师进修学校胶印厂印刷

*

787×1092 毫米 32 开本 12.875 印张 273 千字

1993 年 1 月第 1 版 1993 年 1 月第 1 次印刷

印数: 1—3500 册

ISBN7-5304-1170-5/S·090 定价: 6.30 元

序

科学施惠造福于人类,科学的发展有赖于科学的试验研究,而试验研究首先要有正确的试验方案。鉴定一个科研成果,首先看你的试验设计是否科学、方法是否正确,这是判断成果水平和价值的前提。有了科学试验设计和实施方法,才能取得可靠的数据,而在大量数据面前又要有正确的统计方法,才能找出符合客观的规律和正确的结论。

一个科学、准确的试验设计,可以最大限度地节约试验成本,提高试验成效,缩短试验周期,同时又能获得可靠的结论。特别是在电子计算机迅速发展、应用不断深入的今天,在林业生产和科研工作中,试验设计技术和方法已成为制订科研方案、分析统计数据的必备手段。

作为林业工作者,面临的自然条件十分复杂,地形、气候、土壤等变化多端,林业生产周期长,这要求我们必须掌握更多的试验设计的知识,才能适应实际需要,取得事半功倍的效果。可以看出,试验设计在林业上有着广泛的应用前景,是林业工作者尤其是院校师生及研究工作人员必备的基础知识。目前此类书籍很少,洪伟同志编写的《林业试验设计技术与方法》正好应时而出,满足了读者的需要。

该书系统地介绍了多种试验设计和方法的基本原理、计算步骤,原理与方法并重,并结合作者多年的科研实践的实例,既讲清概念原理、又有实际操作范例,通俗易懂,既可做为教材用,也可用于自学。

试验设计近年来发展很快,象回归设计、回归旋转设计、混料设计、三次设计等在林业生产和科研中尚少见报导,该书

也就这部分内容作了较深入介绍,并例举其在林业科研中的应用,这也是本书的一个重要特色。

可以预期该书的出版将有利于林业科研事业的发展,促进林业生产水平的提高。

俞新妥

1991.12.6

前 言

试验设计是在数理统计的基础上,逐渐发展起来的一门应用统计学分支学科。它越来越受到人们的重视,特别是目前电子计算机应用已深入各学科,因而试验设计的应用范围日益广泛。在林业的科研和生产中,试验设计已成为制定科研方案和分析实验数据的必要手段,应用于林木育种、地理种源、栽培技术、营养诊断、生理测定、病虫害防治、规划设计、森林生态等领域。同时林业试验设计也逐渐受到林业工作者的高度重视,应把掌握林业试验设计作为林业工作者的基本素质。

为了使更多人了解和掌握试验设计的原理与方法,提高林业科学研究水平,需要一定数量的适用于林业院校师生、研究生以及广大林业工作者的试验设计参考书籍。但此方面的书籍却甚少,与我国林业科学的发展现状和科技兴林的需求不相适应。

本书在编写时主要考虑林业工作实际应用,重点介绍多种方法的基本原理和计算步骤,原理与方法并重,结合编著者和前人的科研实践的实例,力求通俗易懂,既讲清概念、原理,又提供实际操作范例,除介绍了常用试验设计方法外,着重介绍回归设计、组合设计、旋转设计、饱和 D -最优设计、混料试验设计、三次设计等现代的试验设计方法。本书所使用的数学工具没有超出高等林业院校高等数学、数理统计等课程的教学大纲范围。为了让没有掌握上述知识的人看懂其原理方法提供方便,编写了预备知识内容,对已掌握统计基本知识的方差分析和回归分析的读者可以越过这二章学习。

本书编著过程中,得到作者所在的福建林学院领导的大力支持,特别是我国林业界著名专家俞新妥教授在百忙之中为本书写了序言。谨此,对支持和关心本书编写出版的同志们表示衷心感谢!

由于编著者水平有限,书中错误及不妥之处,敬请广大读者批评指正。

洪 伟

1991.12.20

目 录

第一章 试验设计	(1)
第一节 试验设计的几个基本概念.....	(2)
第二节 试验设计的基本原理.....	(3)
第三节 试验的基本要求.....	(7)
第四节 试验设计的基本程序.....	(8)
第五节 试验设计中的小区技术	(10)
第二章 方差分析	(12)
第一节 单向分组的方差分析	(13)
第二节 两向分组的方差分析	(39)
第三节 系统分组的方差分析	(65)
第四节 两个以上因素的方差分析	(70)
第五节 数据转换与漏失数据的弥补	(76)
第三章 回归分析	(82)
第一节 一元回归模型	(83)
第二节 一元非线性回归.....	(103)
第三节 多元回归.....	(114)
第四章 完全随机化试验和随机区组试验	(133)
第一节 完全随机化试验.....	(133)
第二节 随机区组设计.....	(135)
第五章 拉丁方设计和希腊拉丁方设计	(140)
第一节 拉丁方设计.....	(140)
第二节 希腊拉丁方设计.....	(143)
第六章 正交试验设计	(148)

第一节	正交试验设计法.....	(148)
第二节	交互作用的试验设计与统计分析.....	(159)
第三节	有重复试验的正交试验.....	(164)
第四节	水平数不等的试验设计.....	(168)
第五节	单纯形最优化方法的应用.....	(172)
第七章	平衡不完全区组设计.....	(180)
第一节	平衡不完全区组设计.....	(180)
第二节	平衡格子设计.....	(189)
第三节	部分平衡格子设计.....	(192)
第八章	裂区设计.....	(196)
第一节	裂区设计.....	(196)
第二节	林木多点多年试验设计.....	(205)
第九章	协方差分析.....	(212)
第一节	协方差分析方法.....	(213)
第二节	随机区组设计的协方差分析.....	(222)
第三节	多个协变量的协方差分析.....	(228)
第十章	回归正交设计.....	(232)
第一节	一次回归正交设计.....	(233)
第二节	二次回归正交设计.....	(234)
第十一章	回归旋转设计.....	(263)
第一节	回归旋转设计的几个重要问题.....	(263)
第二节	二次正交旋转组合设计的统计方法.....	(270)
第三节	通用旋转组合设计的统计方法.....	(277)
第十二章	饱和 D—最优设计	(284)
第一节	D —优良性	(284)
第二节	一次饱和 D —最优设计	(289)
第三节	二次饱和 D —最优设计	(293)

第十三章	混料试验设计	(302)
第一节	单形格子设计.....	(302)
第二节	单形重心设计.....	(317)
第三节	极端顶点设计.....	(322)
第十四章	三次设计	(328)
第一节	三次设计.....	(328)
第二节	三次设计在参数优化中的应用.....	(330)
参考文献		(335)
附表		(336)

第一章 试验设计

在林业生产和科学试验中,要经常研究哪些生产措施和外界条件对林木生长有影响,以寻求优质、高产、低消耗的生产方案,培育林木良种,确定防治森林病虫害的措施等。这些都离不开试验,而试验与数据是分不开的。有经验的试验者为了使试验结果具有真实性、客观性,总是在试验前就根据统计学的要求制定出一套完整的试验方案。试验过程就是方案的实施过程,依靠合理的试验设计得出正确的判断和结论。

进行试验的目的是为了获得条件与结果之间规律性的认识。一个好的试验设计可以最大限度地节约成本,缩短试验周期,同时又能迅速获得确切的科学结论。

试验设计的广义理解是指整个试验研究的课题设计。它应当包括三个组成部分:

1. 确定(设计)试验处理方案(包括小区技术)。
2. 观察资料的搜集与整理。
3. 统计分析方法。

首先要明确几点:试验的目的是什么,考核和评价的指标是什么,考察的因素是什么。然后根据试验的目的来合理设计试验方案,组织试验具体实施,最后对试验结果进行统计分析。

试验设计作为统计学的一个分支,内容十分丰富,本书将阐述在林业生产和科研中常用的试验设计方法:有完全随机化试验、全随机区组试验设计、拉丁方设计、裂区设计、BIB 设

计、回归正交设计等。

试验设计是为实验服务的,试验者必须有实事求是的科学精神。为了达到不同的目的,应该采取不同的试验设计方法,根据试验要求灵活掌握,不能生搬硬套。

第一节 试验设计的几个基本概念

要掌握试验设计具体原理和方法,首先涉及到几个基本概念,这些基本概念有助于我们对后续知识了解。

一、指标

在试验设计中把判断试验结果好坏所采用的标准称为试验指标,简称为指标。例如试验目的为判断杀虫剂杀虫效果时,可用昆虫的死亡率作为试验指标;当试验目的在于了解不同的树种的速生丰产性质时,可用单位面积蓄积量作为试验指标。

二、因素

有可能影响试验指标的条件称为因素或称为因子。例如施肥量、栽植密度等都是可能影响苗木产量这一指标的条件。又如药剂种类、昆虫对药剂的抗性等都可能影响昆虫死亡率这一指标的条件。因此,施肥量、栽植密度、药剂种类、昆虫对药剂的抗性都可以作为分析试验的因素。

因素可分为数量因素和非数量因素。数量因素——依数量划分水平的因素,如栽植密度、施肥量等。非数量因素——不依数量划分水平的因素。如施肥种类、药剂种类等。

三、水平

能影响试验指标的因素通常可以人为地加以控制或分

组,所划分的组通常也叫做多因素类别和等级,统计上称其为因素的水平。例如杉木栽植密度试验其因素分为 180 株/亩、200 株/亩、220 株/亩、250 株/亩、280 株/亩五个等级,就说该栽植密度这一因素分成五个水平。但是在科研试验中,因素和水平这两个概念并不能严格区分。例如,杉木施肥试验中施肥种类是因素,如果该因素分为三个水平即人类尿、过磷酸钙、尿素。这时,每一个水平代表一种肥料。在试验中还可能称每一种肥料为因素,施肥量分为几个等级,每一个等级为一个水平。对于一个确定的试验方案中的因素和水平是不会混淆不清的。

第二节 试验设计的基本原理

统计学的理论认为,欲对试验资料检验其显著性,必须计算试验的偶然误差,即从试验中获得适当而合理的误差估计量。如果没有一个合理的试验设计,这种合理的偶然误差就无法估计。因此试验设计既应注意试验方案的设计和实施,还应考虑到具体的统计分析方法与试验方案的联系。近代试验设计遵循三个基本原理:重复、随机排列、局部控制。

一、设置重复

所谓重复就是将一基本试验重做一次或几次。例如,品种比较试验,每品种各设置 4 个小区,就称为 4 次重复。重复有两个作用:估计试验误差,要有误差估计值,才能判断处理之间差异的显著性,而误差估计值可以从重复试验中得到;更精确地估计处理效应(或者说可以减少试验误差)。例如比较两种不同药物防治某种林木害虫,每种药物只做一次试验,甲药

物 240 小时后害虫死亡率达 60%，乙药物 284 小时后害虫死亡率达 60%。这时我们就无法正确判断药效有无差异。两种药物之间差异，可能是药效不同，也可能是由于试验误差造成。如果上述结果是几次重复的平均值，由于平均数方差只是样本方差的 $1/n$ ，当 n 足够大时，两种药物之间差异，就可以认为是药物效应。

重复试验不但能测定误差估计值，还能减少误差估计值，提高试验精度，同时还能扩大试验的代表性。

重复次数多少合适是值得研究的问题。重复次数太少，试验误差大且不准确；重复次数太多，则费地、费时、又费工。但一般认为试验误差自由度至少应当 10 至 15，同时还应根据试验地形变化、土壤肥力差异、试验准确度要求等方面加以综合考虑。如果试验地形变化大，土壤肥力差异大，试验准确度要求较高时，重复次数应该多些，否则可以少些。

二、随机排列

随机排列是指试验材料的配置和处理在试验地上的顺序是随机排列的。随机排列是试验设计中使用统计学方法分析的基础。统计学方法要求观察值或误差是独立的随机变量。通过随机排列，才能使这一假设得到可靠依据。随机排列可以使各处理在试验地中占据任何一个小区的机会均等，防止系统误差，达到获得合理的误差估计的目的。

顺序排列在理论上的缺点就是在估计试验误差时，不是过低就是过高估计，不能真实地代表试验的真正偶然误差。例如有一试验包括 5 个处理，重复 4 次，可用顺序排列如图 1.1、1.2 所示。

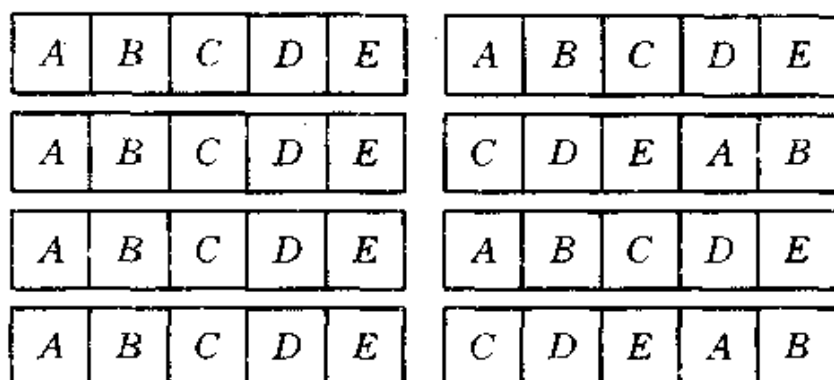


图 1.1 顺序排列(I) 图 1.2 顺序排列(II)

图 1.1 的顺序排列使同处理间的试验小区位置较近,而异处理间试验小区距离较远,其结果增加处理间的方差使试验误差的估计过低。图 1.2 是一种阶梯式的顺序排列,它使同处理的试验小区之间的距离比异处理间的小区之间的距离远,其结果增加试验的方差,以致使试验误差估计过高。

三、局部控制

增加重复可以提高试验精度,但是增加重复也就增大了试验地的面积。土壤肥力的差异是客观存在的。一般说来相邻地区的土壤肥力比较相近。因此增大试验面积必然增大土壤的变异程度,增大土壤差异必然增加试验误差,这样必然会使试验的重复区增多到一定程度后,继续增多会降低试验精确度。为了解决这一矛盾,我们将试验地划分为若干个区组,每一区组按供试的处理划分相同数目的试验小区,这样每一小区配置一个不同处理,由于不同处理配置在较小面积而相邻在一个区组内,完成一个区组的试验就是一次重复(图 1.3),这样可以达到控制土壤变异的目的。这一原理称为“局部控制”原理。

林业土壤肥力差异的表现形式有两种:一是肥力趋向,一般来说土壤的水肥状况的变化趋向为由山顶向山脚,由山坡

向山凹,由阳坡、半阳坡向阴坡;二是小地形变化表现出大斑块状的差异。因此要使用同一重复区组内的各试验小区处在同一坡向和同一坡位区组内(图 1.4)。

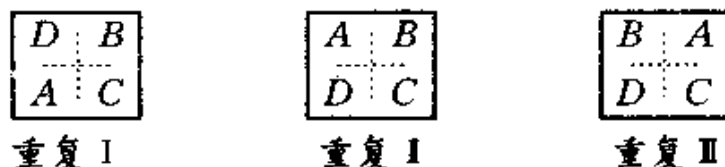


图 1.3 4 个处理 3 个区组的林地配置

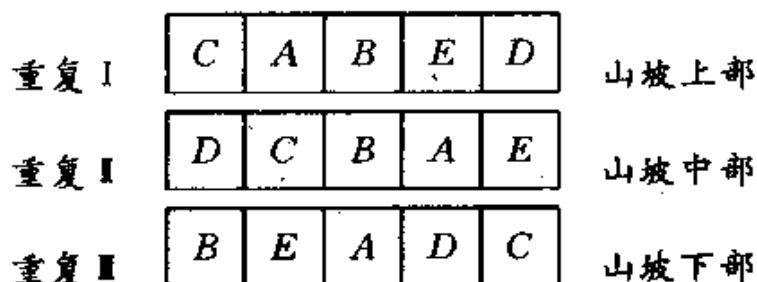


图 1.4 5 个处理 3 个区组的林地配置

综上所述,试验设计的三个基本原则,其最终目的都是为了提高试验结果的精确度,我们只有正确地应用这三个原则,并贯穿实施在试验中,才能得到精确度高可靠试验结果。它们三者关系如下(图 1.5)。

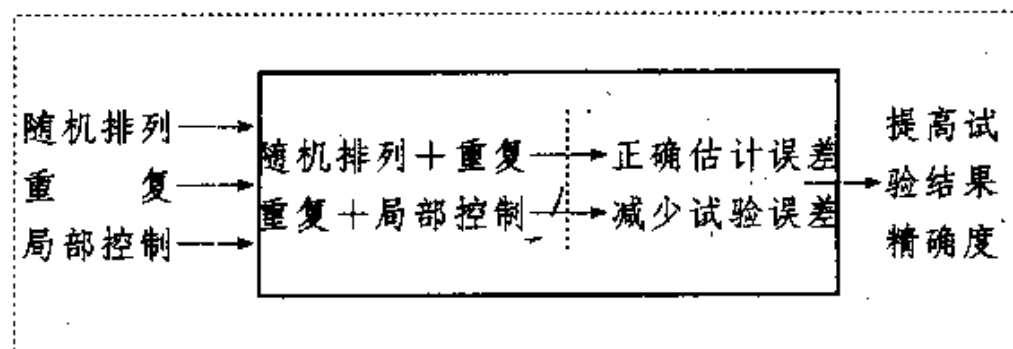


图 1.5 试验实施

第三节 试验的基本要求

为了使试验设计所得结果是可靠和正确的,设计者必须注意以下基本要求:

一、试验的结果要可靠

这里包括试验的精确度和准确度两个方面。

准确度是指试验中每一个处理的实际值或所研究特性的观察值与其相应真值的接近程度。愈是接近则试验愈准确。但在试验中真值往往未知,故准确度不易确定;精确度是指试验误差尽可能小,使处理间差异能精确地比较。

只有能正确反映实际的试验结果,才能起指导生产和促进生产的作用。因此在进行试验的全过程中,必然尽最大努力准确地执行各项试验技术,力求避免人为差错,特别要注意试验条件的一致性。除试验处理以外,其它条件应尽可能一致,这样才能提高试验结果的准确度和精确度。

二、试验条件要有代表性

试验条件要有代表将来准备采用这些试验结果的地区的自然条件,例如要代表本地区的土质、土壤肥力、海拔高度、气候和其它自然条件。只有这样,得出的试验结果才能在本地区或类似地区推广。

三、试验的重演性

在相同条件下再进行试验或实践时能重复获得与原试验相类似的结果,这对在林业生产中推广科研成果有重要意义。

林业试验是受复杂的自然条件影响的,不同年份、不同地区进行相同试验往往结果不同。为了保证重演性,首先要保证

准确性和代表性,在此基础上还要了解和掌握林木生长发育的生物学特性。

四、应当选择适当的试验指标,并有相应的数据分析方法

分析试验结果最基本的统计方法是方差分析。因此试验结果数据须满足方差分析的基本模型要求,如正态、独立、等方差等要求。若不能满足则需采取相应措施,如数据转换等。

第四节 试验设计的基本程序

一、实验目的

实验目的是试验设计首先要考虑的问题,对其应当深入了解,认真分析,提出实验目的及预期效果,避免盲目性。

例如进行间伐强度的试验,其目的是确定间伐强度对林分生长的影响方式和强度,至于间伐强度对林地植被的影响和演变不一定作为实验目标。

二、因素和水平的确定

试验设计之前必须了解哪些因素可能对试验结果产生影响,并根据实验要求选出适当因素加以研究。例如在林地挖掘马尾松树根后,对林地水土保持功能影响的试验,除了挖掘树根这一因素外,通过认真研究还应考虑林地的坡度和降雨强度这两个因素。注意不要贪大求全,不要包罗万象,也不能过少而忽略了起重要作用的因素,在因素和水平的确定中专业知识和经验是特别重要的。

三、指标的确定

在选择试验指标时,必须考虑指标对所研究问题能提供什么信息,以及如何测定该指标。如上述间伐强度试验中,试

验指标可以是间伐前后林分平均高、平均胸径、材积的生长量或生长率。选择其中之一还是两个,选择哪一个这就要结合专业实践,反复用目标来衡量,并根据测定的难易程度多方面综合后确定。

四、实验计划的确定

实验计划的确定在整个实验设计中是至关重要的。首先须确定希望分辨出不同试验处理间的最小差异程度和允许冒多大的风险,以便决定重复数,还要考虑以怎样的方式收集数据以及怎样作随机排列。努力做到统计分析方法正确,而试验经费又比较节省。

五、试验设计的实施

试验设计的实施过程也是收集数据的过程,设计者应亲临现场,认真监督试验计划的执行。由于林业工作中许多试验实施过程时间较长,应特别注意这一环节,以使试验不至于半途而废。

六、数据分析

实验所得数据应做统计分析,如比较两个处理平均数的差异,简单地比较两个平均数是不够的,要考虑统计显著性,数据分析中应注意使用电子计算机,以减轻工作量,对可疑数据的处理要切实查找理论根据,或在统计上提出根据。

七、结论与应用

对数据分析的结果应从中归纳出有关结论,给予生物学的解释,评价这些结论的实际意义,假若一次试验不能得到明确的结论则应进一步安排试验,继续探讨。

第五节 试验设计中的小区技术

一、小区的面积和形状

小区面积大小对于减少土壤差异的影响和提高试验精度有直接关系,对不同试验方案和试验目的来说小区面积要求不同。一般来说,在一定范围内,试验误差增减与小区面积大小成反比,当小区面积增大到一定面积以后,试验误差下降就不明显。有人做过试验,以小区面积 $16M^2$ 为基础,逐渐增大小区面积,不增加重复数,试验指标的变动系数,小区面积不同对试验指标测定的影响如表 1.1。

如果小区面积不断增

表 1.1

大,那么整个试验地将占很大面积,易受土壤趋向性差异的影响,且试验工作量增大。如果当试验地面积一定,增大小区面积,可能造成重复数减少,也会降低试验精

一个小区面积	变动系数(%)
16	21.1
32	14.6
64	12.9
96	7.8
192	5.3

确性。所以试验小区的面积究竟多大为宜,这难以作硬性规定,应根据试验地大小、地形情况以及试验水平及要求而灵活确定。

二、小区形状和方向

小区形状对提高试验的精确性有一定影响。小区形状分为长方形和正方形两种。一般的情况下试验小区以长方形为宜,这是因为长方形小区较狭长,不论是呈梯度或呈斑状土壤肥力差异,采用它能全面地包括不同肥力的土壤。相应减少小

区之间的土壤差异。有人在小麦试验中曾进行小区形状的研究,分小区面积为 16m^2 、 64m^2 ,分别计算出测定指标的变异系数,小区形状与变异系数见表 1.2。

表 1.2

小区面积(m^2)	16		64	
小区形状(宽 $\times$ 长)	2 $\times$ 8	4 $\times$ 4	4 $\times$ 16	8 $\times$ 8
变异系数 (%)	21.1	23.4	12.9	14.5

从表 1.2 中可以看出,长方形的变异系数比正方形变异系数小。狭长形小区用于某些试验时,如肥料试验因其边界长,边际效应明显,可能处理效应会扩及邻区,试验效果不如方形小区好。

二、设置对照区和保护行

在林间试验设计中,应设置对照区和保护行。对照区设于每一个重复内,作为与处理比较的共同标准,还可利用对照估计和矫正试验地土壤差异。

设置保护行是为了多处理能在较均匀的环境条件下生长发育或观察对比;保护试验材料不受人畜等外来因素的损害,同时消除因试验地环境条件差异所产生的边际影响,使处理间有正确比较。

在三种情况下应设置保护行:

- (1)在试验地周围设置保护行,其宽度与小区宽度相等或稍宽。
- (2)试验的各重复较分散时,应在重复区的周围设置保护行。
- (3)小区的两端或两侧设置保护行,以消除小区之间的边际影响。

第二章 方差分析

在上一章的试验设计基本程序中所阐述的数据分析工作是其中极为重要的一环。试验设计得到结果进行数据分析主要依据数理统计的理论和原理,其中最重要的是方差分析和回归分析。由于考虑系统地学习林业试验设计基本知识的需要,因此以初步掌握概率论知识为基础,阐述了预备知识。

对于已经了解与掌握方差分析和回归分析的读者,可以直接阅读第三章之后的内容。

方差分析是 R·Fishar 于 1923 年提出的,该方法广泛应用于科学试验结果进行数据分析。林业科学试验周期长,试验面积大,试验点布设广。例如,国家重点科研攻关课题之一的“全国杉木种源试验”的周期长达十几至二十年,试验点分布我国南方十几个省区,像这样的林业科学试验应当按照试验设计的要求设计试验,用方差分析的方法,进行结果数据分析就显得特别重要。

方差分析是一类特定情况下的统计假设检验。它使影响试验指标的各主要因素造成差异得到定量描述。方差分析可以判断随机因素和人为因素(被考察的因素)对试验结果的影响。在试验中一般这两类因素通常是混杂在一起作用于试验结果的,通过方差分析就可以将这两类因素造成的差异、波动从混杂中分离开来,分别给予定量描述,从而确定因素对试验指标的影响程度。例如上述的林木种源试验,人为因素为不同种源,随机因素为土壤因素等不可知的因素。通过正确的试验

设计,这两类因素共同作用于林木的胸径、树高、材积生长量等,试验的目的是主要考察不同种源这一人为因素对林木胸径、树高、材积生长量等影响程度,要得到正确结果必须应用方差分析排除随机因素的影响。

本章将介绍单向、双向、系统分组的方差分析数学模型以及多重比较方法。

第一节 单向分组的方差分析

一、数据结构形式及两种不同处理效应

下面从一个实际的试验例子分析单向分组的方差分析的数据结构。

例 2.1 对某种树种的种子进行 5 种不同的处理方法,每种方法处理 36 粒种子并分成 6 组育苗试验。一年后观察苗高获得下表 2.1。在这个例子中种子处理方法是试验因素,5 种处理方法为 5 个水平。我们认为 5 个处理方法处理种子为 5 个总体,表 2.1 中数据是从总体中抽出几个样本,通过比较这 5 个样本,判断这 5 个总体即 5 种处理方法是否存在差异。

表 2.1

处理方法	苗高 X_{ij} (cm)						$x_{i\cdot}$	$\bar{x}_{i\cdot}$
	1	2	3	4	5	6		
1	39.2	29.0	25.8	33.5	41.7	37.2	206.4	34.4
2	37.3	27.7	23.4	34.4	29.2	35.6	186.6	31.1
3	20.8	33.8	28.6	23.4	22.7	30.9	160.2	26.7
4	31.0	27.4	19.5	29.6	23.2	18.7	149.4	24.9
5	20.7	17.6	29.4	27.7	25.5	19.5	140.4	23.4

该例为单因素实验,如果单因素有 a 个水平,也可称为 a

个处理,每一个处理所得结果都是一个随机变量 x_i ,其重复 n 次。单向分组的方差分析数据结构形式见表 2.2。

表 2.2

处理	1	2	3	...	n	平均数
x_1	x_{11}	x_{12}	x_{13}	...	x_{1n}	$\bar{x}_{1\cdot}$
x_2	x_{21}	x_{22}	x_{23}	...	x_{2n}	$\bar{x}_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	...	$\vdots$	$\vdots$
x_a	x_{a1}	x_{a2}	x_{a3}	...	x_{an}	$\bar{x}_{a\cdot}$

表 2.2 中数据 x_{ij} 表示第 i 水平的第 j 次试验的观察值。下列用“ $\cdot$ ”表示对下个下标的和,以后各章节都使用这一个符号。

$$x_{i\cdot} = \sum_{j=1}^n x_{ij} \quad (2.1)$$

$$\bar{x}_{i\cdot} = \frac{1}{n} x_{i\cdot} \quad (2.2)$$

$$x_{\cdot\cdot} = \sum_{i=1}^a \sum_{j=1}^n x_{ij} \quad (2.3)$$

$$\bar{x}_{\cdot\cdot} = \frac{1}{an} x_{\cdot\cdot} \quad (2.4)$$

单向分组的每一个观察值常用线性统计模型来描述即:

$$x_{ij} = \mu + \alpha_i + \epsilon_{ij} \quad \begin{cases} i=1, 2, \dots, a \\ j=1, 2, \dots, n \end{cases} \quad (2.5)$$

2.5 式中 x_{ij} 是第 i 水平下的第 j 次观察值。 μ 是总平均数,它是所有观察值的参量; α_i 是第 i 水平的处理效应,它仅对第 i 水平的参数; ϵ_{ij} 是随机误差成份。

模型 2.5 要求误差 ϵ_{ij} 服从正态分布 $N(0, \sigma^2)$ 的独立随机变量,同时各水平方差均为 σ^2 。模型 2.5 中有两类不同的处理

效应即固定效应和随机效应。

固定效应是由固定因素所引起的效应。该因素的 a 个水平是经过人为选择确定下来的,例如不同无性系杂种,不同种源,不同药物的对比试验。随机效应是由随机因素所引起的效应,该因素的 a 个水平是从该因素全部水平总体中随机抽取出来的。固定效应和随机效应两种模型的不同点见表 2.3。

表 2.3

项目	固定效应	随机效应
1. 效应来源	固定因素引起	随机因素引起
2. 水平确定方式	a 个水平经过特意选择	a 个水平是从全部水平总体中随机抽取
3. 结论是否适合其它水平	不适合	适合
4. 水平可控性	可控	不可控
5. 实例	温度、浓度	农家肥

二、固定效应模型

对于固定效应模型 2.5 式中 $\alpha_i (i=1, 2, \dots, a)$ 是固定的常量, α_i 是处理平均数与总平均数的离差,即

$$\sum_{i=1}^a \alpha_i = 0$$

需检验 a 个处理效应相等性,就要判断各 α_i 是否都等于 0。若各 α_i 都等于 0,则各处理效应之间无差异,因此零假设为

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_a = 0$$

方差分析就是将总的变差分解为构成总离差的各个部分:

总的离差平方和 L_T ,它表示 an 个观察值总波动

$$\begin{aligned}
L_T &= \sum_{i=1}^a \sum_{j=1}^n (x_{ij} - \bar{x}_{.j})^2 \\
&= \sum_{i=1}^a \sum_{j=1}^n x_{ij}^2 - \frac{(\sum_{i=1}^a \sum_{j=1}^n x_{ij})^2}{an} \\
&= \sum_{i=1}^a \sum_{j=1}^n x_{ij}^2 - C
\end{aligned}
\tag{2.6}$$

2.6 式中 $(\sum_{i=1}^a \sum_{j=1}^n x_{ij})^2 / an$ 通常称为校正项, 用 C 表示。

处理离差平方和 L_A , 表示处理间波动

$$\begin{aligned}
L_A &= n \sum_{i=1}^a (\bar{x}_i - \bar{x})^2 = n (\sum_{i=1}^a \bar{x}_i^2 - a\bar{x}^2) \\
&= n \sum_{i=1}^a \left[\frac{\sum_{j=1}^n x_{ij}}{n} \right]^2 - na \left[\frac{\sum_{i=1}^a \sum_{j=1}^n x_{ij}}{na} \right]^2 \\
&= \sum_{i=1}^a \frac{x_i^2}{m} - C
\end{aligned}
\tag{2.7}$$

2.7 式中 $m = na$

由于

$$\begin{aligned}
&\sum_{i=1}^a \sum_{j=1}^n (x_{ij} - \bar{x}_{.j})^2 \\
&= \sum_{i=1}^a \sum_{j=1}^n [(x_{ij} - \bar{x}_{i.}) + (\bar{x}_{i.} - \bar{x}_{.j})]^2 \\
&= \sum_{i=1}^a \sum_{j=1}^n (x_{ij} - \bar{x}_{i.})^2 + 2 \sum_{i=1}^a \sum_{j=1}^n (x_{ij} - \bar{x}_{i.})(\bar{x}_{i.} - \bar{x}_{.j}) \\
&\quad + \sum_{i=1}^a \sum_{j=1}^n (\bar{x}_{i.} - \bar{x}_{.j})^2 \\
&= \sum_{i=1}^a \sum_{j=1}^n (x_{ij} - \bar{x}_{i.})^2 + \sum_{i=1}^a \sum_{j=1}^n (\bar{x}_{i.} - \bar{x}_{.j})^2
\end{aligned}$$

即

$$L_T = L_A + L_e \quad (2.8)$$

$$L_e = \sum_{i=1}^a \sum_{j=1}^n (x_{ij} - \bar{x}_{i.})^2$$

L_A 表示处理间的波动, L_e 表示随机误差的波动, 称为误差离差平方和。

自由度也可以分割 L_T 具有 $an-1$ 的自由度, L_A 有 $a-1$ 的自由度, 对于 L_e 项的自由度为 $(an-1)-(a-1)=an-a$ 。

S_e^2 称为该差均方, 它为 L_e 除以自由度即

$$S_e^2 = \frac{L_e}{na-a} \quad (2.9)$$

为了估计 σ^2 , 先证明 S_e^2 的数学期望为 σ^2 。

$$\begin{aligned} E(S_e^2) &= E\left(\frac{L_e}{na-a}\right) = \frac{1}{na-a} E(L_e) \\ &= \frac{1}{na-a} E\left[\sum_{i=1}^a \sum_{j=1}^n (x_{ij} - \bar{x}_{i.})^2\right] \\ &= \frac{1}{na-a} E\left[\sum_{i=1}^a \sum_{j=1}^n (\mu + \alpha_i + \epsilon_{ij} - \mu - \alpha_i - \bar{\epsilon}_{i.})^2\right] \\ &= \frac{1}{na-a} E\left[\sum_{i=1}^a \sum_{j=1}^n (\epsilon_{ij} - \bar{\epsilon}_{i.})^2\right] \\ &= \frac{1}{na-a} E\left[\sum_{i=1}^a \sum_{j=1}^n (\epsilon_{ij}^2 - n \sum_{i=1}^a \bar{\epsilon}_{i.}^2)\right] \\ &= \frac{1}{na-a} (na\sigma^2 - n\sigma^2) \\ &= \frac{1}{na-a} (na\sigma^2 - a\sigma^2) \\ &= \sigma^2 \end{aligned} \quad (2.10)$$

$$E(S_A^2) = E\left(\frac{L_A}{a-1}\right)$$

$$\begin{aligned}
&= \frac{1}{a-1} E(L_A) \\
&= \frac{1}{a-1} E\left[n \sum_{i=1}^a (\bar{x}_{i.} - \bar{x}_{..})^2\right] \\
&= \frac{n}{a-1} E\left[\sum_{i=1}^a (\mu + \alpha_i + \bar{\epsilon}_{i.} - \mu - \bar{\alpha} - \bar{\epsilon}_{..})^2\right] \\
&= \frac{n}{a-1} E\left\{\sum_{i=1}^a [(\bar{\epsilon}_{i.} - \bar{\epsilon}_{..}) + (\alpha_i - \bar{\alpha})]^2\right\} \\
&= \frac{n}{a-1} E\left[\sum_{i=1}^a (\bar{\epsilon}_{i.} - \bar{\epsilon}_{..})^2 + 2 \sum_{i=1}^a (\bar{\epsilon}_{i.} - \bar{\epsilon}_{..})\right. \\
&\quad \left.(\alpha_i - \bar{\alpha}) + \sum_{i=1}^a (\alpha_i - \bar{\alpha})^2\right]
\end{aligned}$$

由于 $E(\epsilon_{ij})=0$ 因此包含 ϵ_{ij} 的乘积项的数学期望值为 0,

又因为 $\sum_{i=1}^a \alpha_i = 0$, 可知 $\bar{\alpha} = 0$ 因此

$$\begin{aligned}
E(S_A^2) &= \frac{n}{a-1} E\left[\sum_{i=1}^a (\bar{\epsilon}_{i.} - \bar{\epsilon}_{..})^2 + 2 \sum_{i=1}^a (\bar{\epsilon}_{i.} - \bar{\epsilon}_{..})\right. \\
&\quad \left.(\alpha_i - \bar{\alpha}) + \sum_{i=1}^a (\alpha_i - \bar{\alpha})^2\right] \\
&= \frac{n}{a-1} E\left[\sum_{i=1}^a (\bar{\epsilon}_{i.} - \bar{\epsilon}_{..})^2\right] + \frac{n}{a-1} \sum_{i=1}^a \alpha_i^2 \\
&= \frac{n}{a-1} E\left[\sum_{i=1}^a \bar{\epsilon}_{i.}^2 - a E(\bar{\epsilon}_{..}^2)\right] + \frac{n}{a-1} \sum_{i=1}^a \alpha_i^2 \\
&= \frac{n}{a-1} \left(\frac{a\sigma^2}{n} - \frac{\sigma^2}{n}\right) + \frac{n}{a-1} \sum_{i=1}^a \alpha_i^2 \\
&= \sigma^2 + \frac{n}{a-1} \sum_{i=1}^a \alpha_i^2 \tag{2.11}
\end{aligned}$$

从 2.10 与 2.11 中可以看出误差的均方 S_A^2 反映了随机因素所造成的方差大小, 它是 σ^2 的无偏估计量, 对于处理的

均方 S_A^2 , 只有当 $H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_a = 0$ 时 S_A^2 才是 σ^2 的无偏估计量。当 $\alpha_i = 0$ 时 2.11 式中 $\frac{n}{a-1} \sum_{i=1}^a \alpha_i^2$ 项等于 0, 这时 $E(S_A^2) = \sigma^2$, 所以用 S_A^2 与 S_r^2 比较, 就可反映出 α_i 的大小。如果 S_A^2 与 S_r^2 相差不大, 就可认为 α_i 与 0 差异不大, 如果 S_A^2 与 S_r^2 比较差异显著, 则认为 α_i 间差异显著。

我们可证明 S_A^2/S_r^2 是遵从第一自由度为 d_{fA} , 第二自由度为 d_{fr} 的 F 分布, 因此可以 F 检验来实现上述的比较。

$$F = \frac{S_A^2}{S_r^2}$$

当 $F < F_\alpha$ 时则认为 S_A^2 与 S_r^2 差异不大, $\frac{n}{a-1} \sum_{i=1}^a \alpha_i^2$ 近等于 0, 接受零假设处理间无显著差异。

当 $F > F_\alpha$ 时则认为 S_A^2 与 S_r^2 差异大, $\frac{n}{a-1} \sum_{i=1}^a \alpha_i^2$ 不为 0, 拒绝零假设处理间差异显著。

如果令

$$\eta_a^2 = \frac{1}{a-1} \sum_{i=1}^a \alpha_i^2$$

那么

$$\begin{aligned} E(S_A^2) &= \sigma^2 + \frac{n}{a-1} \sum_{i=1}^a \alpha_i^2 \\ &= \sigma^2 + n\eta_a^2 \end{aligned}$$

这时零假设也可以写为 $H_0: \eta_a^2 = 0$, 备择假设 $H_A: \eta_a^2 > 0$ 。

对于单向分组固定模型的方差分析, 其计算可用方差分析表示, 见表 2.4。

表 2.4

变差来源	离差平方和	自由度	均方	F 值	均方的期望
处理	L_A	$a-1$	S_A^2	$F = \frac{S_A^2}{S_e^2}$	$\sigma^2 + n\eta_a^2$
误差	L_e	$na-a$	S_e^2		σ^2
总和	L_T	$na-1$	—	—	—

根据上述实例 2.1, 试以 95% 的可靠性判断种子处理方法对苗木生长是否有显著影响?

先假设: 各种处理方法对苗高生长无显著影响, 计算校正数 C ,

$$C = \frac{x^2 \dots}{an} = \frac{843.0^2}{30} = 23688.3$$

再计算

$$\begin{aligned} L_T &= \sum_{i=1}^a \sum_{j=1}^n x_{ij}^2 - C \\ &= 24897.16 - 23688.3 \\ &= 1208.86 \end{aligned}$$

$$\begin{aligned} L_A &= \frac{1}{n} \sum_{i=1}^a x_{i.}^2 - C \\ &= 16(206.4^2 + 186.6^2 + 160.2^2 + 149.2^2 \\ &\quad + 140.4^2) - 23688.3 \\ &= 497.86 \end{aligned}$$

$$\begin{aligned} L_e &= L_T - L_A \\ &= 1208.86 - 497.86 \\ &= 711.00 \end{aligned}$$

列出方差分析见表 2.5

表 2.5

变差来源	离差平方和	自由度	均方	F 值	F_{α}
处理	497.86	4	124.47	4.38**	
误差	711.00	25	28.44		$F_{4,25}^{0.01}=4.18$
总和	1208.86	29	—	—	

因 $F > F_{\alpha}$, 故推翻假设, 即认为不同处理方法对苗木生长无显著影响, 习惯上用“*”表示在 $\alpha=0.05$ 水平上差异显著, 用“**”表示在 $\alpha=0.01$ 水平上差异显著, 常常称为极显著差异。

三、随机效应模型

如果参加实验的 a 个水平是从该因素总体中随机选出来, 那么这因素为随机因素。通过试验设计取得的数据进行方差分析, 实质上是对 a 个水平测验, 从而推断该因素的水平总体。如从许多杉木无性系中随机抽取 a 个无性系构成样本, 通过资料分析去检验整个杉木无性系高生长变异是否显著; 从全部试验点中随机抽取 a 全试验点, 每个试验点重复若干次试验构成样本资料, 由此去分析所有试验点环境条件差异是否显著, 这些都属于随机模型处理问题。对随机模型作主差分析的目的不是选择最优处理水平, 而是确定差异的程度, 由 a 个水平推断出这个因素的所有水平。

随机效应模型为

$$x_{ij} = \mu + \alpha_i + \epsilon_{ij} \quad \begin{cases} i=1, 2, \dots, a \\ j=1, 2, \dots, n \end{cases} \quad (2.13)$$

其中 α_i 和 ϵ_{ij} 都是随机变量, 如果 α_i 具有方差 σ_{α}^2 并独立于 ϵ_{ij} , 则

$$V_{\alpha\beta}(x_{ij}) = \sigma_a^2 + \sigma^2 \quad (2.14)$$

方差 σ_a^2 和 σ^2 称为方差分量, 模型 2.13 称为随机效应模型。用 $N(0, \sigma^2)$ 表示独立的正态分布。在模型中要求 ϵ_{ij} 为遵从 $N(0, \sigma^2)$ 的随机主量, α_i 为遵从 $N(0, \sigma_a^2)$ 的随机变量。

在随机效应模型中, 对多个处理效应的检验并非目的, 所要检验的是 α_i 变异性, 因此假设:

$$H_0: \sigma_a^2 = 0$$

$$H_A: \sigma_a^2 > 0$$

如果接受 $H_0: \sigma_a^2 = 0$ 认为处理之间没有差异; 如果拒绝 $H_0: \sigma_a^2 = 0$ 认为处理间存在差异。

对于随机效应模型下列两个式子仍然成立

$$L_T = L_A + L_e$$

$$df_T = df_A + df_e$$

下列推导各个均方的数学期望

$$\begin{aligned} E(S_A^2) &= \frac{1}{a-1} E(L_A) \\ &= \frac{n}{a-1} E\left[\sum_{i=1}^a \bar{x}_{i.} - \bar{x}_{..}\right]^2 \\ &= \frac{n}{a-1} E\left[\sum_{i=1}^a (\bar{\epsilon}_{i.} - \bar{\epsilon}_{..})^2 + \sum_{i=1}^a (\alpha_i - \bar{\alpha})^2\right] \\ &= \frac{n}{a-1} E\left[\sum_{i=1}^a \bar{\epsilon}_{i.}^2 - a\bar{\epsilon}_{..}^2\right] + \frac{n}{a-1} E\left(\sum_{i=1}^a \alpha_i^2 - \bar{\alpha}^2\right) \\ &= \frac{n}{a-1} \left(\frac{a\sigma^2}{n} - \frac{\sigma^2}{n}\right) + \frac{n}{a-1} (a\sigma_a^2 - \sigma_a^2) \\ &= \sigma^2 + n\sigma_a^2 \end{aligned} \quad (2.15)$$

同理可证明

$$E(S_e^2) = \sigma^2$$

仍用统计学知识可以证明

$$\frac{S_A^2}{S_e^2} \sim F_a(df_a, df_e)$$

虽然二种模型不同,但都可以通过计算 $F = S_A^2/S_e^2$ 来与 F 分布临界值 F_a 比较后作出推断。

在林木遗传育种试验中,用随机模型中各种不同方差分量来表达遗传育种中各种变异, S_A^2 的期望均方分别为 σ^2 和 $n\sigma_a^2$, σ_a^2 是品种间 2 传型差异,称为遗传方差, σ^2 为随机的环境条件影响而产生变异,称为环境方差。遗传方差 σ_a^2 加上环境方差 σ^2 组成表现型方差 σ_P^2

$$\sigma_P^2 = \sigma_a^2 + \sigma^2 \quad (2.16)$$

广义遗传力 η^2 定义为

$$\eta^2 = \frac{\sigma_a^2}{\sigma_P^2} = \frac{\sigma_a^2}{\sigma_a^2 + \sigma^2} \quad (2.17)$$

由于 $E(S_e^2) = \sigma^2$, $E(S_A^2) = n\sigma_a^2 + \sigma^2$

所以

$$\hat{\sigma}^2 = S_e^2, \quad \hat{\sigma}_a^2 = \frac{S_A^2 - S_e^2}{n}$$

由此得到广义遗传力估计值为

$$\begin{aligned} \hat{\eta}^2 &= \frac{\frac{S_A^2 - S_e^2}{n}}{\frac{S_A^2 - S_e^2}{n} + S_e^2} \\ &= \frac{S_A^2 - S_e^2}{S_A^2 + (n-1)S_e^2} \end{aligned} \quad (2.18)$$

例 2.2 8 种杉木无性系品种是从许多无性系品种中随机选出来的,每个无性系重复 4 次,造林 5 年后调查树高生长量,得每小区树高平均生长量见下表 2.6,试问整个杉木的无性系间高生长遗传变异是否显著? 并计算杉木无性系的广义遗传力。

因为本例中的 8 个无性系品种是从许多无性系品种随机抽取而组成的样本, 计算后去推断总体变异程度, 因此采用随机效应模型。

表 2.6

品种	每小区平均生长量				$x_{i.}$	$\bar{x}_i$
A	3.0	2.9	2.9	3.1	11.9	2.975
B	3.2	3.3	3.3	3.5	13.3	3.325
C	3.6	3.8	3.3	3.5	14.2	3.550
D	3.7	3.7	3.7	3.8	14.9	3.725
E	2.2	3.0	2.5	2.6	10.3	2.575
F	4.0	3.7	3.5	3.8	15.0	3.750
G	3.7	3.6	3.6	3.5	14.4	3.600
H	3.2	3.2	3.0	3.3	12.7	3.175
					$x_{..} = 106.70$	

统计假设 $H_0: \sigma_a^2 = 0$ 即高生长遗传变异不显著, 计算离差平方和和自由度。

$$C = \frac{x_{..}^2}{an} = \frac{106.7^2}{8 \times 4} = 355.7778$$

$$L_T = \sum_i \sum_j x_{ij}^2 - C = 361.21 - 355.7778 = 5.4322$$

$$L_A = \frac{1}{a} \sum_j x_{i.}^2 - C = \frac{1}{4} (11.9^2 + \dots + 12.7^2) - C = 4.69$$

$$L_e = L_T - L_A = 0.7422$$

$$df_T = an - 1 = 31$$

$$df_A = a - 1 = 7$$

$$df_e = 24$$

列出方差分析表如表 2.7。

表 2.7

离差来源	离差平方和	自由度	均方	F 值	F_{α}
无性系间	4.69	7	0.67	21.68	$F_{0.05}(7, 24)=2.42$
无性系内	0.7422	24	0.0309		$F_{0.01}(7, 24)=3.5$
总和	5.4322	31			

由于 $F > F_{0.01}(7, 24)$, 推断假设认为无性系间高生长变异极端显著。

广义遗传力 η^2 的估计值为:

$$\begin{aligned}\eta^2 &= \frac{S_A^2 - S_e^2}{S_A^2 + (n-1)S_e^2} \\ &= \frac{0.67 - 0.0309}{0.67 + (4-1) \times 0.0309} \\ &= 0.84\end{aligned}$$

认为无性系间的遗传变异在表型变异中约占 84%。

四、试验次数不等时单向分组的方差分析

上面所阐述的单向分组方差分析的数据结构为各处理试验次数或观察次数都是相同的情况。在林业试验中也出现不同处理的观察次数不等情况,对于这种情况上述方差分析方法仍然适用,只是在计算离差平均和时有所不同。

假设第 i 处理做了 n_i 次观察 ($i=1, 2, \dots, a$), 总观察次数

$$N = \sum_{i=1}^a n_i$$

我们作上述相似离差平方和分解,也有

$$\begin{aligned}& \sum_{i=1}^a \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{..})^2 \\ &= \sum_{i=1}^a \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{i.})^2 + \sum_{i=1}^a \sum_{j=1}^{n_i} (\bar{x}_{i.} - \bar{x}_{..})^2\end{aligned}$$

$$= \sum_{i=1}^a \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{.})^2 + \sum_{i=1}^a n_i (\bar{x}_{i.} - \bar{x}_{.})^2 \quad (2.19)$$

$$\text{即 } L_T = L_A + L_e$$

$$\begin{aligned} L_T &= \sum_{i=1}^a \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{.})^2 \\ &= \sum_{i=1}^a \sum_{j=1}^{n_i} x_{ij}^2 - N \bar{x}_{.}^2 \\ &= \sum_{i=1}^a \sum_{j=1}^{n_i} x_{ij}^2 - C \end{aligned} \quad (2.20)$$

$$\begin{aligned} L_A &= \sum_{i=1}^a n_i (\bar{x}_{i.} - \bar{x}_{.})^2 \\ &= \sum_{i=1}^a \frac{x_{i.}^2}{n_i} - C \end{aligned} \quad (2.21)$$

$$L_e = L_T - L_A \quad (2.22)$$

在假设 $H_0: \alpha_1 = \alpha_2 = \dots = \alpha_a$ 时, 依统计学有关知识一样可以证明:

$$\frac{S_A^2}{S_e^2} = \frac{L_A/(a-1)}{L_e/(N-a)} \sim F_{a-1, N-a} \quad (2.23)$$

根据 F 值与 $F_{a-1, N-a}$ 的临界值比较对假设 H_0 作出判断。

例 2.3 国营林场的森林资源清查工作常用分层抽样的方法, 效率较高。依据什么因素进行合理分层是提高分层抽样效率的技术问题。如何检验其分层的合理性常用单向分组, 试验次数不等的方差分析方法。如某林场的森林资源抽样调查, 按其蓄积量分为三层, 按层面积数量分别于各层中用重复抽样方式抽取 1000 平方米的样地 8 块、13 块、9 块。每木检尺后得各样地蓄积, 见表 2.8, 试问不同层单位面积的蓄积量之间

是否存在差异?

表 2.8

层号	$x_{ij}(M^2/0.1h_p)$								$x_{i.}$
1	0.60	1.20	5.90	7.10	8.20	0.60	7.10	8.80	39.5
2	1.61	21.2	15.3	13.0	13.5	14.1	15.9	11.8	198.6
	15.3	18.8	12.4	17.1	14.1				
3	20.0	26.6	26.5	30.6	21.8	20.6	17.1	21.8	212.0

假设 $H_0: \sigma_a^2=0$

计算离差平方和及自由度: L

$$C = \frac{X_{..}^2}{N} = \frac{450.1^2}{8+13+9} = 6.753$$

$$L_T = \sum_i \sum_j x_{ij}^2 - C = 8542.35 - 6753 = 1789.35$$

$$L_A = \sum_i \frac{x_{i.}^2}{n_i} - C = \frac{39.5^2}{8} + \frac{198.6^2}{13} + \frac{212.0^2}{9} - C \\ = 1469.81$$

$$L_e = L_T - L_A = 319.54$$

相应各项自由度分别为:

$$df_T = N - 1 = 29$$

$$df_A = a - 1 = 2$$

$$df_e = df_T - df_A = 29 - 2 = 27$$

列出方差分析表如表 2.9

表 2.9

变差来源	离差平方和	自由度	均方	F 值	F_{α}
层间	1469.81	2	743.91	62.096	$F_{2,27}^{0.01} = 3.35$
层内	319.54	27	11.835		
总和	1789.35	29	--	--	--

由于 $F = 62.096 > F_{2,27}^{0.01} = 3.35$ 推翻假设,认为层间单位面积蓄积量差异极端显著,这样的分层结果会大大提高分层抽样的效率。

五、方差分析的几点基本条件

1. 可加性:每个处理效应与误差效应是可加的,即

$$x_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

α_i 为处理效应, ϵ_{ij} 为误差效应。有了这一假设条件,不同效应才能被分解和最终判断处理效应是否比误差效应更显著。

2. 正态性:总体的被观察的标志遵从或近似正态分布 $N(\mu, \sigma)$ 。根据线性统计模型和可加性的假设可得到实验误差也应当服从正态分布 $N(0, \sigma^2)$ 的独立随机变量应具备的条件。

3. 独立性:抽样为重复抽样或抽样比很小的非重复抽样,以保证样本各单元互不影响,即要求每次进行的随机试验是相互独立的。

4. 方差齐性:各处理的误差应具备齐性,并具有一个公共总体方差 σ^2 ,即

$$\sigma_1^2 = \sigma_2^2 = \dots = \sigma_s^2 = \sigma^2$$

对于林业科研试验来说,这 4 个应具备的条件在实践中常常可以得到满足。有些不具备条件也可以通过数据变换而实现。方差齐性一般在分析前作检验,具体方法在下节介绍。

六、多个方差的齐性检验

在上节中强调方差分析的 4 条应具备的条件,对于方差齐性的检验,就是各处理的总体方差一致的检验。如果多个方差的齐性检验仍用二个方差齐性的检验方法,则需要进行 C_s^2 次检验,这种方法比较麻烦。对于多个方差的齐性检验,用得

最广泛的是 Bartlett 检验法。

Bartlett 检验法基本原理是：先作假设，假设 a 个总体的总体方差相同，即：

$$H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_a^2$$

$$H_A: \text{至少有两个 } \sigma_i^2 \text{ 不等。}$$

计算样本平均数，样本方差，样本方差的加权平均数。

$$\bar{x}_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$$

$$S_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{i.})^2$$

$$\bar{S}^2 = \frac{\sum_{i=1}^a S_i^2 (n_i - 1)}{\sum_{i=1}^a (n_i - 1)}$$

如令：

$$C = 1 + \frac{1}{3(a-1)} \left[\sum_{i=1}^a \frac{1}{n_i - 1} - \frac{1}{\sum_{i=1}^a (n_i - 1)} \right]$$

那么有

$$X^2 = \frac{2.3026}{C} \left[\lg \bar{S}^2 \sum_{i=1}^a (n_i - 1) - \sum_{i=1}^a (n_i - 1) \lg S_i^2 \right] \quad (2.24)$$

给定显著水平 α 及 $df = a - 1$ 查 X^2 分布表得 $X_{\alpha}^2(a-1)$ 。当 $X^2 > X_{\alpha}^2$ 时推翻假设；如 $X^2 \leq X_{\alpha}^2$ 时，接受假设认为多个方差齐性的。

可以证明，当满足总体分布为正态分布时，Bartlett 检验的精确度是很高的。在正态分布的假设不能满足时，不能使用 Bartlett 检验。

例 2.4 在某树种的抗压强的试验中,以同样加荷速度测定了几种不同比重的该树种木材抗压强度,得表 2.10 中数据,判断这 4 个样本所得出的总体方差是否相同?

表 2.10

比重	抗压强度				S_i^2
0.41—0.27	340	314	335	338	144.25
0.47—0.53	490	475	471		66.34
0.53—0.59	482	504	495		39.00
0.59—0.65	985	570	595		158.33

解:假设 $H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$

具体计算可以列表进行,见表 2.11

表 2.11

样本号	n_i	$n_i - 1$	s_i^2	$\lg s_i^2$	$(n_i - 1)s_i^2$	$(n_i - 1)\lg s_i^2$	$\frac{1}{n_i - 1}$
1	4	3	144.25	2.1591	432.75	6.4773	0.33
2	3	2	66.34	1.8218	132.68	3.6436	0.50
3	3	2	39.00	1.5911	78.00	3.1822	0.50
4	3	2	158.33	2.1996	316.66	4.3992	0.50
Σ	13	9	—	—	960.09	17.7023	1.83

依表 2.11 可计算得

$$\bar{S}^2 = \frac{\sum_{i=1}^4 (n_i - 1) S_i^2}{\sum_{i=1}^4 (n_i - 1)} = \frac{960.09}{9} = 106.68$$

$$\lg \bar{S}^2 = \lg 106.68 = 2.0281$$

$$\begin{aligned}
 C &= 1 + \frac{1}{3(a-1)} \left[\sum_{i=1}^4 \frac{1}{n_i-1} - \frac{1}{\sum_{i=1}^4 (n_i-1)} \right] \\
 &= 1 + \frac{1}{3(4-1)} \left[1.83 - \frac{1}{9} \right] \\
 &= 1.19
 \end{aligned}$$

计算统计量

$$\begin{aligned}
 X^2 &= \frac{2.306}{C} \left[\lg \bar{S}^2 \sum_{i=1}^4 (n_i-1) - \sum_{i=1}^4 (n_i-1) \lg S_i^2 \right] \\
 &= \frac{2.3026}{1.19} [2.0281 \times 9 - 17.7023] \\
 &= 1.065
 \end{aligned}$$

根据 $\alpha = 0.05$, 自由度 $df = (a-1) = 3$ 查 X^2 分布表得 $X_{0.05,3}^2 = 7.815$, 因为 $X^2 < X_{0.05,3}^2$ 所以认为这 4 个总体方差齐性。

多个方差齐性检验除了上面介绍的 Bartlett 检验外, 还可以采用 Cochran 方差齐性检验方法。

Cochran 方差齐性检验方法:

如果怀疑 a 个方差中最大方差 $S_{\max}^2$ 特异, 可计算统计量

$$G_{\max} = \frac{S_{\max}^2}{S_1^2 + S_2^2 + \dots + S_a^2}$$

同时根据显著水平, 查 Cochran 检验的临界值, 当 G 大于表中临界值时认为这一组 a 个方差不齐性, 否则认为这 a 个方差齐性。

七、多重比较

试验结果统计分析, 常采用方差分析。当方差分析得到结论认为某因素各水平之间差异显著时, 一般地说不能断言各水平双双之间差异均显著, 某些水平间十分显著的差异往往可以掩盖某些水平间差异不显著, 因为其结论显著是 a 个水

平组间总变异造成的。

假设水平数为 a , 在方差分析 F 检验显著的条件下, 分析出究竟哪两个水平间差异显著, 哪两个水平之间差异不显著, 这种比较的方法称为多重比较。

读者可能知道要检验各个平均数间差异是否显著, 故然可用两个平均数的 t 检验方法, 但采用 t 检验方法工作量大, 如 $a=5$ 时双双平均数差异检验就得进行 $C_5^2=10$ 次, 十分繁琐。不但如此而且还会增大犯 I 型错误的概率, $a=5$ 时, 犯 I 型错误的概率为 $1-(1-\alpha)^5$, 当给定 $\alpha=0.05$ 时, 采用此法这种概率将为 0.226。在林木种源的试验中, 供试验的种源多达上百人, 只要供试种源为 20 个, 犯 I 型错误的概率为 $1-(1-0.05)^{20}=0.64$, 因此采用 t 检验是十分不可取的。

50 年代以来许多统计学家提出许多不同的检验方法, 来解决多重比较问题, Tukey(1953), Duncan(1955) 等针对上述的缺陷各自提出检验方法。

下面介绍四种方法:

1. 最小显著差数法

此法也称为 LSD 法 (least dignificant difference)。它的计算原理是利用两总体平均数差异显著性的小样本检验法。

对两组平均数差数差异显著性检验, 可以用成组数据的。

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{(\bar{X}_1 - \bar{X}_2)}} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad (2.25)$$

其中当 $n_1 = n_2 = n$ 时, 在上面已证明 $E(S^2) = \sigma^2$, 故以 S^2 代替 σ_1^2, σ_2^2 有

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{2S_e^2}{n}}} \sim t(an - n)$$

所以在显著水平 $\alpha=0.05, \alpha=0.01$ 下分别两组平均数的 LSD 值为

$$LSD_{0.05} = t_{0.05} \sqrt{\frac{2S_e^2}{n}}$$

$$LSD_{0.01} = t_{0.01} \sqrt{\frac{2S_e^2}{n}}$$

作为水平间双双比较的最小显著差数。

如果令 $\bar{x}_i, \bar{x}_j (i, j=1, 2, \dots, a, i \neq j)$ 表示两个样本的平均数。

当 $|\bar{x}_i - \bar{x}_j| < LSD_\alpha$ 时, 认为第 i, j 二个水平间差异不显著, 否则差异显著。

例 2.5 在例 2.1 中已得到 5 种子处理后, 一年生苗高的方差分析, 认为差异显著。现在用 LSD 法进行多重比较。

从表 2.1 和表 2.5 知 $S_e^2 = 28.44, n=6, a=5, df_e=25$

如果 $\alpha=0.05$ 和 0.01 查 t 分布表可得到

$$t_{0.05} = 2.060 \quad t_{0.01} = 2.787$$

所以

$$\begin{aligned} LSD_{0.05} &= t_{0.05} \sqrt{\frac{2S_e^2}{n}} \\ &= 2.060 \sqrt{\frac{2 \times 28.44}{6}} \\ &= 6.34 \\ LSD_{0.01} &= t_{0.01} \sqrt{\frac{2S_e^2}{n}} \end{aligned}$$

$$= 2.787 \sqrt{\frac{2 \times 28.44}{6}}$$

$$= 8.58$$

列表检验

表 2.12

处理号	平均苗高	$\bar{x}_1 - \bar{x}_5$	$\bar{x}_1 - \bar{x}_4$	$\bar{x}_1 - \bar{x}_3$	$\bar{x}_1 - \bar{x}_2$
1	34.4	11.0**	9.5**	7.7*	3.3
2	31.1	7.7*	6.2	4.4	
3	26.7	3.3	1.8		
4	24.9	1.5			
5	23.4				

从表 2.12 中看出处理方法 1 与处理方法 4、处理方法 5 的显著差异；处理方法 1 和处理方法 3 以及处理方法 2 与处理方法 5 的显著差异。

该方法利用方差分析中误差均方 S_e^2 作为公共方差，以误差的自由度 df_e 为公共方差的自由度，依上述公式计算出一个公共 LSD 值作为同一尺度去衡量两两水平间的差异，工作量较省，但它没有解决犯 I 型错误概率增大的问题。

2. Duncan 多范围检验

Duncan 多范围检验在确定各平均数对之间真正的差异时，Duncan 多范围检验较好。

检验方法：

先将需要比较的 a 个平均数依次排列好，如果

$$\bar{x}_1 > \bar{x}_2 > \dots > \bar{x}_a$$

并将每一对 $\bar{x}$ 之间差数排成梯形表，如表 2.13。

Duncan 检验中，不同对的平均数的差有不同的临界值比较，而作出推断如果临界值为 $R_k (k=2, 3, 4, \dots, a)$

$$R_k = r_k S_{\bar{x}} \quad (2.26)$$

$$=r_a \sqrt{\frac{S_e^2}{n}}$$

其中 $r_a = r_a(K, df_e)$ 的值从多重比较中 Duncan 表中查出, df_e 为误差项的自由度, K 为相比较两个平均数之间所包含的平均数的个数(包括相比较的两个平均数的个数)。

如果 $\bar{x}_j - \bar{x}_i > R_k$ (这里 $j - i + 1 = k$), 认为 i, j 处理之间差异显著, 否则认为它们之间差异不显著。

例如上例说用 Duncan 多范围检验法作多重比较。

先计算

$$S_{\bar{x}} = \sqrt{\frac{S_e^2}{6}} = \sqrt{\frac{28.44}{6}}$$

$$= 2.18$$

$$= 6.69$$

$$R_4 = r_{0.01}(4, 25) S_{\bar{x}}$$

$$= 4.25 \times 2.18$$

$$R_2 = r_{0.01}(2, 25) S_{\bar{x}}$$

$$= 3.96 \times 2.18$$

$$= 8.63$$

$$= 9.27$$

$$R'_4 = r_{0.05}(4, 25) S_{\bar{x}}$$

$$= 3.15 \times 2.18$$

$$R'_2 = r_{0.05}(2, 25) S_{\bar{x}}$$

$$= 2.92 \times 2.18$$

$$= 6.37$$

$$= 6.87$$

$$R_5 = r_{0.01}(5, 25) S_{\bar{x}}$$

$$= 4.31 \times 2.18$$

$$R_3 = r_{0.01}(3, 25) S_{\bar{x}}$$

$$= 4.14 \times 2.18$$

$$= 9.03$$

$$= 9.40$$

$$R_5^1 = r_{0.05}(5, 25) S_{\bar{x}}$$

$$= 3.23 \times 2.18$$

$$R'_3 = r_{0.05}(3, 25) S_{\bar{x}}$$

$$= 3.07 \times 2.18$$

$$= 7.04$$

列表检验

表 2.13

处理号	平均苗高	$\bar{x}_i - \bar{x}_5$	$\bar{x}_i - \bar{x}_4$	$\bar{x}_i - \bar{x}_3$	$\bar{x}_i - \bar{x}_2$
1	34.4	11.0**	9.5**	7.7*	3.3
2	31.1	7.7*	6.2	4.4	
3	26.7	3.3	1.8		
4	24.9	1.5			
5	23.4				

应用 Duncan 多范围检验法得到的结论与用 LSD 法一样。

3. Q 检验法

Q 检验法适用于各组内试验次数相同的情况。与 Duncan 检验法相似, Q 检验法也是多重极差检验法, 其具体步骤为:

(1) 根据规定的 α 值及组数 a , 组内方差的自由度, 查多重比较中 q 表得 $q_\alpha(a, df_e)$ 值。

(2) 计算 D 值

$$D = q_\alpha(a, df_e) \sqrt{\frac{S_e^2}{n}} \quad (2.27)$$

式中 S_e^2 为组内均方, n 为各组内单元数

(3) 当 $|\bar{x}_i - \bar{x}_j| > D$ 时判断 $\bar{x}_i$ 与 $\bar{x}_j$ 差异显著

当 $|\bar{x}_i - \bar{x}_j| < D$ 时判断 $\bar{x}_i$ 与 $\bar{x}_j$ 差异不显著。

由于 $q_\alpha(a, df_e)$ 不仅取决于 α 和 S_e^2 相应的自由度 $a(n-1)$, 而且还与组数 a 有关, a 愈大, q 愈大, 从而保证所有比较时, 犯 I 类错误的概率不至增大。一般来说, 当 a 较大时采用该方法更为合理, 例如有 k 个平均数比较时, 从 q 表查出其显著 q_α , 计算 D 值, 但当逐步测验 $(k-1)$ 平均数, $(k-2)$, ..., 2 个平均数时, 则应从其相应处理数目以查出 q_α 值, 规定出 D 值, 这样做则必须根据不同比较而有不同显著水准。

例 2.2 中数据及方差分析结果,用 q 检验法对 8 个杉木无性系进行多重比较。

解: 由于 $\alpha=0.05, df_e=21, a=2, 3, 4, 5, 6, 7, 8$, 分别查“多重比较中的 q 表”得 $q_\alpha(a, df_e)$, 进而计算 D 值, 具体见下表 2.14

表 2.14

a	2	3	4	5	6	7	8
$q_{0.05}(a, 21)$	2.94	3.57	3.95	4.22	4.43	4.60	4.75
D	0.26	0.31	0.35	0.37	0.39	0.41	0.42

根据处理平均数的排秩次, 求平均数差值, 并与表 2.14 中 D 值比较判断其显著性, 具体见表 2.15。

表 2.15

品种	平均值	$\bar{x}_1 - \bar{x}_j$	$\bar{x}_2 - \bar{x}_j$	$\bar{x}_3 - \bar{x}_j$	$\bar{x}_4 - \bar{x}_j$	$\bar{x}_5 - \bar{x}_j$	$\bar{x}_6 - \bar{x}_j$	$\bar{x}_7 - \bar{x}_j$
F	3.75	0						
D	3.73	0.02	0					
G	3.60	0.15	0.13	0				
C	3.55	0.20	0.18	0.05	0			
B	3.33	0.40*	0.40*	0.27	0.22	0		
H	3.18	0.57*	0.55*	0.42*	0.37*	0.15	0	
A	2.98	0.77*	0.75*	0.62*	0.57*	0.35*	0.20	0
E	2.58	1.17*	1.15*	1.02*	0.97*	0.75*	0.60*	0.40*

上述介绍的 Q 检验, 也称为 Newman-Keuls 检验, 对于 Q 检验也可以采用固定的 a , 计算出 D 值对所有平均数差值进行显著测验, 此法也称为 HSD 法, 由 Tukey 提出的, 使用 HSD 法工作量小。

4. S 检验

S 检验是假设试验因素只有 a 个水平,各水平分别作 n_i 次试验($i=1,2,\cdots,a$)方差分析结果试验处理间显著时,如何对各水平间是否存在的显著差异作出判断的方法,这里 $n_i \neq n_j$ ($i, j=1,2,\cdots,a; i \neq j$)。可以用 D_{ij} 尺度来测量。

检验步骤为:

(1)根据规定 α 值以及第一自由度 $df_1=a-1$ (组间自由度),第二自由度 $df_2=n-a$ (组内自由度),查 F 分布表,得 F_{α, df_1, df_2} 临界值。

(2)分别计算求出 D_{ij} 值

$$D_{ij} = \sqrt{\left(\frac{1}{n_i} + \frac{1}{n_j}\right)(a-1)S_e^2 F(a-1, n-a)}$$

其中 n_i, n_j 分别表示第 i 组、第 j 组的试验次数, S_e^2 表示误差均方。

(3)当 $|\bar{x}_i - \bar{x}_j| > D_{ij}$ 时判断 $\bar{x}_i$ 与 $\bar{x}_j$ 差异显著,

$|\bar{x}_i - \bar{x}_j| < D_{ij}$ 时判断 $\bar{x}_i$ 与 $\bar{x}_j$ 差异不显著。

在例 2.3 中用分层抽样方法进行森林调查,分为三层,计算结果得:

$$\bar{x}_1 = \frac{39.5}{8} = 4.99, \quad \bar{x}_2 = \frac{198.6}{13}, \quad \bar{x}_3 = \frac{212.0}{9} = 23.56$$

以及 $df_1=a-1=2, df_2=n-a=27, S_e^2=11.84$

如果 $\alpha=0.05$,查 F 分布表得 $F_{0.05}(2, 27)=3.55$ 据此计算

$$\begin{aligned} D_{12} &= \sqrt{\left(\frac{1}{8} + \frac{1}{13}\right) \times 2 \times 11.84 \times 3.55} \\ &= \sqrt{\left(\frac{1}{8} + \frac{1}{13}\right) \times 79.328} = 4.00 \end{aligned}$$

$$\begin{aligned}
D_{13} &= \sqrt{\left(\frac{1}{8} + \frac{1}{9}\right) \times 2 \times 11.84 \times 3.55} \\
&= \sqrt{\left(\frac{1}{8} + \frac{1}{9}\right) \times 79.328} = 4.33 \\
D_{23} &= \sqrt{\left(\frac{1}{13} + \frac{1}{9}\right) \times 2 \times 11.84 \times 3.55} \\
&= \sqrt{\left(\frac{1}{13} + \frac{1}{9}\right) \times 79.328} = 3.86
\end{aligned}$$

所以比较结果为

$$|\bar{x}_1 - \bar{x}_2| = 10.29 > D_{12} = 4.00$$

$$|\bar{x}_1 - \bar{x}_3| = 18.57 > D_{13} = 4.33$$

$$|\bar{x}_2 - \bar{x}_3| = 8.28 > D_{23} = 3.86$$

因此按蓄积量分成三层,双双之间均有显著差异,分层是有效的。

第二节 两向分组的方差分析

在林业试验中,经常遇到两个或两个以上的因素同时影响某项试验结果。例如在考虑杉木定向培育的试验时,一般可以认为造林(经营)密度,经营措施是影响杉木定向培育的两个主要因素。两向分组试验也称为两因素交叉分组试验,其结果分析应采用两向分组的方差分析。如果研究总体按 A 、 B 两个因素分别分为 a 个水平组和 b 个水平组,通过方差分析可以一次确定两个因素对试验指标有无显著影响,并能对两个因素的联合影响是否显著作出判断。

A 、 B 因素分别分为 a 、 b 水平,共有 $a \times b$ 个交叉组。如果

在两因素实验中,当两个因素都是固定因素时,称为固定模型;两个因素均为随机因素时称为随机模型,一个因素是固定模型,另一个因素是随机因素时,称为混合模型。三种模型以计算方法而论没有多大不同,但是在检验以及对结果的解释上都不同。尤其存在两个因素的联合作用其区别更为明显。

一、主效应和交互作用

由于因素水平的改变而造成因素效应的改变,称为该因素的主效应。例如,一片毛竹用材林中设置两个因素的试验, A 因素为立竹度,分为两个水平 A_1 、 A_2 , B 因素为施钾肥,分为两个水平 B_1 、 B_2 ,试验结果的观察指标为新竹重量。

	A_1	A_2 (kg)
B_1	200	260
B_2	400	460

当因素从第一水平变化到第二水平时, A 因素的主效应(记为 A)为 A_2 水平的平均效应减小 A_1 水平的平均效应。

$$\begin{aligned}
 A &= \frac{A_2B_1 + A_2B_2}{2} - \frac{A_1B_1 + A_1B_2}{2} \\
 &= \frac{260 + 460}{2} - \frac{200 + 400}{2} \\
 &= 360 - 300 \\
 &= 60
 \end{aligned}$$

同样

$$\begin{aligned}
 B &= \frac{A_1B_2 + A_2B_2}{2} - \frac{A_1B_1 + A_2B_1}{2} \\
 &= \frac{400 + 460}{2} - \frac{200 + 260}{2} \\
 &= 430 - 230
 \end{aligned}$$

$$=200$$

从上述可以看出 B 的主效应比 A 的主效应大, 还可以看到, 当因素水平为 A_2B_2 时, 产量为 460kg, 它比 A_1B_1 的产量增加 $460-200=260(\text{kg})$ 。这是由于 A_1 改变为 A_2 时主效应, B_1 改变为 B_2 时主效应, 以及它们的联合作用, 如果联合作用为零, 那么有

$$\begin{aligned} A+B &= \frac{A_2B_1+A_2B_2}{2} - \frac{A_1B_1+A_1B_2}{2} + \frac{A_1B_2+A_2B_2}{2} \\ &\quad - \frac{A_1B_1+A_2B_1}{2} \\ &= \frac{A_2B_1}{2} + \frac{A_2B_2}{2} - \frac{A_1B_1}{2} - \frac{A_1B_2}{2} + \frac{A_1B_2}{2} \\ &\quad + \frac{A_2B_2}{2} - \frac{A_1B_1}{2} - \frac{A_2B_1}{2} \\ &= A_2B_2 - A_1B_1 \end{aligned}$$

有时某一因素在别一个因素的不同水平上所产生的效应不同, 例如进行上述同样试验, 得到下列结果, 见表 2.16。

表 2.16

	A_1	A_2
B_1	200	240
B_2	380	500

$$A(\text{在 } B_1 \text{ 的水平上}) = A_2B_1 - A_1B_1 = 240 - 200 = 40$$

$$B(\text{在 } B_2 \text{ 的水平上}) = A_2B_2 - A_1B_2 = 500 - 380 = 120$$

可以看出 A 的效应依 B 因素水平不同, 则可以说 A 因素和 B 因素间存在交互作用, 其作用大小的估计值为

$$A_1B_1 + A_2B_2 - A_1B_2 - A_2B_1$$

如果用 AB 表示 A 因素和 B 因素的交互作用,那么

$$AB = (200 + 500) - (240 + 380) = 80(\text{kg})$$

在林业生产实践中交互作用的情形普遍存在,例如杂种优势,好的林木种子在特定环境下高、径、材积的生长等,都是交互作用的结果。如何判断交互作用的存在,需要进行方差分析后从结果中得出结论。

两向分组的实验设计方法,在实践中也是极为重要的,该 A 因素划分为 a 个水平: $A_1, A_2, \dots, A_a$, B 因素划分为 b 个水平, $B_1, B_2, \dots, B_b$ 。每个交叉组重复 m 次试验,试验数据结果见表 2.17。

表 2.17

$A_i \backslash B_j$	B_1	B_2	...	B_b	$T_{i\cdot}$	$\bar{x}_{i\cdot}$
A_1	x_{111}	x_{121}	...	x_{1b1}	$T_{1\cdot}$	$\bar{x}_{1\cdot}$
	x_{112}	x_{122}	...	x_{1b2}		
	$\vdots$	$\vdots$	$\vdots$	$\vdots$		
	x_{11m}	x_{12m}	...	x_{1bm}		
A_2	x_{211}	x_{221}	...	x_{2b1}	$T_{2\cdot}$	$\bar{x}_{2\cdot}$
	x_{212}	x_{222}	...	x_{2b2}		
	$\vdots$	$\vdots$	$\vdots$	$\vdots$		
	x_{21m}	x_{22m}	...	x_{2bm}		
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
A_a	x_{a11}	x_{a21}	...	x_{ab1}	$T_{a\cdot}$	$\bar{x}_{a\cdot}$
	x_{a12}	x_{a22}	...	x_{ab2}		
	$\vdots$	$\vdots$	$\vdots$	$\vdots$		
	x_{a1m}	x_{a2m}	...	x_{abm}		
$T_{\cdot j}$	$T_{\cdot 1}$	$T_{\cdot 2}$	...	$T_{\cdot b}$	$T = \sum_i \sum_j \sum_k x_{ijk}$	
$\bar{x}_{\cdot j}$	$\bar{x}_{\cdot 1}$	$\bar{x}_{\cdot 2}$	...	$\bar{x}_{\cdot b}$	$\bar{x} = \frac{T}{abm}$	

上表中 A 、 B 可以是固定型因素,也可以是随机型因素,现就固定效应模型、随机效应模型、混合模型加以讨论。

二、固定效应模型

1. 有重复实验的固定效应模型

当 A 、 B 两因素均为固定效应时,观察值的线性可加模型为

$$x_{ijk} = \mu + \alpha_i + \beta_j + (\alpha, \beta)_{ij} + \epsilon_{ijk} \quad (2.28)$$

上式中 μ 是总平均效应; α_i 为 A 因素第 i 水平处理效应; β_j 为 B 因素第 j 水平的处理效应; $(\alpha, \beta)_{ij}$ 为 α_i 和 β_j 之间的交互作用的效应; ϵ_{ijk} 是随机成份(试验误差)。当两因素均为固定型因素时,各处理效应为距总平均效应的离差。所以,

$$\sum \alpha_i = 0$$

$$\sum \beta_j = 0$$

交互作用的效应也是固定的

$$\sum_{i=1}^a (\alpha, \beta)_{ij} = \sum_{j=1}^b (\alpha, \beta)_{ij} = 0$$

ϵ_{ijk} 为相互独立且遵从 $N(0, \sigma^2)$ 的随机变量,所以

$$E(\epsilon_{ijk}) = 0$$

如果试验重复 n 次,实验的总次数为 abn 次。现以 $x_{i..}$ 表示 A 因素第 i 水平的所有观察值的总和。

$x_{.j.}$ 表示 B 因素第 j 水平的所有观察值的总和。

$x_{ij.}$ 表示 A 因素第 i 水平和 B 因素第 j 水平的所有观察值的总和。

$x_{...}$ 表示所有观察值总和。

用公式表示为

$$x_{i..} = \sum_{j=1}^b \sum_{k=1}^n x_{ijk}, \quad \bar{x}_{i..} = \frac{1}{bn} x_{i..}$$

$$x_{.j.} = \sum_{i=1}^a \sum_{k=1}^n x_{ijk}, \quad \bar{x}_{.j.} = \frac{1}{an} x_{.j.}$$

$$x_{ij.} = \sum_{k=1}^n x_{ijk}, \quad \bar{x}_{ij.} = \frac{1}{n} x_{ij.} \quad (i=1, 2, \dots, a)$$

$$(j=1, 2, \dots, b)$$

$$(k=1, 2, \dots, n)$$

两向分组两因素固定效应模型的方差分析零假设为:

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_a = 0$$

$$\beta_1 = \beta_2 = \dots = \beta_b = 0$$

$$(\alpha\beta)_{ij} = 0 \quad \begin{cases} i=1, 2, \dots, a \\ j=1, 2, \dots, b \end{cases}$$

其统计分析过程,基本做法是将总平方和分解为

$$\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (x_{ijk} - \bar{x}_{...})^2$$

$$= \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n [(x_{i..} - \bar{x}_{...}) + (\bar{x}_{.j.} - \bar{x}_{...})$$

$$+ (x_{ij.} - \bar{x}_{i..} - \bar{x}_{.j.} + \bar{x}_{...}) + (x_{ijk} - \bar{x}_{ij.})]^2$$

$$= bn \sum_{i=1}^a (\bar{x}_{i..} - \bar{x}_{...})^2 + an \sum_{j=1}^b (\bar{x}_{.j.} - \bar{x}_{...})^2$$

$$+ n \sum_{i=1}^a \sum_{j=1}^b (\bar{x}_{ij.} - \bar{x}_{i..} - \bar{x}_{.j.} + \bar{x}_{...})^2$$

$$+ \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (x_{ijk} - \bar{x}_{ij.})^2 \quad (2.29)$$

如果由 A 因素所引起平方和表示为 L_A , B 因素所引起的平方和表示为 L_B , A 、 B 交互作用所引起平方和为 L_{AB} , 试验误差平方和为 L_e , 那么

$$L_A = bn \sum_{i=1}^a (\bar{x}_{i..} - \bar{x}_{...})^2 \quad (2.30)$$

$$L_B = an \sum_{j=1}^b (x_{.j} - \bar{x}_{..})^2 \quad (2.31)$$

$$L_{AB} = n \sum_{i=1}^a \sum_{j=1}^b (\bar{x}_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..})^2 \quad (2.32)$$

$$L_e = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (x_{ijk} - \bar{x}_{ij.})^2 \quad (2.33)$$

从上式所得试验误差的平方和公式中可以看出, 试验至少要重复两次。各项平方和的自由度为 $df_A = (a-1)$, $df_B = (b-1)$, $df_{AB} = (a-1)(b-1)$, $df_e = ab(n-1)$ 。

各项均方为

$$S_A^2 = \frac{L_A}{a-1} \quad S_B^2 = \frac{L_B}{b-1}$$

$$S_{AB}^2 = \frac{L_{AB}}{(a-1)(b-1)} \quad S_e^2 = \frac{L_e}{ab(n-1)}$$

从线性模型 2.28 出发, 可以推得

$$\bar{x}_{i.} = \frac{1}{bn} \sum_j \sum_k x_{ijk} = \mu + \alpha_i + \bar{\epsilon}_{i.}$$

$$\bar{x} = \mu + \bar{\epsilon}$$

$$\text{其中 } \bar{\epsilon}_{i.} = \frac{1}{a} \sum_{i=1}^a \epsilon_{ijk} \quad \bar{\epsilon} = \frac{1}{abn} \sum_i \sum_j \sum_k \epsilon_{ijk}$$

于是

$$\begin{aligned} E(L_A) &= E\left[bn \sum_i (\bar{x}_{i.} - \bar{x})^2\right] \\ &= E\left[bn \sum_i (\mu + \alpha_i + \bar{\epsilon}_{i.} - \mu - \bar{\epsilon})^2\right] \\ &= E\left[bn \sum_i (\alpha_i + \bar{\epsilon}_{i.} - \bar{\epsilon})^2\right] \\ &= bn \sum \alpha_i^2 + \sum [E(\bar{\epsilon}_{i.}^2)] - \sum [E(\bar{\epsilon}^2)] \\ &= bn \sum \alpha_i^2 + a \cdot \sigma^2 - \sigma^2 \\ &= bn \sum \alpha_i^2 + (a-1)\sigma^2 \end{aligned} \quad (2.34)$$

所以

$$E(S_A^2) = E\left(\frac{L_A}{a-1}\right) = \sigma^2 + \frac{bn}{a-1} \sum \alpha_i^2$$

同理可推出

$$E(S_B^2) = E\left(\frac{L_B}{b-1}\right) = \sigma^2 + \frac{an}{b-1} \sum \beta_j^2 \quad (2.35)$$

$$\begin{aligned} E(S_{AB}^2) &= E\left[\frac{L_{AB}}{(a-1)(b-1)}\right] \\ &= \sigma^2 + \frac{n}{(a-1)(b-1)} \sum_{i=1}^a \sum_{j=1}^b (\alpha, \beta)_{ij}^2 \end{aligned} \quad (2.36)$$

$$E(S_e^2) = E\left[\frac{L_e}{ab(n-1)}\right] = \sigma^2 \quad (2.37)$$

若令

$$K_A^2 = \sum \alpha_i^2 / (a-1), \quad K_B^2 = \sum \beta_j^2 / (b-1),$$

$$K_{AB}^2 = \sum_{i=1}^a \sum_{j=1}^b (\alpha, \beta)_{ij}^2 / (a-1)(b-1)$$

所以

$$E(S_A^2) = \sigma^2 + bnK_A^2$$

$$E(S_B^2) = \sigma^2 + anK_B^2$$

$$E(S_{AB}^2) = \sigma^2 + nK_{AB}^2$$

对于 A 因素的 F 检验统计量为

$$\begin{aligned} F &= \frac{S_A^2}{S_e^2} = \frac{S_A^2 \text{ 的估计量 } E(S_A^2)}{S_e^2 \text{ 的估计量 } E(S_e^2)} \\ &= \frac{\sigma^2 + bnK_A^2}{\sigma^2} \end{aligned} \quad (2.38)$$

如果当 $K_A^2 = 0$ 时, $E(S_A^2)$ 的无偏估计值为 σ^2 , 即 $\alpha_1 = \alpha_2 = \dots = \alpha_a = 0$ 时 $E(S_A^2)$ 的无偏估计值为 σ^2 , 所以接受假设, 当 $K_A^2 \neq 0$ 可能 $F > F_\alpha$, 从而拒绝原假设。对于 B 因素, $A \times B$ 也是同理作出推断。

两因素固定模型方差分析表如下

表 2.18

变差来源	平方和	自由度	均方	F	均方期望
A	L_A	$a-1$	$S_A^2 = \frac{L_A}{(a-1)}$	$\frac{S_A^2}{S_e^2}$	$\sigma^2 + bnK_A^2$
B	L_B	$b-1$	$S_B^2 = \frac{L_B}{(b-1)}$	$\frac{SB_2}{S_e^2}$	$\sigma^2 + anK_B^2$
$A \times B$	$L_{A \times B}$	$(a-1) \times (b-1)$	$S_{AB}^2 = \frac{L_{AB}}{(a-1)(b-1)}$	$\frac{S_{AB}^2}{S_e^2}$	$\sigma^2 + nK_{AB}^2$
e	L_e	$ab(n-1)$	$S_e^2 = \frac{L_e}{ab(n-1)}$		σ^2
总和					

在具体统计分析时仍按下式计算各项平方和

$$L_T = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n x_{ijk}^2 - \frac{x_{...}^2}{abn}$$

$$L_A = \frac{1}{bn} \sum_{i=1}^a x_{i..}^2 - \frac{x_{...}^2}{abn}$$

$$L_B = \frac{1}{an} \sum_{j=1}^b x_{.j.}^2 - \frac{x_{...}^2}{abn}$$

L_{AB} 的计算是由重复间的平方数求出,可以证明其总平方和 L_{ST}

$$L_{ST} = L_A + L_B + L_{AB}$$

而 L_{ST} 由下列公式计算得到

$$L_{ST} = \frac{1}{n} \sum_{i=1}^a \sum_{j=1}^b x_{ij.}^2 - \frac{x_{...}^2}{abn}$$

所以 $L_{AB} = L_{ST} - L_A - L_B$

而 $L_e = L_T - L_{ST}$

也可以通过计算重复间平方和得到误差平方和

$$L_e = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n x_{ijk}^2 - \frac{1}{n} \sum_{i=1}^a \sum_{j=1}^b x_{ij.}^2$$

例 2.6 用 3 种深翻和 4 种施肥方案组成 12 种育苗作业方式试验,取得苗高的资料如表 2.19。设苗高的分布为正态分布,且等方差,试分析深翻、施肥以及它们交互作用对苗高生长有无显著影响。

表 2.19

深翻 B \ 施肥 A						
	A ₁	A ₂	A ₃	A ₄	x _{·j·}	$\bar{x}_{·j}$
B ₁	52	48	34	45		
	43	134	37	114	42	114
	58	145	507	42.25		
B ₂	39	29	36	42		
	41	50	36	44		
	47	141	41	121	39	119
B ₃	46	150	531	44.25		
	53	30	44	60		
	39	36	37	43		
X _{i..}	38	129	48	131	40	109
	56	140	509	42.42		
	42	47	32	41		
X _{...}	404	366	342	435	x _{...} =1547	
x _{...}	44.89	40.67	38.0	48.33	$\bar{x}$ =42.97	

解 作统计假设:假设施肥 A 及深翻 B 因素和它们的交互作用对试验结果均无显著影响,则

$$\sum_{i=1}^4 \sum_{j=1}^3 \sum_{k=1}^3 x_{ijk}^2 = 68367, \sum_{i=1}^4 \sum_{j=1}^3 x_{ij.}^2 = 201439$$

$$\sum_{j=1}^3 x_{.j}^2 = 507^2 + 531^2 + 509^2 = 798091$$

$$\sum x_{i..}^2 = 404^2 + 366^2 + 342^2 + 435^2 = 603361$$

$$C = \frac{x_{...}^2}{abn} = \frac{1547^2}{4 \times 3 \times 3} = 66478.03$$

$$L_T = \sum \sum \sum x_{ijk}^2 - C = 1888.97$$

$$L_A = \frac{1}{bn} \sum_i x_{i..}^2 - C = \frac{603361}{9} - 66478.03 = 562.08$$

$$L_B = \frac{1}{an} \sum_j x_{.j}^2 - C = \frac{798091}{9} - 66478.03 = 29.56$$

$$\begin{aligned} L_{ST} &= \frac{1}{n} \sum_{i=1}^4 \sum_{j=1}^3 x_{ij.}^2 - C \\ &= \frac{1}{3} (134^2 + 141^2 + \cdots + 140^2) - 66478.03 \\ &= \frac{1}{3} \times 201439 - 66478.03 \\ &= 67146.33 - 66478.03 \\ &= 668.30 \end{aligned}$$

$$\begin{aligned} L_{AB} &= L_{ST} - L_A - L_B \\ &= 668.30 - 562.68 - 29.56 \\ &= 76.66 \end{aligned}$$

$$\begin{aligned} L_e &= \sum_{i=1}^4 \sum_{j=1}^3 \sum_{k=1}^3 x_{ijk}^2 - \frac{1}{n} \sum_{i=1}^4 \sum_{j=1}^3 x_{ij.}^2 \\ &= 68367 - \frac{1}{3} \times 201439 \\ &= 68367 - 67146.33 \\ &= 1220.67 \end{aligned}$$

列出方差分析

表 2.20

变差来源	自由度	平方和	均方	F	F_{α}
A	3	562.08	187.36	386*	$F_{0.05}(3,24)=3.01$
B	2	29.56	14.78	0.29	$F_{0.05}(2,24)=3.40$
$A \times B$	6	76.66	12.78	0.25	$F_{0.05}(6,24)=2.51$
e	24	1220.67	50.86		
总和	35	1888.97	—	—	—

因为 $F_A > F_{\alpha}$, 所以 A 因素即施肥对试验结果的苗木高生长有显著影响, 而深翻与施肥的交互作用对苗木高生长均无显著影响。

对 A 因素各水平进行 q 检验, 取 $\alpha=0.05$, 组数 $a=4$, 该差项自由度为 24, 则查 q 表得 $q_{0.05}(5,24)=3.90$

所以

$$D = q_{0.05}(a, f_e) \sqrt{\frac{S_e^2}{a}}$$

$$= 9.28$$

列表表示各水平的差异性

表 2.21

$\begin{matrix} \bar{x}_i - \bar{x}_j \\ \bar{x}_i \end{matrix}$	$\bar{x}_{A_1} - \bar{x}_{A_3}$	$\bar{x}_{A_1} - \bar{x}_{A_2}$	$\bar{x}_{A_1} - \bar{x}_{A_4}$
$\bar{x}_{A_1} = 48.33$	10.33*	7.66	3.44
$\bar{x}_{A_2} = 44.89$	6.89	4.22	
$\bar{x}_{A_3} = 40.67$	2.67		
$\bar{x}_{A_4} = 38.00$			

从表 2.21 看出, 仅 $\bar{x}_{A_1}$ 与 $\bar{x}_{A_3}$ 有显著差异, 其余各组双双之间无显著差异。

二、无重复实验的固定模型

如果能够准确判断两因素间没有交互作用,可以在交叉组试验条件下不进行重复试验即 $n=1$ 。

如何准确判断两因素间没有交互作用,可以用 *TuTey* (1949)的 F 检验法。

其做法是将残余项平方和 $(L_T - L_A - L_B)$ 分解为具有自由度为 1 的交互作用成份 L_N 和具有 $(a-1)(b-1)-1$ 自由度的误差成份 L_e

$$L_N = \frac{\left[\sum_{i=1}^a \sum_{j=1}^b x_{ij} x_{i.} x_{.j} - x_{..} (L_A + L_B + \frac{x_{..}^2}{ab}) \right]}{abL_AL_B} \quad (2.39)$$

$$L_e = L_{\text{残余}} - L_N \quad (2.40)$$

为检验交互作用存在,建立 F 统计量。

$$F = \frac{\frac{L_N}{1}}{\frac{L_e}{(a-1)(b-1)-1}} \quad (2.41)$$

当 $F > F_{1, (a-1)(b-1)-1}^{\alpha}$ 时,认为两因素存在交互作用,否则认为两因素无交互作用存在。

在交叉分组的两因素实践中,若因素间不存在交互作用,那么观察值的线性模型为

$$x_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij} \begin{cases} i=1, 2, \dots, a \\ j=1, 2, \dots, b \end{cases} \quad (2.42)$$

$$\sum_{i=1}^a \alpha_i = 0, \quad \sum_{j=1}^b \beta_j = 0$$

ϵ_{ij} 为相互独立且服从正态分布 $N(0, \sigma^2)$ 的随机变量

零假设为 $H_0: \alpha_i = 0, \beta_j = 0$

方差分析如表 2.22

表 2.22

变差来源	平方和	自由度	均方	F	均方期望
A	L_A	$a-1$	$S_A^2 = \frac{L_A}{(a-1)}$	$\frac{S_A^2}{S_e^2}$	$\sigma^2 + bK_A^2$
B	L_B	$b-1$	$S_B^2 = \frac{L_B}{(b-1)}$	$\frac{S_B^2}{S_e^2}$	$\sigma^2 + aK_B^2$
e	L_e	$(a-1)(b-1)$	$S_e^2 = \frac{L_e}{(a-1)(b-1)}$		σ^2
总和	L_T				

例 2.7 将落叶苗木栽在 4 块不同苗床上, 每块苗床上苗木又分别使用 3 种不同的肥料以观察肥效差异, 一年后于每一苗床的各施肥小区内用重复抽样方式各抽取苗木若干株测其平均高, 得表 2.23 数据, (表中数据均减去一个常数值 50)

表 2.23

B(苗床) A(肥料)	B_1	B_2	B_3	B_4	$x_{..}$
A_1	0	-3	-3	3	-3
A_2	13	4	7	8	32
A_3	2	-8	9	-2	-7
$x_{.j}$	15	-7	-13	9	$x_{..} = 30$

解 建立统计假设: 假设肥料与苗床两因素对苗高生长无显著影响, 则

$$C = \frac{x_{..}^2}{ab} = 75$$

$$\sum_{i=1}^3 \sum_{j=1}^4 x_{ij}^2 = 0^2 + 13^2 + \cdots + (-2)^2 = 478$$

$$L_T = \sum_i \sum_j x_{ij}^2 - C = 478 - 75 = 403$$

$$L_A = \frac{1}{b} \sum_{i=1}^3 x_{i.}^2 - C = 183.5$$

$$L_B = \frac{1}{a} \sum_{j=1}^4 x_{.j}^2 - C = 99.67$$

$$L_e = L_T - L_A - L_B = 119.83$$

列出方差分析

表 2.24

变差来源	自由度	平方和	均方	F	F _α
A	2	187.5	91.75	4.59	F _{0.05} (2,6)=5.14
B	3	99.67	33.22	1.66	F _{0.05} (3,6)=4.76
e	6	119.83	19.97		
总和	11	466.0	—	—	—

因 $F_A < F_{\alpha}$, $F_B < F_{\alpha}$ 故判断肥料、苗床对试验结果均无显著影响。

三、随机模型

如果因素 A 和 B 都是随机效应,那么以方差分析的模型为随机模型,例如在杉木的无性系中选取 a 种并随机地安排在若干试验点上进行试验,以考察整个无性系在某一标志上的变异程度以及全部可能试验点上的差异程度时就属于随机效应试验,方差分析的模型为随机效应模型。

此时每一观察值的线性可加模型可表述为

$$x_{ijk} = \mu + \alpha_i + \beta_j + (\alpha, \beta)_{ij} + \epsilon_{ijk} \quad \begin{cases} i=1, 2, \dots, a \\ j=1, 2, \dots, b \\ k=1, 2, \dots, n \end{cases} \quad (2.43)$$

其中模型参量 α_i 、 β_j 、 $(\alpha, \beta)_{ij}$ 和 ϵ_{ijk} 都是随机变量。 α_i 服从

$N(0, \sigma_a^2)$, β_j 服从 $N(0, \sigma_\beta^2)$, $(\alpha, \beta)_{ij}$ 服从 $N(0, \sigma_{\alpha\beta}^2)$, ε_{ijk} 服从 $N(0, \sigma^2)$, 且它们互相独立, 因此任何观察值的方差可表示为

$$\sigma^2(x_{ijk}) = \sigma_a^2 + \sigma_\beta^2 + \sigma_{\alpha\beta}^2 + \sigma^2$$

其中 $\sigma_a^2, \sigma_\beta^2, \sigma_{\alpha\beta}^2, \sigma^2$ 称为方差分量

统计假设 H_0 : $\sigma_a^2 = 0$

$$\sigma_\beta^2 = 0$$

$$\sigma_{\alpha\beta}^2 = 0$$

方差分析的方法与固定模型的分析一样, 分别计算出 $L_T, L_A, L_B, L_{AB}, L_e$, 为了得到检验的统计量, 必须求出均方的数学期望值。这里勿须证明可以给出它们的数学期望为

$$\begin{aligned} E(S_A^2) &= bn\sigma_a^2 + n\sigma_{\alpha\beta}^2 + \sigma^2 \\ E(S_B^2) &= an\sigma_\beta^2 + n\sigma_{\alpha\beta}^2 + \sigma^2 \\ E(S_{AB}^2) &= n\sigma_{\alpha\beta}^2 + \sigma^2 \\ E(S_e^2) &= \sigma^2 \end{aligned} \quad (2.44)$$

由此可见, 两向分组的随机模型, 在检验假设 H_0 时与固定模型并不一样

$$F_A = \frac{S_A^2}{S_{AB}^2} = \frac{bn\sigma_a^2 + n\sigma_{\alpha\beta}^2 + \sigma^2}{n\sigma_{\alpha\beta}^2 + \sigma^2} \quad (2.45)$$

只有这个均方比才能检验出 σ_a^2 是否不为 0, 类似地

$$F_B = \frac{S_B^2}{S_{AB}^2} = \frac{an\sigma_\beta^2 + n\sigma_{\alpha\beta}^2 + \sigma^2}{n\sigma_{\alpha\beta}^2 + \sigma^2} \quad (2.46)$$

$$F_{AB} = \frac{S_{AB}^2}{S_e^2} = \frac{n\sigma_{\alpha\beta}^2 + \sigma^2}{\sigma^2} \quad (2.47)$$

请注意这里使用的统计量, 与因素 A 和 B 均为固定型因素时, 所使用的统计量是不同的。

随机效应模型的方差分析见表 2.25。

表 2.25

变差来源	平方和	自由度	均方	F	期望均方
A	L_A	$a-1$	S_A^2	$F_A = \frac{S_A^2}{S_{AB}^2}$	$\sigma^2 + n\sigma_{a\beta}^2 + bn\sigma_a^2$
B	L_B	$b-1$	S_B^2	$F_B = \frac{S_B^2}{S_{AB}^2}$	$\sigma^2 + n\sigma_{a\beta}^2 + an\sigma_b^2$
$A \times B$	$L_{A \times B}$	$(a-1)(b-1)$	S_{AB}^2	$F_{AB} = \frac{S_{AB}^2}{S_e^2}$	$\sigma^2 + n\sigma_{a\beta}^2$
e	L_e	$ab(n-1)$	S_e^2		σ^2
总和	L_T	$abn-1$	—	—	—

从表 2.25 看出方差分析结果 $\sigma_a^2 \neq 0, \sigma_b^2 \neq 0, \sigma_{a\beta}^2 \neq 0$ 条件下这三项方差估计量分别为

$$\hat{\sigma}_a^2 = \frac{S_A^2 - S_{AB}^2}{bn}$$

$$\hat{\sigma}_b^2 = \frac{S_B^2 - S_{AB}^2}{an}$$

$$\hat{\sigma}_{a\beta}^2 = \frac{S_{AB}^2 - S_e^2}{n}$$

$$\hat{\sigma}^2 = S_e^2$$

例 2.8 36 个泡桐无性系安排在 4 个区组内进行速生对比试验, 生长 6 年后随机抽取 10 个品种测其胸径, 得下表 2.26 数据, 问整个泡桐无性系生长变异是否显著? 并估计整个无性系遗传力。

表 2.26

无性系A \ 区组B	I	II	III	IV
1	15.0	14.8	10.7	13.1
2	17.4	17.0	13.8	13.5
3	17.9	18.3	15.0	16.1
4	20.1	19.8	17.8	18.3
5	21.2	23.2	18.4	18.1
6	23.0	23.2	20.3	20.8
7	24.2	24.0	21.8	21.5
8	27.2	26.5	25.0	24.8
9	30.7	29.8	28.9	27.0
10	30.5	31.0	27.0	27.5

解 由于在 36 个无性系中随机抽取 10 个无性系,同时考察 36 个无性系生长是否有差异性,所以认定为随机因素,采用随机效应模型进行分析。

统计假设 $H_0: \sigma_\alpha^2=0, \sigma_\beta^2=0, \sigma_{\alpha\beta}^2=0$, 计算过程略, 现列出方差分析表 2.27。

表 2.27

变差来源	平方和	自由度	均方	F	F _α
A	2043.29	9	227.63	$F_A = \frac{S_A^2}{S_{AB}^2} = 613.50^{**}$	$F_{9,27}^{0.01} = 3.15$
B	108.94	3	36.31	$F_B = \frac{S_B^2}{S_{AB}^2} = 58.14^{**}$	$F_{3,27}^{0.01} = 4.60$
A×B	9.995	27	0.37	$F_{AB} = \frac{S_{AB}^2}{S_e^2} = 0.62$	$F_{27,40}^{0.01} = 2.25$
e	24.315	40	0.6		
总和		79			

从方差分析表可以得到下列结论: H_0 中 $\sigma_\alpha^2=0, \sigma_\beta^2=0$ 两

个假设均应推翻,而 $\sigma_{\alpha\beta}^2$ 则不能推翻。

在本例中 σ_{α}^2 是由无性系引起胸径变异, $\sigma_{\alpha\beta}^2$ 及 σ^2 均由环境因素造成,故广义遗传力 H^2 的估计值

$$\begin{aligned} H^2 &= \frac{\sigma_{\alpha}^2}{\sigma_{\alpha}^2 + \sigma_{\alpha\beta}^2 + \sigma^2} \\ &= \frac{S_A^2 - S_e^2}{S_A^2 + (b-1)S_{AB}^2 + b(n-1)S_e^2} \\ &= \frac{227.03 - 0.6}{227.03 + (4-1) \times 0.37 + 4 \times (2-1) \times 0.6} \\ &= 0.98 \end{aligned}$$

例 2.9 为了研究不同杉木林分中,施用不同数量的造纸废液对杉木高生长量的影响。设计两向分组的试验,试验结果见表 2.28

表 2.28

		林 分 B					
		I		II		III	
施肥量 A	100 斤	8.69	8.49	8.80	8.74	9.49	9.37
	200 斤	8.88	8.72	9.68	9.54	9.39	9.59
	300 斤	10.82	10.86	11.00	10.92	11.07	11.01
	400 斤	11.16	11.42	10.97	11.13	11.00	10.90

解 按随机效应模型进行分析,原因是造纸废液对植物生长有利因素成份值是随机量,难以控制,同时在施用废液的林分中随机抽取 3 片林分进行分析,因此认定为随机因素。

将上表中每一 x_{ijk} 减去 9.5,列成表 2.29

表 2.29

施 用 量	林 分	x_{ij1}	x_{ij2}	x_{ij}	x_{ij}^2	$\sum_{k=1}^2 x_{ijk}^2$
100	I	-0.81	-1.03	-1.84	3.3856	1.7170
	II	-0.70	-0.76	-1.46	2.1316	1.0676
	III	-0.01	-0.13	-0.14	0.0196	0.0170
200	I	-0.62	-0.78	-1.40	1.9600	0.9928
	II	0.18	0.04	0.22	0.0484	0.0340
	III	-0.11	0.09	-0.02	0.0004	0.0202
300	I	1.32	1.36	2.68	7.1824	3.5920
	II	1.50	1.42	2.92	8.5264	4.2664
	III	1.57	1.51	3.08	9.4864	4.7450
400	I	1.66	1.92	3.58	12.8164	6.4420
	II	1.47	1.63	3.10	9.6100	4.8178
	III	1.50	1.40	2.90	8.4100	4.2100
				13.62	63.5772	32.9218

利用 x_{ij} 列成表 2.30

表 2.30

林 分						
				$x_{i..}$	$x_{i..}^2$	
	I	II	III			
施 用 量	100	-1.84	-1.46	-0.14	-3.44	11.8336
	200	-1.40	0.22	-0.02	-1.20	1.44007
	300	2.68	2.92	3.08	8.68	5.34249
	400	3.58	3.10	2.90	9.58	1.7764
						$\sum_i x_{i..}^2 = 180.3924$
$x_{.j}$	3.02	4.78	5.82	$x_{...} = 13.62$		
$x_{.j}^2$	9.1204	22.8484	33.8724			$\sum_j x_{.j}^2 = 65.8412$

由表 2.30 计算出

$$C = \frac{X_{...}^2}{abn} = \frac{13.62^2}{4 \times 3 \times 2} = 7.7294$$

$$L_T = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n x_{ijk}^2 - C = 32.9218 - 7.7294 = 25.1924$$

$$\begin{aligned} L_{ST} &= \frac{1}{n} \sum_{i=1}^a \sum_{j=1}^n x_{ij.}^2 - C = \frac{1}{2} (63.5772) - 7.7294 \\ &= 24.0592 \end{aligned}$$

由表 2.29 计算出

$$\begin{aligned} L_A &= \frac{1}{bn} \sum_{i=1}^a x_{i..}^2 - C \\ &= \frac{180.3924}{3 \times 2} - 7.7284 = 22.3360 \end{aligned}$$

$$\begin{aligned} L_B &= \frac{1}{an} \sum_{j=1}^n x_{.j.}^2 - C \\ &= \frac{65.8412}{4 \times 2} - 7.7294 = 0.5008 \end{aligned}$$

$$L_e = L_T - L_{ST} = 25.1924 - 24.0592 = 1.1332$$

$$L_{AB} = L_{ST} - L_A - L_B = 1.2229$$

列成方差分析表 2.31

表 2.31

变差来源	平方和	自由度	均方	F
A	22.3360	3	7.4553	36.53**
B	0.5008	2	0.2504	1.23
A×B	1.2229	4	0.2038	2.16
e	1.1327	12	0.0944	
总和	25.1924	23	—	—

从表 2.31 看出随机抽取不同林分对高生长没有显著影

响,但不同的施用量对高生长影响极为显著。

四、混合模型

采用混合模型进行分析的林业生产和科研实际问题颇多,例如用白僵菌、敌敌畏、松毛虫病毒等防治马尾松毛虫,考虑不同的药剂浓度试验,随机抽取几种浓度水平做试验,试验观察指标为死亡率。根据实际情况认定为混合模型。在两因素交叉分组实验中,若一个因素是固定因素(例如药剂),另一个因素是随机因素(例如浓度)。

现设 A 因素为固定型的, B 因素是随机的。则混合模型中每一个观察值 x_{ijk} 的线性统计模型为

$$x_{ijk} = \mu + \alpha_i + \beta_j + (\alpha, \beta)_{ij} + \epsilon_{ijk} \quad \begin{cases} i=1, 2, \dots, a \\ j=1, 2, \dots, b \\ k=1, 2, \dots, n \end{cases} \quad (2.48)$$

其中 α_i 是固定效应, β_j 是随机效应,并遵循正态分布 $N(0, \sigma_\beta)$, $(\alpha, \beta)_{ij}$ 被认定为随机效应,并遵循数学期望为 0, 方差为 $(\frac{a-1}{a})\sigma_{\alpha\beta}^2$ 的正态分布。

各项分量的期望均方如下

$$\begin{aligned} E(S_A^2) &= \sigma^2 + n\sigma_{\alpha\beta}^2 + \frac{bn}{a-1} \sum \alpha_i^2 \\ E(S_B^2) &= \sigma^2 + a n \sigma_\beta^2 \\ E(S_{AB}) &= \sigma^2 + n \sigma_{\alpha\beta}^2 \\ E(S_e^2) &= \sigma^2 \end{aligned} \quad (2.49)$$

由期望均方可以得到方差分析表中应计算的 3 个 F 值

$$\begin{aligned} F_A &= \frac{S_A^2}{S_{AB}^2}, \text{具有 } (a-1), (a-1)(b-1) \text{ 的自由度} \\ F_B &= \frac{S_B^2}{S_e^2}, \text{具有 } (b-1), ab(n-1) \text{ 的自由度} \end{aligned}$$

$F = \frac{S_{AB}^2}{S_e^2}$, 具有 $(a-1)(b-1), ab(n-1)$ 的自由度

固定因素的效应统计

$$\hat{\mu} = \bar{x}_{..}$$

$$\hat{\alpha}_i = \bar{x}_{i.} - \bar{x}_{..} \quad i = 1, 2, \dots, a$$

方差分量 $\sigma_\beta^2, \sigma_{a\beta}^2, \sigma^2$ 的统计值为

$$\hat{\sigma}_\beta^2 = \frac{S_B^2 - S_e^2}{an}$$

$$\hat{\sigma}_{a\beta}^2 = \frac{S_{AB}^2 - S_e^2}{n}$$

$$\hat{\sigma}^2 = S_e^2$$

混合模型方差分析表如下

表 2.32

变差来源	平方和	自由度	均方	F	均方期望
A	L_A	$a-1$	S_A^2	$\frac{S_A^2}{S_{AB}^2}$	$\sigma^2 + n\sigma_{a\beta}^2 + \frac{bn}{a-1} \sum \alpha_i^2$
B	L_B	$b-1$	S_B^2	$\frac{S_B^2}{S_e^2}$	$\sigma^2 + an\sigma_\beta^2$
AB	L_{AB}	$(a-1)(b-1)$	S_{AB}^2	$\frac{S_{AB}^2}{S_e^2}$	$\sigma^2 + n\sigma_{a\beta}^2$
e	L_e	$ab(n-1)$	S_e^2		σ^2
总和	L_T	$abn-1$			

例 2.10 考察马尾松木材密度、加荷速度以及二者交互作用对马尾松木材抗压强度是否有显著影响, 试验结果见表 2.33。其中密度单位为 g/cm^3 , 加荷速度单位为 kg/cm^3 , 抗压强度单位为 P_a 。

解 首先, 确认因素的类型。因素 A 马尾松木材密度, 难以控制, 而试样是从众多马尾松林中随机抽取的, 所以定为随机因素。因素 B 为加荷速度是严格控制的, 所以定为固定因

素。

表 2.33

B	A			$X_{..}$
	$A_1(0.428 \sim 0.466)$	$A_2(0.492 \sim 0.542)$	$A_3(0.524 \sim 0.556)$	
B_1	3.750	5.244	5.243	28.427
	3.685	5.199	5.306	
B_2	4.047	5.342	6.075	29.881
	3.807	5.138	5.472	
B_3	4.159	5.069	5.742	29.185
	3.854	5.069	5.292	
$x_{i..}$	23.302	31.061	33.130	$x_{..} = 87.493$

计算各项离差平方和和自由度

$$C = \frac{x_{..}^2}{abn} = \frac{87.493^2}{3 \times 3 \times 2} = 425.2792$$

$$L_T = \sum_i \sum_j \sum_k x_{ijk}^2 - C = 9.7062$$

$$L_A = \frac{1}{bn} \sum_{i=1}^3 x_{i..}^2 - C = 8.9484$$

$$L_B = \frac{1}{an} \sum_{j=1}^3 x_{.j.}^2 - C = 0.1763$$

$$L_e = \sum_i \sum_j \sum_k x_{ijk}^2 - \frac{1}{n} \sum_i \sum_j x_{ij.}^2 = 0.3842$$

$$L_{AB} = L_T - L_A - L_B - L_e = 0.3842$$

$$df_T = abn - 1 = 17, \quad df_A = a - 1 = 2,$$

$$df_B = b - 1 = 2, \quad df_{AB} = 4, \quad df_e = 9$$

列出方差分析表 2.34

表 2.34

变差来源	自由度	离差平方和	均方	F	F_{α}
A	2	8.9484	4.4742	104.782	$F_{(2,9)}^{0.05}=4.26$
B	2	0.1763	0.0882	1.789	$F_{(2,4)}^{0.05}=6.94$
AB	4	0.1973	0.0493	1.155	$F_{(4,9)}^{0.05}=3.63$
e	9	0.3842	0.0427		
Σ	17	9.7062			

从上表可以看出木材密度对抗压强度有显著的影响,认为不同木材密度造成的木材抗压强度的差异显著。

而加荷速度,以及交互作用对抗压强度未造成显著影响。

在随机效应模型和混合效应模型中,最好应设置重复,这样可以提高可靠程度。从上述叙述中可以知道,如果没有设置重复,因素间交互作用和实验误差无法区分,全部归于该误差项。

将固定模型、随机模型、混合模型,期望均方及 F 值汇总参见表 2.35

表 2.35

变差来源	平方和	自由度	均方	固定模型期望均方	F	随机模型期望均方	F	A 固定 B 随机期望均方	F
A	SS_A	f_A	S_A^2	$bnK_A^2 + \sigma^2$	$\frac{S_A^2}{S_e^2}$	$bn\sigma_A^2 + \sigma^2 + n\sigma_{AB}^2$	$\frac{S_A^2}{S_{AB}^2}$	$bn\sigma_A^2 + \sigma^2$	$\frac{S_A^2}{S_e^2}$
B	SS_B	f_B	S_B^2	$anK_B^2 + \sigma^2$	$\frac{S_B^2}{S_e^2}$	$an\sigma_B^2 + \sigma^2 + n\sigma_{AB}^2$	$\frac{S_B^2}{S_{AB}^2}$	$anK_B^2 + \sigma^2 + n\sigma_{AB}^2$	$\frac{S_B^2}{S_{AB}^2}$
$A \times B$	SS_{AB}	f_{AB}	S_{AB}^2	$n\sigma_{AB}^2 + \sigma^2$	$\frac{S_{AB}^2}{S_e^2}$	$n\sigma^2 AB + \sigma^2$	$\frac{S_{AB}^2}{S_e^2}$	$n\sigma^2 AB + \sigma^2$	$\frac{S_{AB}^2}{S_e^2}$
e	SS_e	f_e	S_e^2	σ^2	—	σ^2	—	σ^2	—
总和	SS_T	f_T							

第三节 系统分组的方差分析

系统分组的方差分析是上述两种方式分组方差分析的扩展。例如,对3个马尾松的种源 A_1 、 A_2 、 A_3 ,各考察它们对施肥量高生长效应,由于根据以往经验可知,对种源 A_1 、 A_2 、 A_3 的施肥量不能采用一致的施肥量水平。对 A_1 的相应施肥量为 B_1^1 、 B_2^1 、 $\dots$ 、 B_k^1 ,对 A_2 的相应施肥量为 B_1^2 、 B_2^2 、 $\dots$ 、 B_k^2 。像这样每一个因素各水平所对应的另一个因素的各水平不相一致的情况,采用系统分组设计是适应的。系统分组设计采用一个因素套另一个因素的阶梯设计方案分组,如图1为因素的系统设计。

A_1				A_2				A_3			
B_1^1	B_2^1	$\dots$	B_k^1	B_1^2	B_2^2	$\dots$	B_k^2	B_1^3	B_2^3	$\dots$	B_k^3
x_{111}	x_{121}	$\dots$	x_{1k1}	x_{211}	x_{221}	$\dots$	x_{2k1}	x_{311}	x_{321}	$\dots$	x_{3k1}
x_{112}	x_{122}	$\dots$	x_{1k2}	x_{212}	x_{222}	$\dots$	x_{2k2}	x_{312}	x_{322}	$\dots$	x_{3k2}
$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$		$\vdots$
x_{11n}	x_{12n}		x_{1kn}	x_{21n}	x_{22n}		x_{2kn}	x_{31n}	x_{32n}		x_{3kn}

图 1

系统分组的双因素试验。第一因素和第二因素不能等同看待,在某些情况下因素的位置不能互换,因此对试验结果的分析,首先侧重第一因素,其次才是第二因素。

现在先进行离差平方和及自由度的分解。

设 A 为第一因素,分为 a 个水平 A_1 、 A_2 、 $\dots$ 、 A_a , B 为第二

因素。为讨论方便,设在 A 因素的每一个水平下 B 水平数均为 b 个,且每一 A_i 和 B_j 组合下试验均重复 n 次, x_{ijk} 表示在 A 因素第 i 水平下 B 因素第 j 水平的第 k 次重复试验值,共进行试验次数 k 为 $a \times b \times n$ 。令

$$\begin{aligned}x_{ij.} &= \sum_{k=1}^n x_{ijk} \\x_{i..} &= \sum_{j=1}^b \sum_{k=1}^n x_{ijk} \\x_{...} &= \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n x_{ijk} \\\bar{x}_{ij} &= \frac{1}{n} \sum_{k=1}^n x_{ijk} = \frac{1}{n} x_{ij.} \\\bar{x}_i &= \frac{1}{bn} \sum_{j=1}^b \sum_{k=1}^n x_{ijk} = \frac{1}{bn} x_{i..} \\\bar{x} &= \frac{1}{abn} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n x_{ijk} = \frac{1}{abn} x_{...}\end{aligned}$$

应用离差平方和分解原理,有

$$\begin{aligned}L_T &= \sum_i \sum_j \sum_k (x_{ijk} - \bar{x})^2 \\&= \sum_i \sum_j \sum_k [(x_{ijk} - \bar{x}_{ij}) + (\bar{x}_{ij} - \bar{x}_i) + (\bar{x}_i - \bar{x})]^2 \\&= \sum_i \sum_j \sum_k (x_{ijk} - \bar{x}_{ij})^2 + n \sum_i \sum_j (\bar{x}_{ij} - \bar{x}_i)^2 \\&\quad + bn \sum_i (\bar{x}_i - \bar{x})^2\end{aligned} \quad (2.49)$$

式 2.49 中右端第一项表示 $A_i \times B_j$ 分组内 n 次重复试验结果变动,它应该归结于随机变动造成的,令其为 L_e ,右端第三项主要为 A 因素造成的组间变动,令其为 L_A ,第二项在因素 A 各水平变动的前提下 B 因素各水平造成的组间变动,令

其为 L_B , 所以

$$L_A = bn \sum_i (\bar{x}_i - \bar{x})^2 = \frac{1}{bn} \sum_{i=1}^a x_{i..}^2 - \frac{x_{...}^2}{abn} \quad (2.50)$$

$$L_B = n \sum_i \sum_j (\bar{x}_{ij} - \bar{x}_{i.})^2 = \frac{1}{n} \sum_i \sum_j x_{ij.}^2 - \frac{1}{bn} \sum_i x_{i..}^2 \quad (2.51)$$

$$\begin{aligned} L_e &= \sum_i \sum_j \sum_k (x_{ijk} - \bar{x}_{ij})^2 \\ &= \sum_i \sum_j \sum_k x_{ijk}^2 - n \sum_i \sum_j \bar{x}_{ij}^2 \\ &= \sum_i \sum_j \sum_k x_{ijk}^2 - \frac{1}{n} \sum_i \sum_j x_{ij.}^2 \end{aligned} \quad (2.52)$$

$$L_T = \sum_i \sum_j \sum_k (x_{ijk} - \bar{x})^2 = \sum_i \sum_j \sum_k x_{ijk}^2 - \frac{x_{...}^2}{abn} \quad (2.53)$$

如果采用固定模型, 根据方差分析的原理, 有

$$F_A = \frac{L_A/(a-1)}{L_e/ab(n-1)} = \frac{S_A^2}{S_e^2} \sim F[a-1, ab(n-1)] \quad (2.54)$$

$$F_B = \frac{L_B/(b-1)}{L_e/ab(n-1)} = \frac{S_B^2}{S_e^2} \sim F[b-1, ab(n-1)] \quad (2.55)$$

那么其他的随机模型, 混合模型的分析有所不同, 具体见表 2.36。

表 2.36

变差来源	自由度	均方	固定模型 期望均方	F	随机模型 期望均方	F	A 固定、B 随机 期望均方	F
A	a-1	S_A^2	$bnK_A^2 + \sigma^2$	$\frac{S_A^2}{S_e^2}$	$\frac{\sigma_A^2 + \sigma^2 + n\sigma_B^2}{n\sigma_B^2}$	$\frac{S_A^2}{S_B^2}$	$bnK_A^2 + \sigma^2$	$\frac{S_A^2}{S_B^2}$
B	a(b-1)	S_B^2	$nK_B^2 + \sigma^2$	$\frac{S_B^2}{S_e^2}$	$\frac{n\sigma_B^2\sigma_2 + \sigma_2}{n\sigma_B^2}$	$\frac{S_B^2}{S_A^2}$	$n\sigma_B^2\sigma_2$	$\frac{S_B^2}{S_A^2}$
e	ab(n-1)	S_e^2	σ^2		σ_2		σ^2	
T	—	—	—	—	—	—	—	—

例 2.11 研究施肥量不同对不同马尾松 3 种品种 (A_1 、 A_2 、 A_3) 苗高的影响, 根据每一品种生物学特性, 拟订 3 种施肥量, 组成系统分组试验, 试验结果数据如表 2.37, 试问不同品种不同施肥量对苗高的影响是否显著?

表 2.37

A 品种	B 施肥量 kg/亩								
	B_1 (2.5)	B_2 (3.0)	B_3 (3.5)	B_1 (3.5)	B_2 (9.0)	B_3 (4.5)	B_1 (4.0)	B_2 (5.0)	B_3 (6.0)
A_1	68 70	76 73	71 72						
A_2				70 73	74 76	82 80			
A_3							90 89	95 97	104 102

解 我们认为 A、B 两个因素为随机因素, 所以采用随机模型进行分析。

作统计假设 $H_0: \sigma_A^2 = 0, \sigma_{B(A)}^2 = 0$

$$\text{则 } \sum_i \sum_j \sum_k x_{ijk}^2 = 121134 \quad x_{...} = \sum_i \sum_j \sum_k x_{ijk} = 1462$$

$$\sum_i \sum_j x_{ij.}^2 = 242228 \quad \sum_i x_{i..}^2 = 724854$$

$$C = \frac{x_{...}^2}{abn} = \frac{1462^2}{3 \times 3 \times 2} = 118746.8889$$

$$L_T = \sum_i \sum_j \sum_k x_{ijk}^2 - C = 2387.1111$$

$$L_A = \frac{1}{bn} \sum_i x_{i..}^2 - C = \frac{1}{3 \times 2} \times 724854 - C \\ = 2062.1111$$

$$L_B = \frac{1}{n} \sum_j \sum_i x_{ij.}^2 - \frac{1}{bn} \sum_i x_{i..}^2 \\ = \frac{1}{2} \times 242228 - \frac{1}{3 \times 2} \times 724854 = 305$$

$$L_e = \sum_i \sum_j \sum_k x_{ijk}^2 - \frac{1}{n} \sum_i \sum_j x_{ij.}^2 = 20$$

$$df_T = abn - 1 = 17 \quad df_A = a - 1 = 2$$

$$df_B = a(b-1) = 6 \quad df_e = ab(n-1) = 9$$

列出方差分析表 2.38

表 2.38

变差来源	自由度	变差平方和	均方	F	F _α
A	2	2062.1111	1031.06	20.28	F _{0.01} (2, 6) = 10.9
B	6	305.00	50.83	22.90	F _{0.01} (0.9) = 5.8
e	9	20.00	2.22		
T	17	2387.1111	—	—	—

从表 2.38 看出 F_A 和 F_B 值均大于 F_{α} 值, 故推翻假设 H_0 , 认为不同马尾松品种, 不同施肥量对苗木高度有显著影响。

第四节 两个以上因素的方差分析

上述介绍方法均考虑两个因素,可以把3种方式分组的方差分析推广到一般情况。如果某一试验考虑3个因素A、B、C。A因素有a个水平,B因素有b个水平,C因素有c个水平。每一处理有n个重复,那么总的观察次数有abcn次。

如果实验中,每一因素都是固定型,那么进行主效应和交叉效应的显著性检验都将主效应或交叉效应的均方除以误差均方,构成遵循F分布的随机变量,而后进行判断。每一个主效应的自由度是该因素的水平数减小,每一个交互作用的自由度是形成交互作用的两个因素的自由度的乘积。

$$df_A = a - 1 \quad df_B = b - 1 \quad df_C = c - 1$$

$$df_{AB} = (a - 1)(b - 1) \quad df_{AC} = (a - 1)(c - 1),$$

$$df_{BC} = (b - 1)(c - 1)$$

$$df_e = abc(n - 1) \quad df_{\text{总}} = abn - 1$$

$$F_A = \frac{S_A^2}{S_e^2} \sim F[(a - 1), abc(n - 1)]$$

$$F_B = \frac{S_B^2}{S_e^2} \sim F[(b - 1), abc(n - 1)]$$

$$F_C = \frac{S_C^2}{S_e^2} \sim F[(c - 1), abc(n - 1)]$$

$$F_{AB} = \frac{S_{AB}^2}{S_e^2} \sim F[(a - 1)(b - 1), abc(n - 1)]$$

$$F_{BC} = \frac{S_{BC}^2}{S_e^2} \sim F[(b - 1)(c - 1), abc(n - 1)]$$

$$F_{AC} = \frac{S_{AC}^2}{S_e^2} \sim F[(a - 1)(c - 1), abc(n - 1)]$$

$$\text{上式 } S_A^2 = L_A / (a-1) \quad S_B^2 = L_B / (b-1)$$

$$S_C^2 = L_C / (c-1)$$

$$S_{AB}^2 = L_{AB} / [(a-1)(b-1)] \quad S_{AC}^2 = L_{AC} / [(a-1)(c-1)]$$

$$S_{BC}^2 = L_{BC} / [(b-1)(c-1)]$$

那么这些离差平方计算依下列公式进行

$$L_T = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^n x_{ijkl}^2 - \frac{x_{\dots}^2}{abcn}$$

$$L_A = \frac{1}{bcn} \sum x_{i\dots}^2 - \frac{x_{\dots}^2}{abcn}$$

$$L_B = \frac{1}{acn} \sum x_{\dots j\dots}^2 - \frac{x_{\dots}^2}{abcn}$$

$$L_C = \frac{1}{abn} \sum x_{\dots \dots k\dots}^2 - \frac{x_{\dots}^2}{abcn}$$

计算交互作用平方和,要列出 3 个两向表,而后依下列公式计算

$$L_{AB} = L_{T(AB)} - L_A - L_B$$

$$L_{AC} = L_{T(AC)} - L_A - L_C$$

$$L_{BC} = L_{T(BC)} - L_B - L_C$$

而 $L_{T(AB)}$, $L_{T(AC)}$, $L_{T(BC)}$ 称为该两因素的次总平方和

$$L_{T(AB)} = \frac{1}{cn} \sum_{i=1}^a \sum_{j=1}^b x_{ij\dots}^2 - \frac{x_{\dots}^2}{abcn}$$

$$L_{T(AC)} = \frac{1}{bn} \sum_{i=1}^a \sum_{k=1}^c x_{i\dots k\dots}^2 - \frac{x_{\dots}^2}{abcn}$$

$$L_{T(BC)} = \frac{1}{an} \sum_{j=1}^b \sum_{k=1}^c x_{\dots jk\dots}^2 - \frac{x_{\dots}^2}{abcn}$$

而 3 个因素的交互作用的平方和是用三向表 x_{ijk} 中求出

$$L_{ABC} = L_T - L_A - L_B - L_C - L_{AB} - L_{BC} - L_{AC}$$

$$\text{而 } L_{ST} = \frac{1}{n} \sum_i \sum_j \sum_k x_{ijk}^2 - \frac{x_{\dots}^2}{abcn}$$

$$L_e = L_T - L_{T(ABC)}$$

例 2.12 在某项试验中,可能影响试验结果的有 3 个因素 A 、 B 、 C 。

每一处理重复 3 次试验。 A 因素有 3 个水平, B 因素有 3 个水平, C 因素有 2 个水平。试验结果如表 2.39。

表 2.39

A	C_1			C_2		
	B_1	B_2	B_3	B_1	B_2	B_3
1	10.7	10.3	11.2	10.9	10.5	12.2
	10.8	10.2	11.6	12.1	11.1	11.7
	11.3	10.5	12.0	11.5	10.3	11.0
2	11.4	10.2	10.7	9.8	12.6	10.8
	11.8	10.9	10.5	11.3	7.5	10.2
	11.5	10.5	10.2	10.9	9.9	11.5
3	13.6	12.0	11.1	10.7	10.2	11.9
	14.1	11.6	11.0	11.7	11.5	11.6
	14.5	11.5	11.5	12.7	10.9	12.2

试作方差分析检验各因素的主效应和交互效应是否显著?

解 编制计算 3 个两向表

表 2.40

C_1	B			总和
	B_1	B_2	B_3	
A_1	32.8	31.0	34.8	98.6
A_2	34.7	31.6	31.4	97.7
A_3	42.2	35.1	33.6	110.9
总和	109.7	97.7	99.8	307.2

表 2.41

C_1	B			总和
	B_1	B_2	B_3	
A_1	34.5	31.9	34.9	101.3
A_2	32.0	30.0	32.5	94.5
A_3	35.1	32.6	35.7	103.4
总和	101.6	94.5	103.1	299.2

表 2.42

A	B			总和
	B_1	B_2	B_3	
A_1	67.3	62.9	69.7	199.9
A_2	66.7	61.6	63.9	192.2
A_3	77.3	67.7	69.3	214.3
总和	211.3	192.2	202.9	606.4

表 2.43

A	C		总和
	C_1	C_2	
A_1	98.6	101.3	199.9
A_2	97.7	94.5	192.2
A_3	110.9	103.4	214.3
总和	307.2	299.2	606.4

表 2.44

B	C		总和
	C_1	C_2	
B_1	109.7	101.6	211.3
B_2	97.7	94.5	192.2
B_3	99.8	103.1	202.9
总和	307.2	299.2	606.4

计算各离差平方和

$$C = \frac{x_{\dots}^2}{abcn} = \frac{(606.4)^2}{54} = 6809.65$$

$$L_T = \sum \sum \sum x_{ijk}^2 - C$$

$$= 10.7^2 + 10.8^2 + \cdots + 12.2^2 - C$$

$$=63.19$$

$$L_A = \frac{1}{bcn} \sum_i x_{i...}^2 - C$$

$$= \frac{1}{18} (199 \cdot 9^2 + 192 \cdot 2^2 + 214 \cdot 3^2) - C$$

$$=13.98$$

$$L_B = \frac{1}{acn} \sum_j x_{.j.}^2 - C$$

$$= \frac{1}{18} (211 \cdot 3^2 + 19 \cdot 2^2 + 202 \cdot 9^2) - C$$

$$=10.18$$

$$L_C = \frac{1}{abn} \sum_k x_{..k}^2 - C$$

$$= \frac{1}{27} (307 \cdot 2^2 + 299 \cdot 2^2) - C$$

$$=1.18$$

$$L_{AB} = \frac{1}{cn} \sum_i \sum_j x_{ij.}^2 - C$$

$$= \frac{1}{6} (67 \cdot 3^2 + 66 \cdot 7^2 + \dots + 69 \cdot 3^2) - C$$

$$=4.78$$

$$L_{AC} = \frac{1}{bn} \sum_i \sum_k x_{i.k}^2 - C$$

$$= \frac{1}{6} (98 \cdot 6^2 + 97 \cdot 7^2 + \dots + 103 \cdot 4^2) - C$$

$$=2.92$$

$$L_{BC} = \frac{1}{an} \sum_i \sum_k x_{ijk}^2 - C$$

$$= \frac{1}{6} (109 \cdot 7^2 + 101 \cdot 6^2 + \dots + 103 \cdot 1^2) - C$$

$$=3.64$$

$$L_{ABC} = \frac{1}{n} \sum_i \sum_j \sum_k x_{ijk}^2 - C$$

$$= \frac{1}{3}(32.8^2 + 34.7^2 + \cdots + 35.7^2) - C$$

$$= 4.89$$

$$L_e = L_T - L_A - L_B - L_C - L_{AB} - L_{BC} - L_{AC} - L_{ABC}$$

$$= 21.62$$

列出方差分析表

表 2.45

变差来源	离差平方和	自由度	均方	F 值	F_{α}
A	13.98	2	6.99	11.65**	$F_{2,36}^{0.05} = 3.26$
B	10.18	2	5.09	8.48**	
C	1.18	1	1.18	1.97	$F_{2,36}^{0.01} = 5.25$
A×B	4.78	4	1.20	2.00	
A×C	2.92	2	1.46	2.43	$F_{1,36}^{0.05} = 4.11$
B×C	3.64	2	1.82	3.03	
ABC	4.89	4	1.22	2.03	$F_{4,36}^{0.05} = 2.63$
e	21.62	36	0.60		
总和	63.19	53	—	—	—

从上表中看出,因素 A、B 对试验结果有显著的影响,因素 C 和交互作用效应均没有对试验结果造成显著影响。

第五节 数据转换与漏失数据的弥补

方差分析是分析试验数据的一种科学方法,但并不是任何资料都可以直接作方差分析。我们知道方差分析是以满足独立、正态、随机、等方差这 4 个条件为前提的。在某些情况下,上述前提往往不能满足。例如,Poisson 分布及二项分布的平均数和标准差是有关联的,上述前提不能满足时所做的方差分析,往往导致错误的结论。

为了对某些不满足以上前提条件的数据进行适当的变换使之满足或近似满足前提条件,从而使采用方差分析能得到正确结论。

一、平方根变换

遵循泊松(Poisson)分布的资料一般用平方根变换,从观察资料来看,各处理内的标准差有极明显的差别,这类资料一般用计数表示。例如每网所获得的昆虫数,每试验小区的杂草数等。当资料中各小区某性状的出现个数频率大于 10 时,可将原始数据 $\{x_{ij}\}$ 变换成 $\{\sqrt{x_{ij}}\}$,如果某性状出现个数小于 10 时,或为 0 时,则变换成 $\{\sqrt{x_{ij}+1}\}$ 。

例 2.13 在某苗圃采用 5 种不同方法对土壤进行处理,来控制金针虫的蔓延,每种方法重复 3 次,处理后抽样检查,得存活金针虫数目如表 2.46,问不同处理方法对金针虫的存活是否有显著影响?

表 2.46

处理方法	x_{ij}			$x_{i.}$	$\bar{x}$	R
1	3	4	6	13	4.3	3
2	2	6	4	12	4	4
3	5	4	5	14	4.7	1
4	1	0	1	2	0.67	1
5	4	6	9	19	6.3	5

用 Bartlett 检验法不难验证上表的数据是否满足等方差。从极差的大小也可以粗略看出来,变换前方差不齐性。现在对上述数据实行 $x'_{ij} = \sqrt{x_{ij} + 1}$ 的变换,得表 2.47。

表 2.47

处理方法	$x'_{ij} = \sqrt{x_{ij} + 1}$			$\bar{x}_{i.}$	$\bar{x}$	R
1	2.00	2.24	2.65	6.89	2.3	0.65
2	1.73	2.65	2.24	6.62	2.2	0.92
3	2.65	2.24	2.65	7.54	2.5	0.41
4	1.41	1.00	1.41	3.82	1.3	0.41
5	2.24	2.65	3.16	8.05	2.7	0.92

从上表可以看出变换后的极差大小差别不大,所以其方差可以认为已趋于齐性。

二、反正弦变换

反正弦变换的方法是对每个观察值平方根的反正弦,使它变成一个角度,进行方差分析。这种变换方法适用于数据近似遵从二项分布的数据,此时,方差与期望也存在一定比例关系,满足似方差齐性的条件,例如不同采种区马尾松种子的发芽率、木麻黄木毒蛾在不同处理下的死亡率等。当百分数特别

大时或特别小(如 0~20 或 80~100)时应作数据变换。

变换公式为

$$x'_{ij} = \arcsin \sqrt{x_{ij}}$$

其中 x_{ij} 为百分率, x'_{ij} 为相应的角度值。各种百分率的相应角度值可查表。

例 2.14 在一片林分中, 设置 4 个区组, 施行 6 种不同处理防治害虫, 各小区的校正死亡率如表 2.48, 试问不同处理对各小区的校正死亡率是否有显著影响?

表 2.48

区组	处 理					
	1	2	3	4	5	6
1	19.3	10.1	25.2	14.0	3.3	3.1
2	29.2	34.7	36.5	30.2	35.8	9.6
3	1.0	14.0	23.4	7.2	1.1	1.0
4	6.4	5.6	12.9	8.9	2.0	1.0

解 直接观察数据, 各组变动甚大, 用 Bartlett 检验发现不满足方差齐性条件, 现作反正弦变换, 结果见表 2.49

表 2.49

区组	处 理					
	1	2	3	4	5	6
1	26.1	18.5	30.1	22.0	10.5	10.1
2	32.7	36.1	37.2	33.3	36.8	18.0
3	5.7	22.0	28.9	15.6	6.0	5.7
4	14.6	13.7	21.0	17.4	8.1	5.7
$\bar{x}$	19.8	22.6	29.3	22.1	15.4	9.9

对上表的数据仍用 Bartlett 检验可知变换后数据满足方差齐性。现对其进行方差分析, 可得表 2.50

表 2.50

变差来源	自由度	平方和	均方	F 值	F_{α}
区组	3	1392.85	464.19	19.19**	$F_{3,15}^{0.01}=5.42$
处理	5	885.72	177.12	7.30**	$F_{5,15}^{0.01}=4.56$
e	15	364.15	24.28		
T	23	2642.72	—	—	—

三、对数变换

对于前述的两类数据(百分率数据、计算数据),有时也可以用对数变换 $\lg x_{ij}$ 或 $\lg(x_{ij}+1)$ 变换成 x'_{ij} ,从而满足方差齐性的要求,对于标准差与平均数成比例的一类数据,用对数变换也常常是有效的,例子不再累述。

应当注意,原则上数据转换前后都应作统计检验,看是否正态,是否等方差,而正态分布和方差齐性的具体检验,请读者参看有关的统计书籍。

四、漏失数据的弥补

林业试验数据进行方差分析,试验前要进行周密设计。由于林业许多试验周期长,少则一、二年,多则几十年,并且许多试验均在野外进行,所以试验中,可能出现意外,造成试验数据漏失。由于试验数据的漏失会使试验结果不能按设计取得数据,必须进行弥补。

这里介绍弥补的方法。

统计上当试验无重复时常按补上缺少的数据后所得的剩余离差平方和应最小原则进行弥补。当试验有重复并且每一处理至少有一个以上数据未丢失时,丢失的数据可以用未丢失的数据平均值来代替,但丢失几个数据,作方差分析时剩余项的自由度就要相应减小几个。

例 2.15 试根据表 2.51 中资料对丢失的数据 a 和 b 进行弥补。

表 2.51

B	A				$x_{.j}$
	A_1	A_2	A_3	A_4	
B_1	3.5	y_1	2.0	1.4	$6.9 + y_1$
B_2	2.3	2.0	1.5	y_2	$5.8 + y_2$
B_3	2.0	1.9	1.2	0.3	5.4
$x_{i.}$	7.8	$3.9 + y_1$	4.7	$1.7 + y_2$	$x_{..} = 18.1 + y_1 + y_2$

由表 2.51 中的数据作下列计算

$$C = \frac{x_{..}^2}{a \times b} = \frac{(18.1 + y_1 + y_2)^2}{12}$$

$$\begin{aligned} L_A &= \frac{1}{3} \sum_{i=1}^4 x_{i.}^2 - C \\ &= \frac{1}{3} [7.8^2 + (3.9 + y_1)^2 + 4.7^2 + (1.7 + y_2)^2] - C \end{aligned}$$

$$\begin{aligned} L_B &= \frac{1}{4} \sum_{j=1}^4 x_{.j}^2 - C \\ &= \frac{1}{4} [(6.9 + y_1)^2 + (5.8 + y_2)^2 + 5.4^2] - C \end{aligned}$$

$$\begin{aligned} L_T &= \sum_i \sum_j x_{ij}^2 - C \\ &= [3.5^2 + 2.3^2 + 2.0^2 + y_1^2 + \cdots + y_2^2 + 0.3^2] - C \end{aligned}$$

$$L_e = L_T - L_A - L_B$$

我们采用这样的原则来确定 y_1 和 y_2 使 L_e 成为最小, 则 y_1 和 y_2 应该是方程组

$$\begin{cases} \frac{\partial L_e}{\partial y_1} = 0 \\ \frac{\partial L_e}{\partial y_2} = 0 \end{cases} \quad \text{的解, 不难得出}$$

$$\frac{\partial L_e}{\partial y_1} = 2y_1 - \frac{2}{3}(3.9 + y_1) - \frac{1}{2}(6.9 + y_1)$$

$$+ \frac{1}{6}(18.1 + y_1 + y_2)$$

$$= 0$$

$$\frac{\partial L_e}{\partial y_2} = 2y_2 - \frac{2}{3}(1.7 + y_2) - \frac{1}{2}(5.8 + y_2)$$

$$+ \frac{1}{6}(18.1 + y_1 + y_2)$$

$$= 0$$

$$\text{即} \begin{cases} 6y_1 + y_2 = 18.2 \\ y_1 + 6y_2 = 6.1 \end{cases}$$

$$\text{得 } y_1 = 2.95, \quad y_2 = 0.53$$

第三章 回归分析

林业试验直接应用回归分析,主要是通过试验设计得到表达回归关系的数学模型,这也是试验者所希望的。

回归分析主要是研究两个或多个变量之间相互联系的规律性的一种数学分析方法。要建立变量之间的关系式,需要收集大量资料。许多试验也由于收集资料的困难而无法进行,这使人们考虑到是否可以用较少的资料来研究变量间关系,而把回归分析引入到试验设计中正是解决的途径。

在林业生产和科研工作中,回归分析的应用是很广泛的,常用的有:

- (1)编制立木材积表,某地区某树种地位指数表、生长量表。
- (2)研究林木密度与林分出材量之间规律性。
- (3)研究施用药剂浓度与杀虫效果之间规律性。
- (4)研究林分平均高、密度、年龄与优势木平均高之间的关系。

回归分析主要解决以下几个方面问题:

- (1)确定几个特定变量之间的关系,并用数学表达式给予描述。
- (2)根据一个或几个变量数值,预测控制另一个变量取值,并计算出预测控制的精确度。
- (3)进行因素分析,找出哪些变量是重要的,哪些是次要的,以及它们之间的关系。

在生产实践和科学试验中,研究两个变量之间关系时,可以分为两种关系,确定性关系与非确定性关系。

例如,圆半径 r 取一个特定值,就会有一个特定圆面积与它对应,这两个变量均为非随机变量,这种关系是确定性关系,其关系式称为函数关系。另外还有一种关系,例如种子园单位面积产量与单位面积施肥量之间关系,施肥量相同时,而单位面积的产量不一定相同。因为决定种子园产量的因素不止一个施肥量,还有许多因素。如果认为施肥量是重要的,其他因素都是次要的,这就是造成不确定性的原因。我们的目的,是通过样本资料去从变量间联系的现象中寻求其统计规律。施肥量是非随机变量,单位面积产量是随机变量。我们称一个变量为随机变量、一个变量为非随机变量的变量间关系为回归关系。如果两个变量均为随机变量的变量间关系称为相关关系。我们主要研究回归关系。

第一节 一元回归模型

一、一元线性回归

如果回归关系为一次函数。设总体两变量 Y 与 X 存在着回归关系,其表达式为

$$\bar{Y}_x = A + BX + \epsilon \quad (3.1)$$

3.1 式称为线性回归模型,其中 x 为可以精确测定变量, A 、 B 为总体回归系数。 ϵ 为随机变量,对于不同 x 、 ϵ 作为不同随机变量彼此间互相独立,且具有相同方差,服从正态分布,由于 y 是 ϵ 的线性函数,因此与不同 X 对应的 Y 也是相互独立的正态变量。回归分析必须满足 4 条基本假设:正态、线性、

独立、等方差。也就是：

$$(1) E(\varepsilon_i) = 0$$

$$(2) E[\varepsilon_i - E(\varepsilon_i)]^2 = \sigma^2 > 0$$

$$(3) E(\varepsilon_i \varepsilon_j) = 0$$

$$(4) \varepsilon_i \sim N(0, \sigma)$$

现在的问题就是如何根据样本资料对参数 A, B 进行估计。

1. 直线回归方程的确定

假设从总体中抽取几个单元组成样本，样本各单元在自变量 X 与因变量 Y 的各观察值为 $(x_i, y_i) (i=1, 2, \dots, n)$ 。将这 n 对观察值点绘在直角坐标纸上，成一散点图，如图 3.1 所示。我们可以根据这些样本资料去估计 A, B ，从而得到经验的直线回归方程。

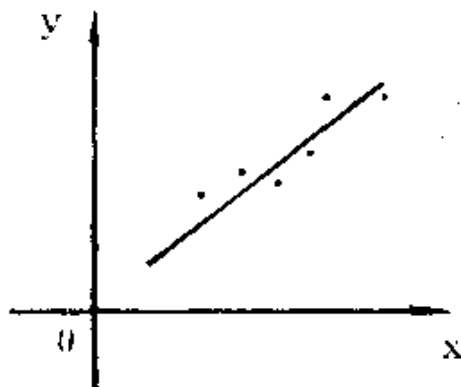


图 3.1

设 a, b 是 A, B 的估计值，于是经验直线回归方程：

$$\hat{y} = a + bx \quad (3.2)$$

a, b 为样本回归系数， $\hat{y}$ 为 Y_x 的估计值，称为回归值。

3.2 式中参数如何求得？这里介绍最小二乘法。

我们从 n 个样点分别测两个指标值 $(x_i, y_i) i=1, 2, \dots, n$ 。假设数据满足上述的基本假设。

设 a, b 是 A, B 的估计值，于是经验回归直线方程为

$$\hat{y}_i = a + bx_i$$

回归值 $\hat{y}_i$ 与实测值 y_i 之差 $(y_i - \hat{y}_i)$ 表示 y_i 与回归直线上对应于同一 x_i 的值 $\hat{y}_i$ 的偏离程度。同时考虑到正、负偏离关系,而采用其差值的平方和来描述其偏离的程度。

$$Q = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3.3)$$

在确定直线回归方程参数 a, b 时,希望所确定的 a, b 能使 Q 实现最小,用此原则求 a, b 的方法称为最小二乘法。

依此原则并根据微积分有关知识,得知需令

$$\frac{\partial Q}{\partial a} = 0 \quad \frac{\partial Q}{\partial b} = 0$$

亦即

$$\frac{\partial Q}{\partial a} = 2 \sum_{i=1}^n (y_i - a - bx_i)(-1) = 0$$

得

$$na + b \sum_{i=1}^n x_i = \sum_{i=1}^n y_i$$

$$\frac{\partial Q}{\partial b} = 2 \sum_{i=1}^n (y_i - a - bx_i)(-x_i) = 0$$

得

$$a \sum_{i=1}^n x_i + b \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i$$

联立得方程组

$$\begin{cases} na + b \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \\ a \sum_{i=1}^n x_i + b \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i \end{cases} \quad (3.4)$$

上述方程称为标准方程或正规方程。

由 3.4 式第一式可得

$$\begin{aligned} na + b \sum_{i=1}^n x_i &= \sum_{i=1}^n y_i \\ a &= \frac{1}{n} \sum_{i=1}^n y_i - b \frac{1}{n} \sum_{i=1}^n x_i \\ a &= \bar{y} - b\bar{x} \end{aligned} \quad (3.5)$$

将 3.5 式代入 3.4 式中第二式得

$$\begin{aligned} (\bar{y} - b\bar{x}) \sum_{i=1}^n x_i + b \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i y_i &= 0 \\ \bar{y} \sum_{i=1}^n x_i - b\bar{x} \sum_{i=1}^n 1 + b \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i y_i &= 0 \\ b \left(\sum_{i=1}^n x_i^2 - n\bar{x} \right) &= \sum_{i=1}^n x_i y_i - n\bar{x}\bar{y} \\ \text{则 } b &= \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \end{aligned} \quad (3.6)$$

$$\text{若令 } L_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad L_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2$$

$$\text{那么 } b = \frac{L_{xy}}{L_{xx}}$$

2. A、B 的估计值

根据直线回归方程关系,有

$$E(y_i) = E(A + Bx_i + \epsilon_i) = A + Bx_i$$

$$\begin{aligned}
E(\bar{y}) &= E\left[\frac{\sum_{i=1}^n y_i}{n}\right] = \frac{1}{n} E\left(\sum_{i=1}^n y_i\right) \\
&= \frac{1}{n} \sum_{i=1}^n E(y_i) = \frac{1}{n} \sum_{i=1}^n (A + Bx_i) \\
&= \frac{1}{n} \left(nA + B \sum_{i=1}^n x_i \right) = A + B\bar{x}
\end{aligned} \tag{3.7}$$

由此可得到

$$\begin{aligned}
E(b) &= E\left[\frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}\right] \\
&= \frac{\sum_{i=1}^n (x_i - \bar{x}) E(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \\
&= \frac{\sum_{i=1}^n (x_i - \bar{x})(A - Bx_i - A + B\bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \\
&= B \cdot \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} = B
\end{aligned} \tag{3.8}$$

$$E(a) = E(\bar{y} - b\bar{x}) = A + B\bar{x} - B\bar{x} = A \tag{3.9}$$

从 3.8、3.9 式可说明 a, b 是 A, B 无偏估计值,

又由于

$$E(\hat{y}_i) = E(a + bx_i) = A + Bx_i = Ey_i \tag{3.10}$$

从 3.10 式说明 $\hat{y}_i$ 是 Ey_i 的无偏估计值。

例 3.1 在某林分内随机抽取 6 块面积为 0.08 公顷的样地,测定了样地上平均树高 x_i 与每公顷平均断面积 y_i 后,得表 3.1 数据,试估计该林分每 0.08 公顷林地上每公顷平均断面积 y_i 对于平均高 x_i 的总体回归直线方程。

表 3.1

样地号	1	2	3	4	5	6
林分高 $x_i(m)$	20	22	24	26	28	30
断面积 $y_i(m^2)$	24.3	26.5	28.7	30.5	31.7	32.9

解 列计算表 3.2

表 3.2

$x(m)$	$y(m^2)$	x^2	xy
20	24.3	400	486
22	26.5	484	583
24	28.7	576	688.8
26	30.5	676	793
28	31.7	784	887.6
30	32.9	900	987
150	174.6	3820	4425.4

从 3.2 表有关数据,直接计算:

$$\bar{x} = \frac{150}{6} = 25$$

$$\bar{y} = \frac{174.6}{6} = 29.1$$

$$\begin{aligned}
 b &= \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} \\
 &= \frac{4425.4 - 6 \times 25 \times 29.1}{3820 - 6 \times 25^2} \\
 &= 0.863
 \end{aligned}$$

$$\bar{a} = \bar{y} - b\bar{x} = 29.1 - 0.863 \times 25 = 7.53$$

所以估计的回归直线方程为

$$\hat{y} = 7.53 + 0.863x$$

二、直线回归的显著性检验

1. 平方和分解

$$\text{令 } L_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2$$

那么

$$\begin{aligned}
 L_{yy} &= \sum_{i=1}^n [(y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})]^2 \\
 &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \\
 &\quad + 2 \sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y})
 \end{aligned} \tag{3.11}$$

可以证明

$$\sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = 0$$

$$\text{则 } \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \tag{3.12}$$

若令 $u = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$, $Q = \sum_{i=1}^n (y_i - \hat{y}_i)^2$, u 称为回归平方和, Q 称为剩余平方和。

于是有

$$L_{yy} = u + Q \quad (3.13)$$

$$\begin{aligned} u &= \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = \sum_{i=1}^n (a + bx_i - \bar{y})^2 \\ &= \sum_{i=1}^n (\bar{y} - b\bar{x} + bx_i - \bar{y})^2 \\ &= b^2 \sum_{i=1}^n (x_i - \bar{x})^2 = b^2 L_{xx} \end{aligned} \quad (3.14)$$

即 $u = b^2 L_{xx}$

又由于 $b = \frac{L_{xy}}{L_{xx}}$

$$\begin{aligned} \text{所以 } u &= b \cdot \frac{L_{xy}}{L_{xx}} \cdot L_{xx} = b L_{xy} \\ &= \frac{L_{xy}^2}{L_{xx}} \end{aligned}$$

回归平方和 u 是由 x 变化引起的, 它的大小反映了 Y 对 X 依赖程度。

从 3.13 式, 还可得到

$$Q = L_{yy} - u$$

若 $Q=0$, 那么 $u=L_{yy}$, u 取最大值, y 的变动完全由 x 所引起, 并且 $y_i = \hat{y}_i$, 观察值与理论值完全拟合, 此时 y 对 x 的关系成为确定性的函数关系。若 $u=0$, 此时 $\hat{y}_i = \bar{y}$, 说明 $\hat{y}_i$ 对 x 的变动毫无响应。

2. 显著性检验

在已知条件为样本抽取方式是重复抽样, 总体两变量 X 和 Y 存在线性回归关系, 总体在变量 Y 上的各条件概率分布是正态分布且条件方差是齐性的。

直线回归 $Y = A + BX + \epsilon$ 是否显著可以用 B 是否不为 0

来检验。这意味着 Y 不随 X 改变,即不存在 Y 对 X 的回归关系。

(1)建立统计假设 $H_0: B=0, H_1: B \neq 0$

(2)计算统计量 F

$$F = \frac{u/1}{Q/(n-2)}$$

统计量 F 遵循 F 分布,第一自由度 $df_1=1, df_2=n-2$ 。

具体计算可采用下列公式

$$u = bL_{xy}$$

$$L_{xy} = \sum_{i=1}^n x_i y_i - \frac{1}{n} \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)$$

$$Q = L_{yy} - u$$

自由度: $df_y = df_u + df_Q$

df_y ——总自由度, df_u ——回归平方和项的自由度, df_Q ——回归剩余平方和项的自由度,则 $df_y = n-1, df_u = 1, df_Q = n-2$ 。

(3)根据给定 α 值及自由度 $df_1 = df_u = 1, df_2 = df_Q = n-2$ 查 F 分布表得 F_{α, df_1, df_2} 值。

(4)若 $F > F_{\alpha}(df_1, df_2)$,则否定假设 H_0 ,即认为线性回归关系是显著的。

(5)列方差分析表

表 3.3

变差来源	自由度	平方和	均方	F 值	F _α
回归	1	$u = bL_{xy}$	$S_u^2 = \frac{u}{1}$	$F = \frac{\frac{u}{1}}{\frac{Q}{n-2}}$	$F_{\alpha}(1, n-2)$
剩余	$n-2$	$Q = L_{yy} - u$	$S_Q^2 = \frac{Q}{n-2}$		
共计	$n-1$	L_{yy}	—	—	—

例 3.2 根据例 3.1 资料进行线性回归显著性检验。

解 (1) 建立 $H_0: B=0, H_1: B \neq 0$

(2) 已知 $\hat{y} = 7.53 + 0.863x$

$$\sum_{i=1}^n x_i = 150, \quad \sum_{i=1}^6 y_i = 174.6, \quad \sum_{i=1}^n x_i y_i = 4425.4$$

$$\sum_{i=1}^6 y_i^2 = 5133.98, \quad n=6$$

$$u = bL_{xy}$$

$$= 0.863 \left[\sum_{i=1}^6 x_i y_i - \frac{1}{6} \left(\sum_{i=1}^6 x_i \right) \left(\sum_{i=1}^6 y_i \right) \right]$$

$$= 0.863 \left[4425.4 - \frac{1}{6} (150)(174.6) \right] = 52.12$$

$$L_{yy} = \sum_{i=1}^6 (y_i - \bar{y})^2 = \sum_{i=1}^6 y_i^2 - \frac{1}{n} \left(\sum_{i=1}^6 y_i \right)^2$$

$$= 5133.98 - \frac{1}{6} (174.6)^2$$

$$= 53.12$$

$$Q = L_{yy} - u$$

$$= 53.12 - 52.12$$

$$= 1.00$$

(3) 列方差分析表

表 3.4

变差来源	自由度	平方和	均方	F 值	F _α
回归	1	52.12	52.12	$F = 208.48$	$F_{1,4}^{0.01} = 21.2$
剩余	4	1.00	0.25		
总和	5	53.12	—	—	—

根据表 3.4, 认为回归是显著的。

3. 相关系数

(1) 总体相关系数

从平方和分解公式得知因变量 Y 在总体上的离差平方和也可分解为

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad (3.15)$$

$$\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

令 $\sigma_s^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$ 称为总体线性回归剩余方差。

$\sigma_h^2 = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - \bar{y})^2$ 称为总体线性回归方差。

$\sigma_y^2 = \frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2$ 称为总体因变量 y 方差。

于是 $\sigma_y^2 = \sigma_s^2 + \sigma_h^2$

由此可看到, 总体在变量 Y 上的差异是由两方面原因引起的, 一方面是由于各 Y_i 所对应的 X_i 不同引起的差异, 用 σ_h^2 表示, 别一方面是其他因素所引起的, 我们把其他因素引起差异看作是随机差异引起的, 用 σ_s^2 表示。

如果 $\sigma_s^2 < \sigma_h^2$, 则说明 $\hat{Y}_i$ 都很靠近 $Y = \bar{Y}$ 直线, 此时随机差异 σ_s^2 值愈大。

如果 $\sigma_u^2 > \sigma_y^2$ 时, 则说明各 $\hat{Y}_i$ 值离 $Y = \bar{Y}$ 直线远, 而且 σ_y^2 值小, 表示线性回归关系显著。

因此, 用 $\rho^2 = \frac{\sigma_u^2}{\sigma_y^2}$ (3.16)

作为判断线性回归关系显著性指标, 并把这个指标称为总体相关系数。

$$\begin{aligned}\rho &= \sqrt{\frac{\sigma_u^2}{\sigma_y^2}} \\ &= \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \\ &= \frac{\text{cov}(x, y)}{\sigma_x \sigma_y} \quad (3.17)\end{aligned}$$

如果 $0 \leq |\rho| \leq 1$, 由于 σ_u^2 愈大, 则 ρ^2 愈大, 所以 $|\rho|$ 愈接近于 1, 便说明两变量线性关系愈紧密, 反之 $|\rho|$ 愈接近于 0, 表明两变量的线性相关关系愈不紧密。

当 $\rho > 0$ 时, 则 $\hat{Y}$ 随 X 增大而增大, 称为正相关; 当 $\rho < 0$ 时, 则 $\hat{Y}$ 随 X 的增大而减小称为负相关。

(2) 样本相关系数

样本相关系数用 r 表示。它是总体相关系数 ρ 的估计值。假定样本容量为 n , 样本回归剩余方差记为 S_e^2 , 样本因变量 y 的方差记为 S_y^2 。那么

$$r = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{L_{xy}}{\sqrt{L_{xx} \cdot L_{yy}}}$$

$$= \pm \sqrt{\frac{u}{L_{yy}}} \quad (3.18)$$

由于样本回归直线中 $b = \frac{L_{xy}}{L_{xx}}$, $L_{xx} > 0$, 故 b 与 L_{xy} 同号,

$r = \frac{L_{xy}}{\sqrt{L_{xx} \cdot L_{yy}}}$, 所以 r 同 L_{xy} 同号。因此 r, b, L_{xy} 三者取同一符号。

例 3.3 根据例 3.1 样本资料求样本相关系数。

解
$$r = \frac{L_{xy}}{\sqrt{L_{xx} \cdot L_{yy}}}$$

$$\begin{aligned} L_{xy} &= \sum_{i=1}^n xy - \frac{1}{n} \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right) \\ &= 4425.4 - \frac{1}{6} (150)(174.6) \\ &= 60.4 \end{aligned}$$

$$\begin{aligned} L_{xx} &= \sum_{i=1}^n x^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \\ &= 3820 - \frac{1}{6} (150)^2 \\ &= 70 \end{aligned}$$

$$\begin{aligned} L_{yy} &= \sum_{i=1}^n y_i^2 - \frac{1}{n} \left(\sum_{i=1}^n y_i \right)^2 \\ &= 5133.98 - \frac{1}{6} (174.6)^2 \\ &= 53.12 \end{aligned}$$

所以

$$\begin{aligned} r &= \frac{L_{xy}}{\sqrt{L_{xx} \cdot L_{yy}}} \\ &= \frac{60.4}{\sqrt{70 \times 53.12}} \end{aligned}$$

$$=0.991$$

4. 相关系数显著性检验

回归直线应用有无意义,主要靠专业知识,靠实践经验,但是在统计学中也给出一些辅助办法,如文中介绍的显著性检验。这里再介绍相关系数显著性检验。它依 $|r|$ 值大小来判断,若 $|r| \rightarrow 1$,则总体两变量 X, Y 的线性关系越紧密。若 $|r| \rightarrow 0$,则总体两变量 X 和 Y 没有线性关系。

在实际中会发现,当 n 较小时,而求得 $|r|$ 并不小(例如0.80),但样本 $(x_i, y_i) (i=1, 2, \dots, n)$ 可能来自总体相关系数 $\rho=0$ 的总体,所以不能根据一个很小样本求得 $|r|$ 值较大时而轻易下结论认为 X 与 Y 的线性相关很密切。此时必需对 r 进行差异显著检验。

检验方法与步骤

(1)建立统计假设: $H_0: \rho=0, H_1: \rho \neq 0$

(2)从总体中重复地抽取 n 个单元组样本,并根据样本资料 $(x_i, y_i) (i=1, 2, \dots, n)$,求出 r 值。

(3)根据给定 α 值及自由度 $df=n-2$,查检验相关系数 $\rho=0$ 的临界值(r_{α})值,得到 $r_{\alpha}, (n-2)$ 值。

(4)若 $|r| > r_{\alpha}, (n-2)$,则否定假设 H_0 ,而认为建立直线回归方程有意义,反之 $|r| < r_{\alpha}, (n-2)$ 时,则接受假设 H_0 ,说明所建立直线方程无实际意义。

在二维正态总体中, r 的显著性检验与回归系数显著性检验相一致。

当 $B=0$ 时,推出 $\rho=0$,而在 $\rho=0$ 假设下,还可以作 r 的 t 检验。

$$t^2 = \frac{\frac{b^2 L_{xx}}{Q}}{n-2} = \frac{r^2 L_{yy} (n-2)}{L_{yy} - u} = \frac{r^2 L_{yy} (n-2)}{L_{yy} - r^2 L_{yy}}$$

$$= \frac{r^2}{\frac{1-r^2}{n-2}}$$

$$\text{那么 } t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} \quad (3.19)$$

如果 $|t| \geq t_{\alpha}(n-2)$ 时, 直线回归显著, 推翻 $\rho=0$ 的假设。

例 3.4 对例 3.2 的相关系数进行检验。

解 (1) 建立统计假设 $H_0: \rho=0$

(2) 已知 $n=6, r=0.991, \hat{y}=7.53+0.863x$

(3) 取 $\alpha=0.01, df=n-2=4$, 查相关系数 $\rho=0$ 的检验表, 得 $r_{0.01}(4)=0.91720$

因 $r=0.991 > r_{0.01}(4)=0.9172$, 否定假设 H_0 , 认为总体两变量 X 和 Y 线性相关系数极端显著。

5. 利用回归方程进行预报

若回归方程拟合得很好且具有实际意义, 那就可以根据回归方程进行预报。所谓预报问题, 就是进行估计。人们普遍关心的不仅是具体估计值, 更关心这种预报在一定概率保证下能产生多大误差。

(1) 总体条件平均数的抽样估计

用 $\bar{Y}_{x_0}$ 表示自变量为 x_0 条件下的总体条件平均数。设双变量总体回归方程为直线方程

$$\bar{y}_x = A + Bx$$

若从总体中抽取 n 个单元组成样本, 其各观察值为 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, 根据这样本资料可得样本的回归方程为

$$\hat{y} = a + bx$$

式中参数 a, b 用前面叙述的最小二乘法计算得到。前面

已证明 $E(b) = B, E(a) = A$, 说明 a 可以作为 A 的估计值, b 可以作为 B 的估计值。

对于自变量的某一特定值 x_0 , 样本回归直线方程的 $\hat{y}_{x_0}$ 为总体相应条件平均数 $\bar{Y}_{x_0}$ 的无偏估计。因为

$$\begin{aligned} E(\hat{y}_{x_0}) &= E(a + bx_0) = E(a) + x_0 E(b) \\ &= A + Bx_0 = \bar{Y}_{x_0} \end{aligned}$$

所以 $\hat{y}_{x_0}$ 是 $\bar{Y}_{x_0}$ 的无偏估计, 可以用 $\hat{y}_0$ 所取得数值作为 $\bar{Y}_{x_0}$ 的估计值。

设对于 X 的任何数值 x_i , 因变量的条件方差都等于 σ_{δ}^2 , 那么有

$$\begin{aligned} \sigma^2(b) &= \sigma^2 \left[\frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \right] \\ &= \frac{1}{\left[\sum_{i=1}^n (x_i - \bar{x})^2 \right]^2} \sigma^2 \left[\sum_{i=1}^n (x_i - \bar{x}) y_i \right] \\ &= \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \sigma^2(y_i)}{\left[\sum_{i=1}^n (x_i - \bar{x})^2 \right]^2} \\ &= \frac{\sigma_{\delta}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \end{aligned} \quad (3.20)$$

$$\begin{aligned} \sigma^2(\bar{y}) &= \sigma^2 \left[\frac{\sum_{i=1}^n y_i}{n} \right] = \frac{1}{n^2} \sum_{i=1}^n \sigma^2(y_i) \\ &= \frac{1}{n^2} n \sigma_{\delta}^2 = \frac{\sigma_{\delta}^2}{n} \end{aligned} \quad (3.21)$$

由式 3.20 和 3.21 可得

$$\begin{aligned}
 \sigma^2(\hat{y}_{x_0}) &= \sigma^2[a + bx_0] \\
 &= \sigma^2[(\bar{y} - b\bar{x}) + bx_0] \\
 &= \sigma^2[\bar{y} + b(x_0 - \bar{x})] \\
 &= \sigma^2(\bar{y}) + (x_0 - \bar{x})^2 \sigma^2(b) \\
 &= \frac{\sigma_b^2}{n} + \frac{(x_0 - \bar{x})^2 \sigma_b^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \\
 &= \frac{\sigma_b^2}{n} \left[1 + \frac{n(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right] \quad (3.22)
 \end{aligned}$$

由于 y_i 都遵循正态分布的随机变量, 所以根据统计学有关定理, 可以推论 $\hat{y}_i$ 也是遵循正态分布的随机变量。

$$\hat{y}_0 \sim N \left[\bar{Y}_{x_0}, \sqrt{\frac{\sigma_b^2}{n} \left[1 + \frac{n(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right]} \right]$$

所以以 $\hat{y}_0$ 作为 $\bar{Y}_{x_0}$ 的估计值时, 误差限为

$$\Delta(\hat{y}_0) = U_{\alpha} \sigma_b \sqrt{\frac{1}{n} + \frac{n(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} \quad (3.23)$$

可靠性为 $P = 1 - \alpha$ 。

通常 3.23 式中 σ_b 未知, 可用它的估计值来代替。这时以 $\hat{y}_0$ 作为 $\bar{y}_{x_0}$ 的估计值的误差限为

$$\Delta(\hat{y}_0) = \frac{t_{\alpha} S_b}{\sqrt{n-2}} \sqrt{1 + \frac{n(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} \quad (3.24)$$

可靠性为 $P = 1 - \alpha$, 式中 t_{α} 是根据给定 α 值及自由度 $df = n - 2$, 查 t 分布表得到的。

$$S_e = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}$$

由 3.24 式可看出用 $\hat{y}_0$ 去估计 $\bar{y}_{x_0}$ 时其误差与下列方面有关。

(1) 与 α 取值有关, α 愈小, $\Delta(\hat{y}_0)$ 也就相应愈大。

(2) 与 n 有关, n 愈大, $\Delta(\hat{y}_0)$ 相应愈小。

(3) 与观察值 x_0 有关, 当 x_0 靠近 $\bar{x}$ 时, $\Delta(\hat{y}_0)$ 相应的减小, 当 x_0 远离 $\bar{x}$ 时, $\Delta(\hat{y}_0)$ 就相应的增大, 若分别作出函数

$$\begin{cases} y = \hat{y}_0 + \Delta(\hat{y}_0) \\ y = \hat{y}_0 - \Delta(\hat{y}_0) \end{cases}$$

其图形见图 3.2

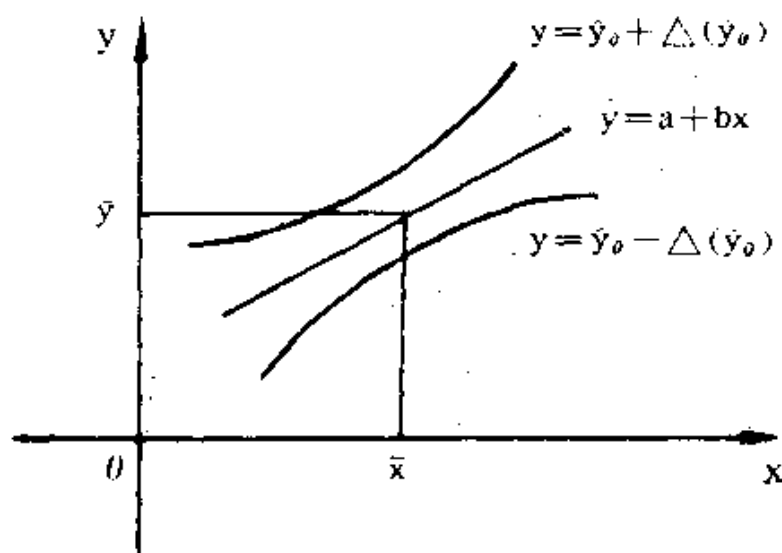


图 3.2

估计精度

这种估计的相差误差限 $\Delta'(\hat{y}_0)$ 为

$$\Delta'(\hat{y}_0) = \frac{\Delta(\hat{y}_0)}{\hat{y}_0} \quad (3.25)$$

$$\text{估计精度 } Pc(\hat{y}_0) = [1 - \Delta'(\hat{y}_0)] \times 100\% \quad (3.26)$$

在研究其误差限 $\Delta(\hat{y}_0)$ 中, 应用总体回归剩余方差 σ^2 的

估计值 $S_b^2 = \frac{Q}{n-2}$, S_b^2 称为样本回归剩余方差。下面证明

$$E(S_b^2) = \sigma^2$$

$$\text{因为 } Q = L_{yy} - u = L_{yy} - b^2 L_{xx}$$

$$\begin{aligned} E(L_{yy}) &= E\left[\sum_{i=1}^n y_i^2 - n\bar{y}^2\right] = \sum_{i=1}^n E(y_i^2) - nE(\bar{y})^2 \\ &= \sum_{i=1}^n [\sigma_y^2 + E(y_i)^2] - n[\sigma_{\bar{y}}^2 + E(\bar{y})^2] \\ &= \sum_{i=1}^n [\sigma^2 + (A + Bx_i)^2] - n\left[\frac{\sigma^2}{n} + (A + B\bar{x})^2\right] \\ &= n\sigma^2 + \sum_{i=1}^n (A + Bx_i)^2 - \sigma^2 - n(A + B\bar{x})^2 \\ &= (n-1)\sigma^2 + \sum_{i=1}^n (A + Bx_i)^2 - n(A + B\bar{x})^2 \\ &= (n-1)\sigma^2 + B^2\left(\sum_{i=1}^n x_i^2 - n\bar{x}^2\right) \\ &= (n-1)\sigma^2 + B^2 L_{xx} \\ E(u) &= E(b_1 L_{xx}) \\ &= [\sigma_b^2 + E(b)^2] L_{xx} \\ &= \left(\frac{\sigma^2}{L_{xx}} + B^2\right) L_{xx} \end{aligned}$$

所以

$$\begin{aligned} E(Q) &= E(L_{yy} - u) \\ &= E(L_{yy}) - E(u) \\ &= (n-2)\sigma^2 \end{aligned}$$

因此

$$E\left(\frac{Q}{n-2}\right) = \sigma^2 \quad (3.27)$$

对总体个别单元在因变量 Y 上的数值的估计,这在实际工作中也是常用的,这里只作简单的叙述。

y_0 的估计值仍然用 $\hat{y}_0$ 表示。其误差限

$$\Delta(\hat{y}_0 - y) = t_\alpha \frac{S_\varepsilon}{\sqrt{n-2}} \sqrt{1 + n + \frac{n(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} \quad (3.28)$$

例 3.5 根据例 3.1 资料,估计 $x_0 = 24.5$ 时相应的 Y_{x_0} 及 y_0 的估计值,取 $\alpha = 0.05$ 。

解 已知 $\hat{y} = 7.53 + 0.863x$, $n = 6$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^6 x_i = \frac{150}{6} = 25$$

$$\sum_{i=1}^6 (x_i - \bar{x})^2 = 70$$

当 $x_0 = 24.5$ 时

$$\hat{y}_0 = 7.53 + 0.863 \times 24.5 = 28.6735$$

根据 $\alpha = 0.05$ 及自由度 $df = n - 2 = 4$,查 t 分布表得

$$t_{0.05}(4) = 2.776$$

由例 3.3 已求得 $Q = 1.00$

$$\text{故 } S_\varepsilon = \sqrt{\frac{Q}{n}} = \sqrt{\frac{1.00}{6}} = 0.408$$

$$\text{而 } \frac{S_\varepsilon}{\sqrt{n-2}} = \frac{0.408}{\sqrt{4}} = 0.204$$

以 $\hat{y}_0 = 28.6735$ 作为当 $x_0 = 24.5$ 时相应的 Y_{x_0} 的估计值时的误差限 $\Delta(\hat{y}_0)$ 为

$$\Delta(\hat{y}_0) = t_\alpha \frac{S_\varepsilon}{\sqrt{n-2}} \sqrt{1 + \frac{n(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

$$= 2.776 \times 0.204 \sqrt{1 - \frac{6 \times (24.5 - 25)^2}{70}}$$

$$= 0.572$$

$\bar{Y}_{x_0}$ 的估计区间为 (28.102, 29.246)

$$Pc(\hat{y}_0) = (1 - \frac{0.572}{28.6735}) \times 100\%$$

$$= 98.0\%$$

以 $\hat{y}_0 = 28.6735$ 作为当 $x_0 = 24.5$ 时相应的 y_0 估计值的误差限 $\Delta(\hat{y}_0 - y_0)$ 为

$$\Delta(\hat{y}_0 - y_0) = t_{\alpha} \frac{S_{\hat{y}}}{\sqrt{n-2}} \sqrt{1 + n + \frac{n(x_0 - \bar{x})^2}{\sum_{j=1}^n (x_j - \bar{x})^2}}$$

$$= 2.776 \times 0.204 \sqrt{1 + 6 + \frac{6(24.5 - 25)^2}{70}}$$

$$= 1.501$$

y_0 的估计区间为 (27.173, 30.175), 估计精度

$Pc(\hat{y}_0 - y_0)$ 为

$$Pc(\hat{y}_0 - y_0) = (1 - \frac{1.501}{28.6735}) \times 100\%$$

$$= 95\%$$

第二节 一元非线性回归

在实际应用中, 回归关系当然不会局限于一元一次, 即一元线性情形, 如材积对于胸径的回归形式常用 $\bar{Y}_x = AX^b$, 这是幂函数回归的形式, 树高对于胸径的关系常取二次抛物线回归形式 $\bar{Y}_x = A + BX + CX^2$, 而林分立木株数对于林龄的关

系则往往取双曲线回归的形式 $\bar{Y}_x = A + \frac{B}{X}$ 。这些都是一元非线性回归问题,现在讨论一些常见的一元非线性回归问题。

一、幂函数曲线

$$\bar{Y}_x = AX^b \quad (3.29)$$

估计该方程为

$$\hat{y} = ax^b \quad (3.30)$$

曲线图形:

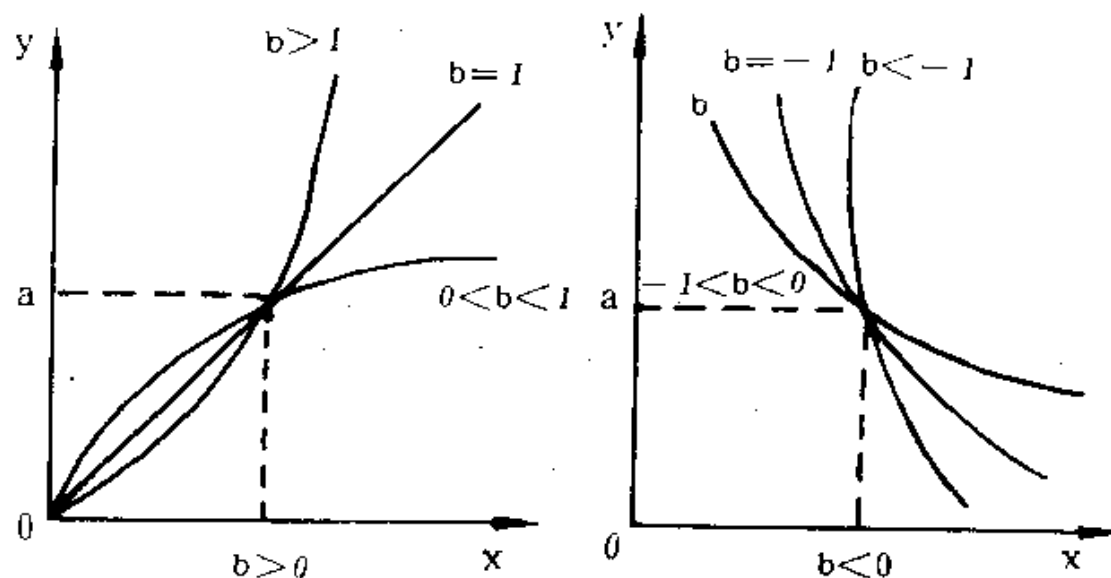


图 3.3

该幂函数曲线在林业科研及生产实践中应用是非常广泛的。例如,常用于编制一元立木材积表,生长率、生长量表,编

制林分密度控制图等。

为了应用最小二乘法确定方程中参数 a, b , 将曲线直线化。将式 3.30 两端取对数, 得

$$\lg \hat{y} = \lg a + b \lg x$$

$$\text{令 } y' = \lg \hat{y}, a' = \lg a, x' = \lg x$$

$$\text{则 } y' = a' + bx' \quad (3.31)$$

按最小二乘法确定 a' 与 b 时, 应解下列一组标准方程

$$\begin{cases} na' + b \sum x' = \sum y' \\ a' \sum x' + b \sum x'^2 = \sum x' y' \end{cases} \quad (3.32)$$

$$a' = \bar{y}' - b\bar{x}'$$

$$b = \frac{\sum (x' - \bar{x}')(y' - \bar{y}')}{\sum (x' - \bar{x}')^2}$$

解得 a', b 之后再由 a' 查反对数得 a 值。

例 3.6 设下列资料系 9 株柞树的胸径与相应材积数据, 试估计材积对于胸径的幂函数回归曲线方程 $V = aD^b$ 中的参数 a 与 b 。

解 对 $V = aD^b$ 两边取对数得

$$\lg V = \lg a + b \lg D$$

$$\text{令 } \lg V = V' \quad \lg a = a' \quad \lg D = D'$$

则上式成为

$$V' = a' + bD'$$

式中 a', b 可用下式解

$$a' = \bar{V}' - b\bar{D}'$$

$$b = \frac{\sum (D' - \bar{D}') (V' - \bar{V}')}{\sum (D' - \bar{D}')^2}$$

$$= \frac{l_{DV}}{l_{D'D}}$$

$$l_{DV} = \sum D' V' - \frac{1}{n} (\sum D') (\sum V')$$

$$l_{D'D} = \sum D'^2 - \frac{1}{n} (\sum D')^2$$

$$\bar{V}' = \frac{1}{n} \sum V' \bar{D}'$$

$$= \frac{1}{n} \sum D' V' \bar{D}'$$

表 3.5

胸径 D_i (cm)	材积 V_i (m ³)
3.0	0.00224
4.0	0.00472
4.9	0.00667
5.9	0.01029
7.1	0.01625
7.9	0.01957
9.1	0.02588
9.8	0.02658
11.1	0.04313

V' 与 D' 均系对数值, 为此列计算表

表 3.6

$D'_i = \lg D_i$	$V'_i = \lg V_i$	D_i^2	$D' V'$	V'^2
0.4771	-2.6498	0.2276	-1.2642	7.0214
0.6021	-2.3261	0.3625	-1.4005	5.4107
0.6902	-2.1759	0.4764	-1.5018	4.7345
0.7709	-1.9876	0.5943	-1.5322	3.9506
0.8513	-1.7891	0.7247	-1.5231	3.2009
0.8976	-1.7084	0.8057	-1.5335	2.9186
0.9590	-1.5870	0.9197	-1.5219	2.5186
0.9912	-1.5754	0.9825	-1.5615	2.4819
1.0453	-1.3652	1.0927	-1.4270	1.8638
7.2847	-17.1645	6.1861	-13.2657	34.1010

$$n=9, \sum_{i=1}^9 D'_i = 7.2847 \quad \bar{D}' = \frac{1}{9} \times 7.2847 = 0.8094$$

$$\sum_{i=1}^9 V'_i = -17.1645 \quad \bar{V}' = \frac{1}{9}(-17.1645) \\ = -1.9072$$

$$l_{DV'} = \sum_{i=1}^9 D'_i V'_i - \frac{1}{n} \left(\sum_{i=1}^9 D'_i \right) \left(\sum_{i=1}^9 V'_i \right) \\ = -13.2657 - \frac{1}{9} (7.2847)(-17.1645) = 0.6274$$

$$l_{DD'} = \sum_{i=1}^9 (V'_i - \bar{V}')^2 = \sum_{i=1}^9 V'^2_i - \frac{1}{n} \left(\sum_{i=1}^9 V'_i \right)^2 \\ = 34.1010 - \frac{1}{9} (-17.1645)^2 = 1.3654$$

$$b = \frac{l_{DV'}}{l_{DD'}} = \frac{0.6274}{0.2898} = 2.1649$$

$$a' = \bar{V}' - b\bar{D}' \\ = -1.9072 - 2.1649 \times 0.8094 \\ = -3.6595 \\ = -4.3405$$

$$a = 0.0002190 \quad (a' \text{ 的反对数值})$$

材积对胸径的回归值为

$$V = 0.0002190 D^{2.1649}$$

设总体在材积与胸径对数的双变量频率分布近似二维正态分布, 现求 $\lg V$ 与 $\lg D$ 之间的相关系数 r

$$r = \frac{l_{DV'}}{\sqrt{l_{DD'} l_{VV'}}} = \frac{0.6274}{\sqrt{0.2898 \times 1.3654}} = 0.997$$

自由度为 $f = n - 2 = 9 - 2 = 7$, $\alpha = 0.05$ 查附表 11 得

$Y_{0.05,7} = 0.6664$, 因 $r = 0.997 > 0.6664$, 故线性回归关系是显著的。

二、指数函数曲线

$$y = ae^{bx} \quad (a > 0) \quad (3.33)$$

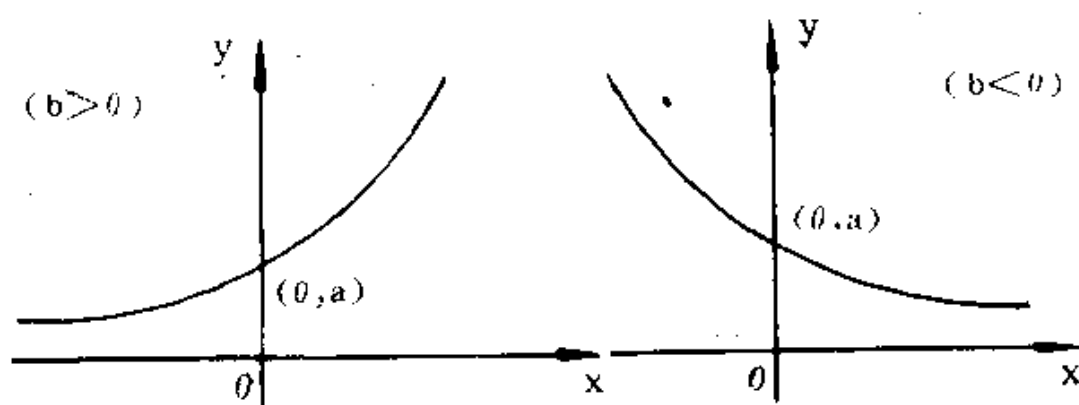


图 3.4

此方程可用于表示害虫的产卵数与温度之间关系。

对 3.33 式两端取对数得

$$\ln y = \ln a + bx$$

令 $y = \ln y$ $a' = \ln a$,

则 $y' = a' + bx$

即成为直线关系形式

三、双曲线

$$1. y = a + \frac{b}{x} \quad (3.34)$$

令 $x' = \frac{1}{x}$ 则成为

$y = a + bx'$ 的直线关系形

式。

常用于年龄(X)与林分平均高(Y)之间关系，或形数(Y)与胸径(X)之间关系。

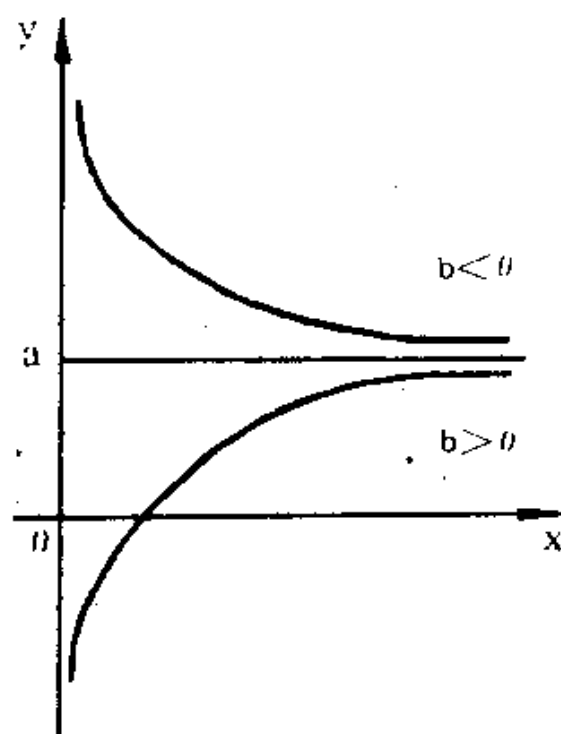


图 3.5

$$2. \frac{1}{y} = a + \frac{b}{x} \quad (3.35)$$

3.35 式中令 $x' = \frac{1}{x}$,

$x' = \frac{1}{x}$ 即成为 $y' = a$

$-bx'$ 的直线关系形式,

我们曾用此方程

作为编制福建闽北

杉木中心产区林分

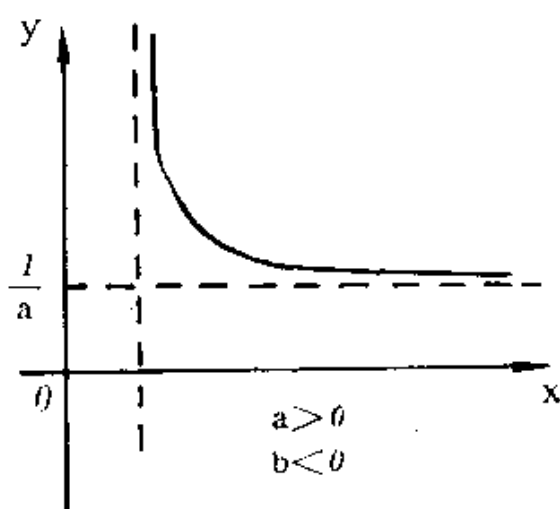
密度控制图,等树高

线的数学模型, M^{-1}

$= a + bN^{-1}$, 式中 N

为密度(株数), M 为

单位面积蓄积量。



四、对数函数曲

线

$$y = a + b \lg x$$

(3.36)

令 $x' = \lg x$ 即成

为 $y = a + bx'$ 的直线

关系形式,此方程可

用来编制杉木或马

尾松地位指数表用,

表明年龄(X)与优势

木平均高(Y)之间的

关系。

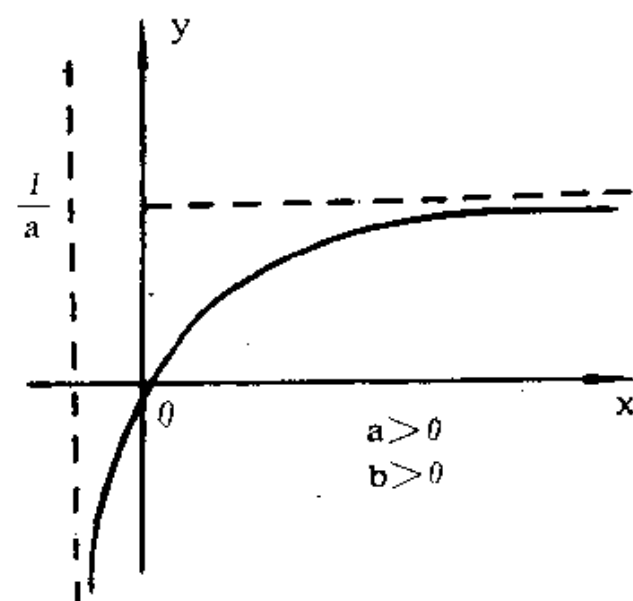


图 3.6

五、多项式回归方程

以上讨论的各种回归问题都要求我们事先确定回归方程的类型,这一点并不总是容易实现的,因此在不易对回归方程的类型作出判断时,常采用多项式进行逼近,因为,任意曲线都可以近似地用多项式表示,多项式的一般形式为

$$y = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$$

下面只讨论二次、三次抛物线回归方程中参数的最小二乘估计方法。

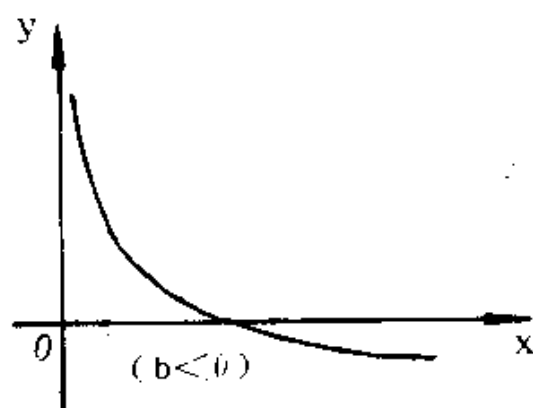
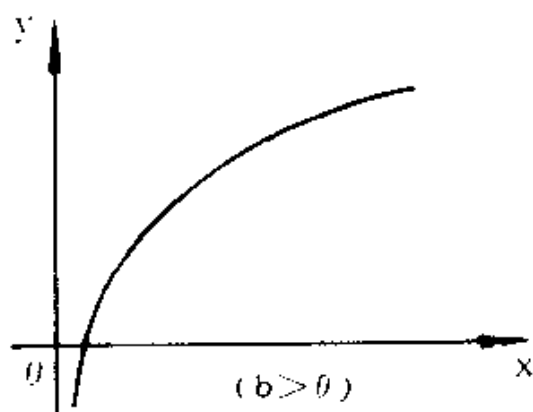


图 3.7

1. 二次抛物线回归

$$\hat{y} = a + bx + cx^2 \quad (3.37)$$

设样本资料为

$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, 使

$$Q = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \text{ 取最小值}$$

其中 $\hat{y}_i = a + bx_i + cx_i^2$

于是

$$\begin{aligned} Q &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\ &= \sum_{i=1}^n [y_i - (a + bx_i + cx_i^2)]^2 \\ &= Q(a, b, c) \end{aligned}$$

回归方程 $\hat{y} = a + bx + cx^2$ 的参数 a, b, c 由方程组

$$\begin{cases} \frac{\partial Q}{\partial a} = 0 \\ \frac{\partial Q}{\partial b} = 0 \\ \frac{\partial Q}{\partial c} = 0 \end{cases} \quad (3.38)$$

确定。对 $Q = Q(a, b, c)$ 分别求三个一阶偏导数, 代入 3.38 式得 3.39 式

$$\begin{cases} na + b \sum x + c \sum x^2 = \sum y \\ a \sum x + b \sum x^2 + c \sum x^3 = \sum xy \\ a \sum x^2 + b \sum x^3 + c \sum x^4 = \sum x^2 y \end{cases} \quad (3.39)$$

解 3.39 方程组即可得到 a, b, c 。

例 3.7 1975 年, 在福建大田县桃源林场场部工区调查了 18 年生杉木林分, 并设置标准地选取优势木实测其平均高与年龄的资料如下表 3.7 所示, 试用最小二乘法估计二次抛物线回归方程。

表 3.7

年龄 x_i (年)	2	4	6	8	10	12	14	16	18
平均高 y_i (米)	0.50	1.95	4.10	5.85	7.60	9.60	11.60	13.35	14.43

解 本题中 $n=9$ 经计算得

$$\sum_{i=1}^9 x_i = 90, \sum_{i=1}^9 y_i = 68.98, \sum_{i=1}^9 x_i^2 = 1140, \sum_{i=1}^9 x_i^3 = 16200$$

$$\sum_{i=1}^9 x_i^4 = 245328, \sum_{i=1}^9 x_i y_i = 907.14, \sum_{i=1}^9 x_i^2 y_i = 13064.12$$

由此得正规方程组

$$\begin{cases} 9a + 90b + 1140c = 68.98 \\ 90a + 1140b + 16200c = 907.14 \\ 1140a + 16200b + 245328c = 13064.12 \end{cases}$$

解得: $a = -1.691190$, $b = 0.9873474$

$$c = -0.004088203$$

于是所求二次抛物线回归方程为

$$\hat{y} = -1.691190 + 0.9873474x - 0.004088203x^2$$

与一元线性回归相似,可以定义二次抛物线回归乘余标准差

$$S_s = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - a - bx_i - cx_i^2)^2} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

但实践中常常用修正的样本回归剩余标准差

$$S'_s = \sqrt{\frac{n}{n-3}} S_s = \sqrt{\frac{1}{n-3} \sum_{i=1}^n (y_i - a - bx_i - cx_i^2)^2} \quad (3.40)$$

用它来描写样本各单元值对样本回归曲线的离散程度。

2. 三次抛物线回归

$$\hat{y} = a + bx + cx^2 + dx^3 \quad (3.41)$$

与二次抛物线回归相似,我们有正规方程组

$$\begin{cases} na + b \sum x + c \sum x^2 + d \sum x^3 = \sum y \\ a \sum x + b \sum x^2 + c \sum x^3 + d \sum x^4 = \sum xy \\ a \sum x^2 + b \sum x^3 + c \sum x^4 + d \sum x^5 = \sum x^2 y \\ a \sum x^3 + b \sum x^4 + c \sum x^5 + d \sum x^6 = \sum x^3 y \end{cases} \quad (3.42)$$

及

$$S'_e = \sqrt{\frac{1}{n-4} \sum_{i=1}^n (y_i - a - bx_i - cx_i^2 - dx_i^3)^2} \quad (3.43)$$

六、相关指数

描写总体两变量 X 与 Y 所存在曲线回归关系的紧密程度的特征数称为相关指数。或称回归指数。

1. 总体相关指数

$$\theta = \sqrt{1 - \frac{\sigma_{\hat{y}}^2}{\sigma_y^2}} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

式中 $\hat{Y} = f(X)$ 为曲线方程。相关指数恒取正值,且在 $[0,1]$ 区间内取值。

2. 样本相关指数

$$\theta' = \sqrt{1 - \frac{S_{\hat{y}}^2}{S_y^2}} = \sqrt{1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

若 $\hat{y} = a + bx + cx^2$, 其回归指数特称为样本二次抛物线回归指数, 记为 θ'_2 。

$$\sigma'_2 = \sqrt{1 - \frac{\sum_{i=1}^n (y_i - a - bx_i - cx_i^2)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (3.44)$$

第三节 多元回归

前面所介绍的是研究因变量 Y (随机变量) 对于一个自变量 X 的回归关系, 称为简回归。若研究因变量 Y 对于多个自变量 $X_i (i=1, 2, \dots, m)$ 的回归关系, 称为多元回归或复回归, 这种关系也是指 Y 的条件平均数 $\bar{Y}_{x_1, x_2, \dots, x_m}$ 与 $X_1, X_2, \dots, X_m$ 之间函数关系。在林业实践中可以遇到许多这样的关系。例如林木生长量与温度、湿度、光照……等多个因子有关; 林木的材积与胸径、树高、形数等因子有关; 森林害虫的发生量与温度、水分、冬季食物储量等有关。掌握 $\bar{Y}_{x_1, x_2, \dots, x_m}$ 与各因子 $X_i (i=1, 2, 3, \dots, m)$ 的关系可以预测林木生长, 编制立木材积表, 预报、预测害虫发生量等。

与讨论一元回归时相似, 我们将重点讨论多元线性回归问题, 至于非线性回归问题将简单提示一下。

一、多元线性回归

1. 回归方程的确定

由样本估计的多元线性回归方程的形式是

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_m x_m \quad (3.45)$$

上式称为经验复回归平面, 其中 $b_i (i=1, 2, \dots, m)$ 称为 y 对于 $x_i (i=1, 2, \dots, m)$ 的经验偏回归系数或称为样本偏回归系数, b_0 为常数项。

当 $m=2$ 时, 则

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 \quad (3.46)$$

上式称为二元线性经验回归方程或称为经验回归平面，其几何图形是三维空间里一个平面方程。

3.46 式中参数 b_0, b_1, b_2 用最小二乘法确定时 b_0, b_1, b_2 应满足下列一组正规方程

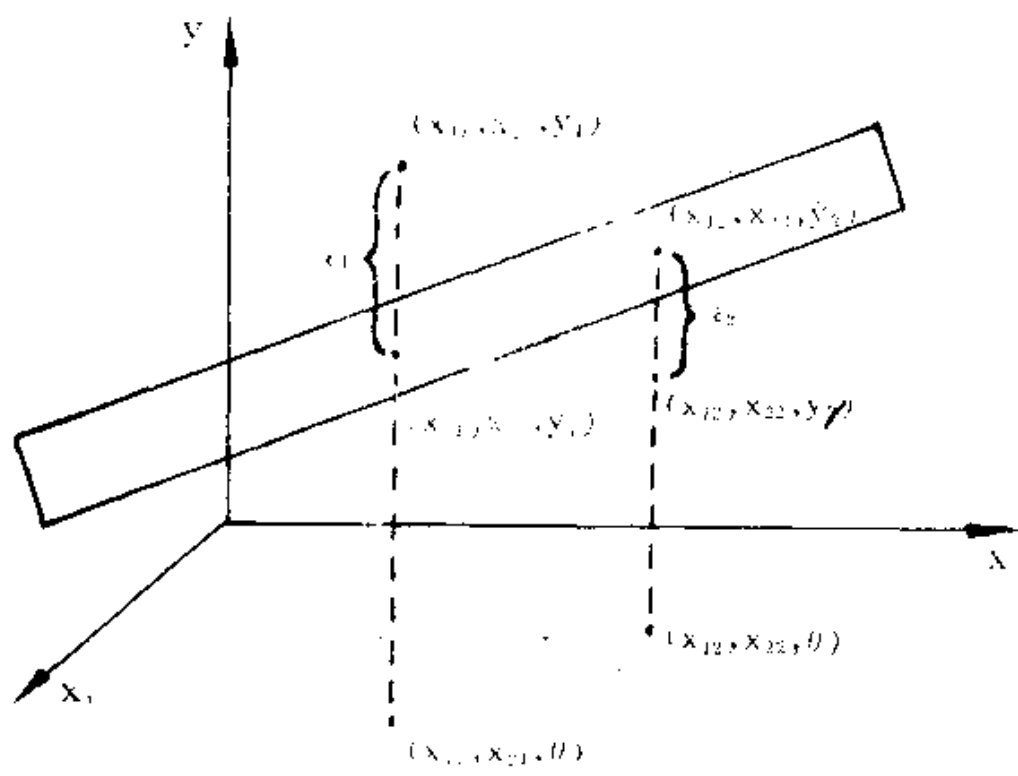


图 3.8

$$\begin{cases} nb_0 + b_1 \sum_{i=1}^n x_{1i} + b_2 \sum_{i=1}^n x_{2i} = \sum_{i=1}^n y_i \end{cases} \quad (3.47)$$

$$\begin{cases} b_0 \sum_{i=1}^n x_{1i} + b_1 \sum_{i=1}^n x_{1i}^2 + b_2 \sum_{i=1}^n x_{1i} x_{2i} = \sum_{i=1}^n x_{1i} y_i \end{cases} \quad (3.48)$$

$$\begin{cases} b_0 \sum_{i=1}^n x_{2i} + b_1 \sum_{i=1}^n x_{1i} x_{2i} + b_2 \sum_{i=1}^n x_{2i}^2 = \sum_{i=1}^n x_{2i} y_i \end{cases} \quad (3.49)$$

由 3.47 式得

$$b_0 = \bar{y} - b_1 \bar{x}_1 - b_2 \bar{x}_2 = \bar{y} - \sum_{i=1}^n b_i \bar{x}_i \quad (3.50)$$

而 3.48、3.49 式可写成

$$\begin{cases} l_{11}b_1 + l_{12}b_2 = l_{1y} \\ l_{21}b_1 + l_{22}b_2 = l_{2y} \end{cases} \quad (3.51)$$

$$\quad (3.52)$$

式中

$$l_{11} = \sum_{i=1}^n (x_{1i} - \bar{x}_1)^2 = \sum_{i=1}^n x_{1i}^2 - \frac{1}{n} \left(\sum_{i=1}^n x_{1i} \right)^2$$

$$l_{22} = \sum_{i=1}^n (x_{2i} - \bar{x}_2)^2 = \sum_{i=1}^n x_{2i}^2 - \frac{1}{n} \left(\sum_{i=1}^n x_{2i} \right)^2$$

$$l_{12} = l_{21} = \sum_{i=1}^n (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)$$

$$= \sum_{i=1}^n x_{1i}x_{2i} - \frac{1}{n} \left(\sum_{i=1}^n x_{1i} \right) \left(\sum_{i=1}^n x_{2i} \right)$$

$$l_{jy} = \sum_{i=1}^n (x_{ji} - \bar{x}_j)(y_i - \bar{y})$$

$$= \sum_{i=1}^n x_{ji}y_i - \frac{1}{n} \left(\sum_{i=1}^n x_{ji} \right) \left(\sum_{i=1}^n y_i \right) \quad (j=1,2)$$

3.51 式推导如下:

由 3.48 式知

$$b_0 \sum_{i=1}^n x_{1i} + b_1 \sum_{i=1}^n x_{1i}^2 + b_2 \sum_{i=1}^n x_{1i}x_{2i} = \sum_{i=1}^n x_{1i}y_i$$

将 3.50 式代入上式得

$$(\bar{y} - b_1 \bar{x}_1 - b_2 \bar{x}_2) \sum_{i=1}^n x_{1i} + b_1 \sum_{i=1}^n x_{1i}^2 + b_2 \sum_{i=1}^n x_{1i}x_{2i} = \sum_{i=1}^n x_{1i}y_i$$

$$b_1 \sum_{i=1}^n (x_{1i} - \bar{x}_1)^2 + b_2 \sum_{i=1}^n (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)$$

$$= \sum_{i=1}^n (x_{1i} - \bar{x}_1)(y_i - \bar{y})$$

$$\text{令 } l_{11} = \sum_{i=1}^n (x_{1i} - \bar{x}_1)^2$$

$$l_{12} = \sum_{i=1}^n (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)$$

$$l_{1y} = \sum_{i=1}^n (x_{1i} - \bar{x}_1)(y_i - \bar{y})$$

$$\text{则有 } l_{11}b_1 + l_{12}b_2 = l_{1y}$$

同理可证 3.52 式

现将 3.51 式两边同乘以 $\frac{1}{\sqrt{l_{11}} \cdot \sqrt{l_{yy}}}$ 得

$$\begin{aligned} & \frac{l_{11}}{\sqrt{l_{11}} \cdot \sqrt{l_{yy}}} b_1 + \frac{l_{12}}{\sqrt{l_{11}} \cdot \sqrt{l_{yy}}} b_2 = \frac{l_{1y}}{\sqrt{l_{11}} \cdot \sqrt{l_{yy}}} \\ & \frac{l_{11}}{\sqrt{l_{11}} \cdot \sqrt{l_{yy}}} \cdot \frac{\sqrt{l_{11}}}{\sqrt{l_{11}}} b_1 + \frac{l_{12}}{\sqrt{l_{11}} \cdot \sqrt{l_{yy}}} \cdot \frac{\sqrt{l_{22}}}{\sqrt{l_{22}}} b_2 \\ & = \frac{l_{1y}}{\sqrt{l_{11}} \cdot \sqrt{l_{yy}}} \\ & \frac{l_{11}}{\sqrt{l_{11}} \cdot \sqrt{l_{11}}} \left[\frac{\sqrt{l_{11}}}{\sqrt{l_{yy}}} b_1 \right] + \frac{l_{12}}{\sqrt{l_{11}} \cdot \sqrt{l_{22}}} \left[\frac{\sqrt{l_{22}}}{\sqrt{l_{yy}}} b_2 \right] \\ & = \frac{l_{1y}}{\sqrt{l_{11}} \cdot \sqrt{l_{yy}}} \end{aligned}$$

$$\text{令 } b'_1 = \frac{\sqrt{l_{11}}}{\sqrt{l_{yy}}} b_1, \quad b'_2 = \frac{\sqrt{l_{22}}}{\sqrt{l_{yy}}} b_2$$

$$\text{则得 } r_{11}b'_1 + r_{12}b'_2 = r_{1y}$$

b'_i 称为标准偏回归系数

同理可把 3.52 式写成

$$r_{21}b'_1 + r_{22}b'_2 = r_{2y}$$

于是得

$$\begin{cases} r_{11}b'_{11} + r_{12}b'_{12} = r_{1y} \\ r_{21}b'_{11} + r_{22}b'_{12} = r_{2y} \end{cases} \quad (3.53)$$

$$\begin{cases} r_{11}b'_{21} + r_{12}b'_{22} = r_{1x} \\ r_{21}b'_{21} + r_{22}b'_{22} = r_{2x} \end{cases} \quad (3.54)$$

这是写成相关系数的正规方程,若以系数组成的矩阵,称为相关系数矩阵。

$$\begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix}$$

2. 复相关系数

用以表示复相关关系紧密程度的特征数称为复相关系数记为

$r_{y,1,2,\dots,m}$ 其定义仿照总体相关系数的定义

$$\begin{aligned} r_{y,1,2,\dots,m} &= \sqrt{1 - \frac{Q}{L_{yy}}} = \sqrt{1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}} = \sqrt{\frac{L_{yy}}{L_{yy}}} \\ &= \sqrt{\frac{\sum_{i=1}^m b_i l_{iy}}{L_{xy}}} \end{aligned} \quad (3.55)$$

式中 $Q = \sum (y_i - \hat{y}_i)^2$ 称为样本多元线性回归剩余平方和,而剩余标准差为

$$S_{\hat{e}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} = \sqrt{\frac{Q}{n}}$$

但作为对总体多元线性回归剩余标准差的无偏估计值则为

$$S'_{\hat{e}} = \sqrt{\frac{1}{n-m-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2} = \sqrt{\frac{Q}{n-m-1}} \quad (3.56)$$

复相关系数恒取正值。对复相关系数作 t 检验

$$t = \frac{r_{y,1,2,\dots,m} \cdot \sqrt{n-m-1}}{\sqrt{1-(r_{y,1,2,\dots,m})^2}} \quad (3.57)$$

统计量 t 是遵循自由度 $df = n - m - 1$ 的 t 分布。式中 m 为自变量个数。

当 $t > t_{(\alpha, n-m-1)}$ 时, 则多元线性回归差异显著。

当 $t < t_{(\alpha, n-m-1)}$ 时, 则多元线性回归差异不显著, 所配回归方程无意义。

3. 偏相关系数

在研究因变量对于两个或两个以上自变量的相关关系时, 我们有时需要分析因变量对于这些自变量的某一个自变量的相关关系。例如我们要分析优势木平均高对于林分平均高的相关关系时, 把年龄、密度等自变量当作固定不变的, 这种在 x_1, x_2 固定不变条件下研究 y 对 x_3 的相关关系叫做 y 对 x_3 的偏相关关系。

描写偏相关关系紧密程度的特征数称为偏相关系数。

其计算公式为

$$r_{y,1,2,\dots,j-1,j+1,\dots,m} = \frac{R_{yj}}{\sqrt{R_{yy}R_{jj}}} = \frac{C_{yy}}{\sqrt{C_{yy}C_{jj}}} \quad (3.58)$$

$$r_{j,1,2,\dots,i-1,i+1,\dots,j-1,j+1,\dots,m} = \frac{R_{ij}}{\sqrt{R_{ii}R_{jj}}} = \frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}} \quad (3.59)$$

式中 C_{ij} 为与 R (相关矩阵) 的元素 r_{ij} 相应的逆阵 R^{-1} 的元素。这样我们便可以计算偏相关系数了。

对于 $\hat{y} = b_0 + b_1x_1 + b_2x_2$ 的情形

$$R = \begin{bmatrix} r_{yy} & r_{y1} & r_{y2} \\ r_{y1} & r_{11} & r_{12} \\ r_{y2} & r_{21} & r_{22} \end{bmatrix} = \begin{bmatrix} 1 & r_{y1} & r_{y2} \\ r_{y1} & 1 & r_{12} \\ r_{y2} & r_{21} & 1 \end{bmatrix} \quad r_{12} = r_{21}$$

$$R_{yy} = \begin{vmatrix} 1 & r_{12} \\ r_{12} & 1 \end{vmatrix} = 1 - r_{12}^2; \quad R_{.1} = \begin{vmatrix} 1 & r_{y2} \\ r_{y2} & 1 \end{vmatrix} = 1 - r_{y2}^2$$

$$R_{y1} = - \begin{vmatrix} r_{y1} & r_{12} \\ r_{y2} & 1 \end{vmatrix} = -(r_{y1} - r_{12} \cdot r_{y2})$$

$$r_{y1 \cdot 2} = \frac{r_{y1}}{\sqrt{R_{yy} R_{11}}} = \frac{r_{y1} - r_{12} r_{y2}}{\sqrt{1 - r_{12}^2} \sqrt{1 - r_{y2}^2}} \quad (3.60)$$

$$\text{同理 } r_{y2 \cdot 1} = \frac{-R_{y2}}{\sqrt{R_{yy} R_{22}}} = \frac{r_{y2} - r_{12} \cdot r_{y1}}{\sqrt{1 - r_{12}^2} \sqrt{1 - r_{y1}^2}} \quad (3.61)$$

偏相关系数的 t 检验, 对偏相关系数作 t 检验, 以考察 x_i 对 y 贡献的显著性。

$$t_i = \frac{r_{yi} \cdot \sqrt{n-m-1}}{\sqrt{1-r_{yi}^2}} \quad (3.62)$$

统计量 t 是遵循自由度 $df = n - m - 1$ 的 t 分布, t_i 的绝对值越大, 则该因素(或自变量)越重要。

4. 例题

例 3.8 实测调查了闽北地区 71 块有固定边界的杉木标准地, 资料如表 3.8。

试建立前三项指标与优势木平均高的回归方程, 并计算复相关系数, 偏相关系数。

表 3.8

序号	年龄(取对数) x_1	密度 x_2	林分平均高 x_3	优势木平均高 $y(x_1)$
1	0.6990	292	3.2	4.2
2	0.6990	367	4.3	6.0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
71	1.3802	143	120	13.5
共计	87.0580	11564	883.0	1045.2

解 (1)回归方程的建立与计算。

这里自变量是 x_1, x_2, x_3 、现选配三元线性回归方程

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3$$

式中参数 $b_i (i=0, 1, 2, 3)$ 用最小二乘法确定, 亦即必须使参数 $b_i (i=0, 1, 2, 3)$ 满足下列一组正规方程:

$$\begin{cases} nb_0 + b_1 \sum x_1 + b_2 \sum x_2 + b_3 \sum x_3 = \sum y & (3.63) \end{cases}$$

$$\begin{cases} b_0 \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_1 x_2 + b_3 \sum x_1 x_3 = \sum x_1 y & (3.64) \end{cases}$$

$$\begin{cases} b_0 \sum x_2 + b_1 \sum x_1 x_2 + b_2 \sum x_2^2 + b_3 \sum x_2 x_3 = \sum x_2 y & (3.65) \end{cases}$$

$$\begin{cases} b_0 \sum x_3 + b_1 \sum x_1 x_3 + b_2 \sum x_2 x_3 + b_3 \sum x_3^2 = \sum x_3 y & (3.66) \end{cases}$$

由 3.63 式得

$$b_0 = \bar{y} - b_1 \bar{x}_1 - b_2 \bar{x}_2 - b_3 \bar{x}_3 = \bar{y} - \sum_{i=1}^3 b_i \bar{x}_i$$

而 3.64、3.65、3.66 式的系数及常数项可用离差来表示

$$\begin{cases} l_{11}b_1 + l_{12}b_2 + l_{13}b_3 = l_{1y} \\ l_{21}b_1 + l_{22}b_2 + l_{23}b_3 = l_{2y} \\ l_{31}b_1 + l_{32}b_2 + l_{33}b_3 = l_{3y} \end{cases}$$

为了计算各 l_{ij} 的值, 列离差平方和及离差积和计算表, 为

表达方便,把 y 记为 x_1 ,见表 3.9。

表 3.9

i	j	$\Sigma(x_i x_j)$	$(\Sigma x_i)(\Sigma x_j)$	$l_{ij} = \Sigma(x_i x_j) - \frac{\Sigma(x_i)\Sigma(x_j)}{n}$
1	1	108.76339	7577.8418	2.03
1	2	14134.5169	1006655.451	-43.73
1	3	1115.30895	76865.8564	32.69
1	4	1319.17873	90985.4962	37.69
2	2	2129.756	133726096.0	246290
2	3	140994.9	10211012.0	2822.2
2	4	168103.4	12086692.8	-2131.7
3	3	11933.76	779689.0	952.22
3	4	14042.12	922911.6	1043.36
4	4	16561.64	1092443.04	1175.12

* y 记为 x_1

i, j 为 x 变量的下标,例如表 3.9 中对于 $i=1, j=1$ 则

$$\begin{aligned}
 l_{11} &= \sum (x_1 x_1) - \frac{\sum (x_1) \sum (x_1)}{n} \\
 &= \sum x_1^2 - \frac{1}{n} (\sum x_1)^2 \\
 &= 108.76339 - \frac{1}{71} (87.0508)^2 \\
 &= 2.03
 \end{aligned}$$

对 $i=2, j=4$ 则

$$\begin{aligned}
 l_{24} &= \sum (x_2 x_4) - \frac{\sum (x_2) \sum (x_4)}{n} \\
 &= 168103.4 - \frac{12086692.8}{71} \\
 &= -2131.7
 \end{aligned}$$

因为 $l_{24}=l_{42}$,故 l_{42} 不予列出,其余类推

$$\text{而 } \bar{x}_1 = \frac{1}{n} \sum x_1 = \frac{87.0508}{71} = 1.22607$$

$$\bar{x}_2 = \frac{1}{n} \sum x_2 = \frac{11564}{71} = 163$$

$$\bar{x}_3 = \frac{1}{n} \sum x_3 = \frac{883.0}{71} = 12.43662$$

$$\bar{y} = \frac{1}{n} \sum y = \frac{1045.2}{71} = 14.721127$$

于是列出正规方程为

$$2.03b_1 - 43.73b_2 + 32.69b_3 = 37.69$$

$$-43.73b_1 + 246290b_2 - 2822.2b_3 = -2131.7$$

$$32.69b_1 - 2822.2b_2 + 952.22b_3 = 1043.36$$

用求解求逆并行方案解参数 b_1, b_2, b_3 ,

列出增广矩阵

$$A^{(0)} = \begin{bmatrix} 2.03 & -43.73 & 32.69 & 37.69 & 1 & 0 & 0 \\ -43.73 & 246290 & -2822.2 & -2131.7 & 0 & 1 & 0 \\ 32.69 & -2822.2 & 952.22 & 1043.36 & 0 & 0 & 1 \end{bmatrix}$$

对增广矩阵 $A^{(0)}$ 施行 $l=1, 2, 3$ 次消去变换得

$$A^{(1)} = \begin{bmatrix} 1 & -21.5419 & 16.1034 & 18.5665 & 0.4926 & 0 & 0 \\ 0 & -5610.5185 & 48.4335 & 30.1804 & -0.4926 & -0.0229 & 0 \\ 0 & -64.7903 & 13.0524 & 13.3503 & -0.4926 & 0 & 0.0306 \end{bmatrix}$$

$$A^{(2)} = \begin{bmatrix} 1 & 0 & 15.91743 & 18.45063 & 0.490704 & 0.000088 & 0 \\ 0 & 1 & -0.008633 & -0.005379 & 0.000088 & 0.000004 & 0 \\ 0 & 0 & -0.192407 & -0.200671 & 0.007515 & -0.000004 & -0.000472 \end{bmatrix}$$

$$A^{(3)} = \begin{bmatrix} 1 & 0 & 0 & 1.84875 & 1.112407 & -0.00024 & -0.03905 \\ 0 & 1 & 0 & 0.003625 & -0.00024 & 0.000004 & 0.000021 \\ 0 & 0 & 1 & 1.0430 & -0.03905 & 0.000021 & 0.002453 \end{bmatrix}$$

求得

$$b_1 = 1.84875, \quad b_2 = 0.003625, \quad b_3 = 1.0430$$

$$b_0 = \bar{y} - b_1\bar{x}_1 - b_2\bar{x}_2 - b_3\bar{x}_3 = 14.721127 - 1.84875$$

$$\times 1.22607 - 0.003625 \times 163 - 1.0430 \times 12.43662$$

$$= -1.1078$$

故求得优势木平均高关于林分年龄、密度、林分平均高的三元线性回归方程为

$$\hat{y} = 1.84875x_1 + 0.003625x_2 + 1.0430x_3 - 1.1078$$

而 5、6、7 列给出系数矩阵的逆矩阵。

$$C^{-1} = \begin{bmatrix} 1.112407 & -0.00024 & -0.03905 \\ -0.00024 & 0.000004 & 0.000021 \\ -0.03905 & 0.000021 & 0.002453 \end{bmatrix}$$

其对角线元素记为

$$C_{11} = 1.112407, \quad C_{22} = 0.000004, \quad C_{33} = 0.002453$$

(2) 线性回归显著性的 F 检验——方差分析

计算平方和

$$\begin{aligned} L_{\text{回}} = u &= \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = b_1 l_{1y} + b_2 l_{2y} + b_3 l_{3y} \\ &= 1.84875 \times 37.6928 + 0.003625 \times (-2131.7) \\ &\quad + 1.0430 \times 1043.36 \\ &= 1150.1817 \end{aligned}$$

$$L_Y = l_{yy} = 1175.12$$

$$L_{\text{余}} = Q = L_Y - L_{\text{回}} = 1175.12 - 1150.1817 = 24.9383$$

自由度

$$df_{\text{总}} = n - 1 = 71 - 1 = 70, \quad df_{\text{回}} = 3,$$

$$df_{\text{余}} = df_{\text{总}} - df_{\text{回}} = 70 - 3 = 67$$

列方差分析表

表 3.10

变差来源	平方和	自由度	均方	F 值	F _{0.05} 值
回归	1150.1817	3	$S_r^2 = \frac{1150.1871}{3}$ = 383.3939	$F = \frac{383.3939}{0.3722}$ = 1030.11	$F_{3,67}^{0.05}$ = 4.13
剩余	24.9383	67	$S_d^2 = \frac{24.9383}{67}$ = 0.3722		
共计	1175.12	70	-	-	-

经方差分析结果,线性回归关系极端显著,说明所配的回归方程效果很好,有实用的意义。

(3) 计算复相关系数

$$r_{y, x_1, x_2, x_3} = \sqrt{\frac{L_{yy}}{L_{yy}}} = \sqrt{\frac{U}{L_{yy}}} = \sqrt{\frac{1150.1871}{1175.12}} = 0.9893$$

对复相关系数作 t 检验

$$t = \frac{r_{y, x_1, x_2, x_3} \sqrt{n-m-1}}{\sqrt{1-(r_{y, x_1, x_2, x_3})^2}} = \frac{0.9893 \sqrt{71-3-1}}{\sqrt{1-(0.9893)^2}} = 55.504$$

根据 $\alpha=0.01, df=n-m-1=71-3-1=67$, 查 t 分布表得 $t_{0.01, 67}=2.655$, 而 $t=55.504 > 2.655$, 表明林分年龄、密度、平均高对于优势高之间存在着显著的线性相关且关系极紧密。

(4) 计算偏相关系数

① 分别求出各简相关系数

$$r_{y1} = \frac{l_{12}}{\sqrt{l_{11}l_{22}}} = \frac{37.69}{\sqrt{2.03 \times 1175.12}} = \frac{37.69}{48.84} = 0.771$$

$$r_{y2} = \frac{l_{2y}}{\sqrt{l_{22}l_{yy}}} = \frac{-2131.7}{\sqrt{246290 \times 1175.12}} = \frac{-2131.7}{17012.36} \\ = -0.125$$

$$r_{y3} = \frac{l_{3y}}{\sqrt{l_{33}l_{yy}}} = \frac{1043.36}{\sqrt{952.22 \times 1175.12}} = \frac{1043.36}{1057.82} \\ = 0.986$$

$$r_{12} = r_{21} = \frac{l_{12}}{\sqrt{l_{11}l_{22}}} = \frac{-43.73}{\sqrt{2.03 \times 246290}} = \frac{-43.73}{707.08} \\ = -0.062$$

$$r_{31} = r_{13} = \frac{l_{13}}{\sqrt{l_{11}l_{33}}} = \frac{32.69}{\sqrt{2.03 \times 952.22}} = \frac{32.69}{43.97} \\ = 0.743$$

$$r_{23} = r_{32} = \frac{l_{23}}{\sqrt{l_{22}l_{33}}} = \frac{-2822.2}{\sqrt{246290 \times 952.22}} = \frac{-2822.2}{15314.02} \\ = -0.188$$

②令

$$R = \begin{vmatrix} r_{yy} & r_{y1} & r_{y2} & r_{y3} \\ r_{1y} & r_{11} & r_{12} & r_{13} \\ r_{2y} & r_{21} & r_{22} & r_{23} \\ r_{3y} & r_{31} & r_{32} & r_{33} \end{vmatrix} = \begin{vmatrix} 1 & 0.771 & -0.125 & 0.986 \\ 0.771 & 1 & -0.062 & 0.743 \\ -0.125 & -0.062 & 1 & -0.188 \\ 0.986 & 0.743 & -0.188 & 1 \end{vmatrix}$$

分别求出 R_{y1} 、 R_{y2} 、 R_{11} 、 R_{22} 、 R_{33} ，其分别表示 $R = |r_{ij}|$ 中各元素 r_{yx} 、 r_{xy} 、 r_{ij} 的代数余子式。

$$R_{y1} = - \begin{vmatrix} 0.771 & -0.062 & 0.743 \\ -0.125 & 1 & -0.188 \\ 0.986 & -0.188 & 1 \end{vmatrix} = -0.031$$

$$R_{y2} = \begin{vmatrix} 0.771 & 1 & 0.743 \\ -0.125 & -0.062 & -0.188 \\ 0.986 & 0.743 & 1 \end{vmatrix} = 0.024$$

$$R_{v3} = - \begin{vmatrix} 0.771 & 1 & -0.062 \\ -0.125 & -0.062 & 1 \\ 0.986 & 0.743 & -0.188 \end{vmatrix} = -0.400$$

$$R_{11} = \begin{vmatrix} 1 & -0.125 & 0.986 \\ -0.125 & 1 & -0.188 \\ 0.986 & -0.188 & 1 \end{vmatrix} = 0.023$$

$$R_{22} = \begin{vmatrix} 1 & 0.771 & 0.986 \\ 0.771 & 1 & 0.743 \\ 0.986 & 0.743 & 1 \end{vmatrix} = 0.012$$

$$R_{33} = \begin{vmatrix} 1 & 0.771 & -0.125 \\ 0.771 & 1 & -0.062 \\ -0.125 & -0.062 & 1 \end{vmatrix} = 0.398$$

$$R_{yy} = \begin{vmatrix} 1 & -0.062 & 0.743 \\ -0.062 & 1 & -0.188 \\ 0.743 & -0.188 & 1 \end{vmatrix} = 0.427$$

③计算偏相关系数

$$r_{(y, x_1), x_2 x_3} = \frac{-R_{y1}}{\sqrt{R_{11}R_{yy}}} = \frac{-(-0.031)}{\sqrt{0.023 \times 0.427}} = 0.313$$

$$r_{(y, x_2), x_1 x_3} = \frac{-R_{y2}}{\sqrt{R_{22}R_{yy}}} = \frac{-(-0.024)}{\sqrt{0.012 \times 0.427}} = 0.335$$

$$r_{(y, x_3), x_1 x_2} = \frac{-R_{y3}}{\sqrt{R_{33}R_{yy}}} = \frac{-(-0.400)}{\sqrt{0.398 \times 0.427}} = 0.970$$

从偏相关系数来看自变量 x_3 (林分平均高) 对于因变量 (优势木平均高) 的偏相关关系都比 x_1 、 x_2 来得紧密, 应指出: 有了 R_{yy} 、 R_{y1} 、 R_{y2} 、 R_{y3} , 又已知 S_y 、 S_{x1} 、 S_{x2} 、 S_{x3} 的值可求方程中参数 b_1 、 b_2 、 b_3

$$b_1 = - \frac{S_y}{S_{x1}} \cdot \frac{R_{y1}}{R_{yy}} \quad (3.67)$$

$$b_2 = -\frac{S_y}{S_{x2}} \cdot \frac{R_{y2}}{R_{yy}} \quad (3.68)$$

$$b_3 = -\frac{S_y}{S_{x3}} \cdot \frac{R_{y3}}{r_{yy}} \quad (3.69)$$

$$\text{而 } b_0 = \bar{y} - \sum_{i=1}^3 b_i l_{iy}$$

$$\text{而 } b'_1 = b_1 \sqrt{\frac{l_{11}}{l_{yy}}} = b_1 \sqrt{\frac{\frac{l_{11}}{n}}{\frac{l_{yy}}{n}}} = b_1 \frac{S_{x1}}{S_y}$$

$$b'_2 = b_2 \frac{S_{x2}}{S_y} \quad b'_3 = b_3 \frac{S_{x3}}{S_y}$$

5. 多元回归中确定主导因素的方法

在多元回归分析中,实际上各个自变量对因变量的作用是不一致的。如何差别所观察的多因子中,哪些是重要因素,哪些是次要因素,一般所用方法有:

- (1) 计算分析标准偏回归系数
- (2) 计算分析偏回归平方和
- (3) 检验偏相关系数

下面分别加以介绍

(1) 标准偏回归系数(b')

在多元线性回归方程中的偏回归系数 $b_i (i=1, 2, \dots, m)$, 是指在其他因素都保持不变的情况下, 某一自变量(x_i)变动一个单位所引起的因变量(y)将变动多少个单位的变动系数。因此, b_i 的绝对值越大, 则该因素就越显著。但是, 因为它们与各自变量所取的单位有关, 因此不能用来相互比较, 为了消除这个干扰, 可以用标准偏回归系数(b'_i)来比较分析; b'_i

$= b_i \sqrt{\frac{l_{ii}}{l_{yy}}}$, 标准偏回归系数 b'_i 与 x_i 及 y 所取的单位无关, 因

此可以用来互相比较。它的绝对值越大,则其相应的因素 x_i 对 y 的影响就越大,其中最大的一个,其相应的自变量 x_i 也常常就是影响因变量的主导因素。

仍以上例作计算说明

$$b'_1 = b_1 \sqrt{\frac{l_{11}}{l_{yy}}} = 1.84875 \sqrt{\frac{2.03}{1175.12}} = 0.0768$$

$$b'_2 = b_2 \sqrt{\frac{l_{22}}{l_{yy}}} = 0.003625 \sqrt{\frac{246290}{1175.12}} = 0.0525$$

$$b'_3 = b_3 \sqrt{\frac{l_{33}}{l_{yy}}} = 1.0430 \sqrt{\frac{952.22}{1175.12}} = 0.9389$$

可见 y 对 x_3 的标准偏回归系数比 y 对 x_1 及 y 对 x_2 的标准偏回归系数都大,这说明林分平均高这个因素是重要的。

比较标准偏回归系数来判别多元回归中各因素的主次关系的方法比较简便直观,但这种方法只有当各自变量之间的相关性都很小时,即 r_{ij} (不包括 r_{iy}) 的绝对值都不大时才正确,如果在各自变量彼此之间相关密切,应用这个方法就可能得出不正确的结论。例如,在上例中因 x_1 和 x_3 之间相关关系密切 ($r_{13} = 0.743$), 因此应当用偏回归平方和来作比较判别。

(2) 偏回归平方和 (P_i)

偏回归平方和的计算公式为

$$P_i = \frac{b_i^2}{C_{ii}} \quad (3.70)$$

式中 b_i 为 y 对第 i 个自变量的偏回归系数;

C_{ii} 为多元回归标准方程的系数矩阵 C 的逆矩阵 C^{-1} 中的一个元素。

P_i 为 y 对第 i 个自变量的偏回归平方和。

凡是对 y 作用显著的因素,一般具有较大的 P_i 值, P_i 值越大,则该因素 x_i 对 y 的作用也就越大,这样通过比较各个因素的 P_i 值,就可以大致看出各个因素对因变量 y 作用的重要性。偏回归平方和大的,一定是对 y 有重要影响的因素。但究竟要大到什么程度才算是显著,这就要作 F 检验:

$$F_i = \frac{P_i}{S_e^2} = \frac{b_i^2}{C_{ii}S_e^2} \quad (3.71)$$

式中 S_e^2 系方差分析计算中的剩余方差,统计量 F_i 是遵循第一自由度 $f_1=1$,第二自由度 $f_2=n-m-1$ 的 F 分布。

根据规定 α 值及 $f_1=1, f_2=n-m-1$ 查 F 分布表得 $F_{\alpha}(1, n-m-1)$ 临界值。

当 $F_i > F_{\alpha}(1, n-m-1)$ 时,则表示此 x_i 是显著的重要影响因素,一般 P_i 最大的一个因素也常常是影响 y 的主导因素。

现仍以上例作具体计算。

$$F_1 = \frac{b_1^2}{C_{11}S_e^2} = \frac{1.84875^2}{1.112407 \times 0.3740} = \frac{3.41788}{0.4604} = 8.22$$

$$F_2 = \frac{b_2^2}{C_{22}S_e^2} = \frac{0.003625^2}{0.000004 \times 0.3740} = \frac{0.000013}{0.0000015} = 8.67$$

$$F_3 = \frac{b_3^2}{C_{33}S_e^2} = \frac{1.0430^2}{0.002453 \times 0.3740} = \frac{1.08785}{0.000917} = 1186.31$$

根据 $\alpha=0.01, f_1=1, f_2=n-m-1=71-3-1=67$ 查 F 分布表得 $F_{0.01}(1, 67)=6.96$

可见 x_3 对 y 的影响要比 x_1 与 x_2 大得多,是个主导因素。

(3) 偏相关系数的检验

对偏相关系数作 t 检验,以考察 x_i 对 y 贡献的显著性,现

仍以上例依 3.62 式作具体计算。

$$t_1 = \frac{r_{(y, x_1)x_2x_3} \sqrt{n-m-1}}{\sqrt{1-r^2}} = \frac{0.313 \sqrt{67}}{\sqrt{1-0.313^2}} = 2.697$$

$$t_2 = \frac{r_{(y, x_2)x_1x_3} \sqrt{n-m-1}}{\sqrt{1-r^2}} = \frac{0.335 \sqrt{67}}{\sqrt{1-0.335^2}} = 2.910$$

$$t_3 = \frac{r_{(y, x_3)x_1x_2} \sqrt{n-m-1}}{\sqrt{1-r^2}} = \frac{0.970 \sqrt{67}}{\sqrt{1-0.970^2}} = 32.661$$

根据 $\alpha=0.01$ 及自由度 $df=n-m-1=71-3-1=67$, 查 t 分布表得 $t_{0.01, 67}=2.655$, 可见 x_3 对 y 的影响都比 x_1 与 x_2 来得显著。

二、多元非线性回归

对于多元非线性回归问题, 这里不作仔细的讨论, 仅仅指出对于一些比较简单的情形。例如对 $\hat{y}=ax_1^{b_1}x_2^{b_2}\cdots x_m^{b_m}$ 这一类回归方程, 为确定 $a, b_i (i=1, 2, \cdots, m)$ 只需对于等式两边取对数, 即得

$$\lg y = \lg a + b_1 \lg x_1 + b_2 \lg x_2 + \cdots + b_m \lg x_m \quad \text{令 } \lg y = \hat{Y}$$

$$\lg a = a' \quad \lg x_1 = x'_1 \quad \lg x_2 = \lg x'_2 \cdots, \lg x_m = x'_m$$

则上式可化为多元线性回归的形式

$$\hat{Y} = a' + b_1 x'_1 + b_2 x'_2 + \cdots + b_m x'_m$$

这一类方程在编制二元、三元材积表中常被用到, 因为材积 V 与胸径 D , 树高 H 的关系常取 $V=D^b H^c$ 的形式。

再如, $\hat{y} = \frac{1}{a+bx_1+cx_2}$ 这样一类方程, 只要令 $1/y=y'$, 即可化为二元线性回归式, 这是比较明显的。

对于很广泛的一类一般的多元非线性回归问题, 即多元多项式回归问题, 如

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_1^2 + b_4 x_1 x_2 + b_5 x_2^2 + \cdots$$

则可以令 $x_1 = X_1, x_2 = X_2, x_1^2 = X_3, x_1 x_2 = X_4, x_2^2 = X_5, \dots$ 化为多元线性回归进行处理, 由于任意函数可以用多项式逐段逼近, 因此一般性问题在原则上就可以用上述方法处理, 但由于“元”数迅速增加, 因子间的相关性即使利用电子计算机也会感到计算量过大, 对结果的分析也很复杂, 为此常用正交多项式回归克服这些缺点, 这里就不再讨论了。

第四章 完全随机化试验和 随机区组试验

第一节 完全随机化试验

完全随机化试验设计是最基本的一种试验设计。这种试验设计一般只考虑单一因素,假定因素 A ,选择它的 a 个水平 $A_1, A_2, \dots, A_a$,其试验安排完全随机化试验就称为完全随机化单因子试验设计。这 a 个水平可以看成 a 个方法或条件。在各水平上试验重复 n 次,总共进行 an 次试验。当然各水平的重复数也可以不相同,例如为了比较不同温度对某种真菌生长的影响,只考虑一个因素温度。选定温度为 $17.5^\circ(A_1)$ 、 $21^\circ(A_2)$ 、 $24.5^\circ(A_3)$ 、 $27.5^\circ(A_4)$ 、 $30^\circ(A_5)$ (单位 $^\circ\text{C}$) 5 个水平,各水平重复 4 次,共做 20 次试验,这 20 次试验的安排完全随机化,按试验条件配制 20 组培养基,并同时接种,可以将 20 组培养基,利用随机数表分为 5 小组,分别按 5 种选定温度加以控制,10 天后测其菌丝长度(单位:毫米)得表 4.1。如大家已经熟悉的,该问题可用单因素方差分析法进行统计分析。

完全随机化试验的数据结构模型为

$$x_{ij} = \mu + \alpha_i + \epsilon_{ij} \quad (i=1, 2, \dots, a, j=1, 2, \dots, n) \quad (4.1)$$

其中 x_{ij} 是第 i 水平(处理)的第 j 个观察值。 μ 称为总平均数, α_i 称为第 i 个处理的效应, ϵ_{ij} 称为误差,是一个随机变量。我们还假设 $\epsilon_{ij} \sim N(0, \sigma)$,按随机化条件,它们相互独立。 σ^2 即为试验误差的方差。

表 4.1

重复号	水 平				
	A_1	A_2	A_3	A_4	A_5
1	3.9	3.1	4.0	3.8	3.5
2	3.5	3.2	3.9	3.7	4.0
3	4.0	4.0	3.5	3.4	3.9
4	3.6	3.8	3.8	3.5	3.3

现在对表 4.1 的试验数据作方差分析。对这一实际问题，要检验 4 个处理效应是否存在

$$H_0: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$$

$$H_1: \alpha_i (i=1, 2, 3, 4,) \text{ 中至少有一个不为 } 0$$

若接受 H_0 ，则不存在处理效应，每一个观察值都是由总平均数(μ)加上随机误差(ϵ_{ij})构成，若拒绝 H_0 ，则存在处理效应，每一观测值由总平均数，处理效应以及随机误差三部分构成。

先计算校正项 C

$$C = \frac{x_{..}^2}{na} = \frac{73.4^2}{4 \times 5} = 269.378$$

再计算

$$L_T = \sum_i \sum_j x_{ij}^2 - C = 270.9 - 269.378 = 1.522$$

$$L_A = \frac{1}{n} \sum_{i=1}^a x_{i.}^2 - C$$

$$= \frac{1}{4} (15.0^2 + 14.1^2 + 15.2^2 + 14.4^2 + 14.7^2) - C$$

$$= 269.575 - 269.378 = 0.197$$

$$L_e = L_T - L_A = 1.522 - 0.197 = 1.325$$

将上面的计算结果列出方差分析表，表 4.2 中，从分析结

果有如下结论：不同温度对该真菌生长无显著影响。

表 4.2

变差来源	离差平方和	自由度	均方	均方比	F 值
处理间	0.197	4	0.04925	0.5588	$F_{0.95}(4, 15) = 2.99$
误差	1.325	15	0.08813		
总 和	1.522				

第二节 随机区组设计

随机区组设计，其全名应为完全随机化区组设计。它是一种比较充分运用林间试验设计基本原则的试验设计方案。实践中经常需要人们设计一种能够系统地控制外来因素引起的变差的试验方案。例如进行松木的抚育间伐强度试验，选择林分测树因子一致的试验地并把它分成 9 个试验小区，分别按强度、中度、不间伐三个水平加以处理，测定按不同强度间伐三年后的平均胸径。试验只有一个因素，即间伐强度，本来完全可以按单因素完全随机化试验设计方案进行，但由于试验地大，会存在土壤肥力的差异，如果真是这样，土壤肥力的差异与真正的试验误差混杂在一起，会增大试验误差。而试验设计的一个基本原则应总是将试验误差压缩到尽可能小。这时，最好在环境条件大致相同或时间空间上相距较近的范围的试验地进行试验，这种范围，今后简称为区组。按此思路，我们可以将间伐的 3 个水平对应的 3 个处理配置在一个区组里视为一个重复，配置在另一个区组里的 3 个水平为另一重复等等，由于每一区组内条件相近，会减少或避免 3 个处理处在更大

范围内的不同小区内而造成土壤差异与试验误差的混杂,至于区组内不同处理在不同小区上的配置应完全随机。注意,这时区组与区组之间的差异客观存在,我们只能在方差分析中将这种差异造成的平均胸径的差异从总变动中排除。

一般地,随机区组设计是指这样的试验设计:试验对象由于多种原因无法置于同等条件下做试验,为此,先将它们所处范围划分若干条件比较一致的区组,将每区组再分成若干小区与处理相应,从而容纳全部需比较的处理,各处理在每小区内的配置按随机化原则进行。

随机区组试验结果的统计分析采用两向分组方差分析法,其中一个因素即所研究因素(如 A)另一因素为区组因素。两向分组方差分析可以判断这两个因素各水平间差异显著程度。如果区组之间无显著差异,说明区组条件相似,否则只有用统计分析方法才能真正对各处理间的差异显著性作出判断。

随机区组设计的统计模型为

$$x_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij} \quad \begin{cases} i=1, 2, \dots, a \\ j=1, 2, \dots, b \end{cases} \quad (4.2)$$

在随机区组设计的试验中,处理一般都是固定型:

$$\sum_{i=1}^a \alpha_i = 0$$

区组可能为固定型,也可能为随机型,当区组为固定型时

$$\sum_{j=1}^b \beta_j = 0$$

当区组为随机型时

$$\beta_j \sim N(0, \sigma_\beta)$$

随机区组实验的目的是检验处理效应,因此统计假设为

$$H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_a = 0$$

$$H_1: \text{至少有一个 } \alpha \neq 0 \quad (i=1, 2, \cdots, a)$$

随机区组设计试验也可以检验 $H_0: \beta_j = 0$, 检验的目的是确定区组之间是否存在显著差异, 以便提供在以后的类似实验中是否还有必要设置区组的依据。如果区组间没有显著差异, 区组可以安排配置另一因素, 提高试验的效率。

例 4.1 福建省“五一”林场于 1977 年进行杉木抚育间伐强度试验, 采用随机区组设计。设置 3 个区组, 3 个处理(对照、中度、强度)。3 年后测定每一小区的平均胸径, 结果见表 4.3

表 4.3

处理	区 组			总 和 平 均	
	I	II	III	$X_{i..}$	$\bar{X}_i$
中度	11.7	11.7	11.0	34.4	11.5
强度	11.2	12.0	11.9	35.1	11.7
对照	10.7	10.8	10.1	31.6	10.5
合计	33.6	34.5	33.0	101.1	11.2

试比较不同间伐强度是否造成了胸径的差异。

解: 在随机区组设计中, 区组数用 b 表示, 处理数用 t 表示, 小区数用 k 表示, 重复数用 r 表示, 现在 $b=3, t=3, k=3, r=3, t=k, b=r$ 。

本例中间伐强度是固定因素, 区组是随机因素, 零假设 $H_0: \alpha_i = 0 \quad i=1, 2, 3,$

先计算校正项 C

$$C = \frac{X_{..}^2}{a \cdot b} = \frac{101.1^2}{3 \times 3} = 1135.69$$

于是各离差平方和为

$$L_T = \sum_i \sum_j x_{ij}^2 - C = 1138.9 - 1135.69 = 3.28$$

$$L_A = \frac{1}{b} \sum_i x_{i.}^2 - C = \frac{1}{3} (34.4^2 + 35.1^2 + 31.6^2) - C \\ = 2.29$$

$$L_B = \frac{1}{a} \sum_j x_{.j}^2 - C = \frac{1}{3} (33.6^2 + 34.5^2 + 33.0^2) - C \\ = 0.38$$

$$L_e = L_T - L_A - L_B = 3.28 - 2.29 - 0.38 = 0.61$$

将以上结果列入方差分析表 4.4 中

表 4.4

变差来源	离差平方和	自由度	均方	均方比	F 值
处 理	2.29	2	1.145	7.508	$F_{0.10}(2,4)=4.32$
区 组	0.38	2	0.19	1.246	$F_{0.05}(2,4)=6.94$
误 差	0.61	4	0.1525		
总 和	3.28	8			

因 $F_{\text{处理}} > F_{0.05}(2,4)$ 拒绝 H_0 , 认为不同的间伐强度对三年后林分平均胸径有显著影响, 而区组间则没有显著差异。

为了进一步分析经哪一种处理后平均胸径显著高于其它处理结果, 用 Duncan 多范围检验的 (LSR 检验) 作多重比较。这里有

$$\bar{x}_{\text{强度}} > \bar{x}_{\text{中度}} > \bar{x}_{\text{对照}}$$

$$LSR = SSR \times S_e$$

$$S_e^2 = 0.1525 \quad f_e = 4 \quad \text{重复数 } r = 3$$

$$S_e = \sqrt{\frac{S_e^2}{r}} = \sqrt{\frac{0.1525}{3}} = 0.2255$$

为比较 $\bar{x}_1 = 11.7$ 与 $\bar{x}_2 = 11.5$ 与 $\bar{x}_3 = 11.5$ 及 $\bar{x}_4 = 10.5$ 可

采用 $LSR_2 = 6.51 \times S_x = 1.468 (\alpha = 0.01)$ 或

$$LSR_2 = 3.93 \times S_x = 0.886 \quad (\alpha = 0.05) \quad (a = 2)$$

现在 $\bar{x}_1 - \bar{x}_2 = 0.2$ $\bar{x}_2 - \bar{x}_3 = 1.0$, 可见在显著水平 $\alpha = 0.05$ 上, 中度、强度间伐后平均胸径无显著差异, 但中度间伐对照相比已达显著。

为比较 $\bar{x}_1 = 11.7$ 与 $\bar{x}_3 = 10.5$ 需采用 $(a = 3)$

$$LSR_3 = 6.80 \times \sqrt{\frac{0.1525}{3}} = 1.533 \quad (\alpha = 0.01)$$

$$\text{或 } LSR_3 = 4.01 \times \sqrt{\frac{0.1525}{3}} = 0.904 \quad (\alpha = 0.05)$$

而 $\bar{x}_1 - \bar{x}_3 = 11.7 - 10.5 = 1.2$, 故在 $\alpha = 0.05$ 水平上认为差异显著。

随机区组设计有许多优点。例如把试验地分成 b 个区组, 从总离差平方和 L_T 中分解出区组平方和 L_B , 从而比完全随机化设计和精度要高些。这种设计对区组数没有什么限制, 对结果的统计分析是简单的。如果在一个完整的试验中, 需要取消某些处理时并不影响对试验结果的分析。或者由于某些意外的原因丢失了一、二个数据, 也可以通过适当的方法补救。

但这种设计也有缺点。主要是必须保证区组内条件的高度一致, 这是比较困难的。方差分析只能鉴别区组之间的差异而分辨不出区组内的差异当实际上区组内存在较大变差时, 就会造成较大的试验误差。特别当处理增多, 小区数也随之增大, 从而区组面积较大, 这样就很难实现局部控制的原则。因此在试验设计中不能无限增大处理数。为了排除区组内差异对试验结果的影响, 当试验处理数不太多时可以采用拉丁方设计。

第五章 拉丁方设计和 希腊拉丁方设计

第一节 拉丁方设计

如果在试验地上进行土壤理化性质的测定,发现试验地东部和北部肥力高,而西部和南部肥力低,此时采用完全随机区组设计,不论区组内的小区是从东向西走,还是从南向北走,都不可能做到区组内各小区的肥力大致相同,为了消除两个方向的土壤肥力差异给实验带来的干扰,应当设计成两个方向上的区组,这就是拉丁方设计。

由 n 行 n 列 $n \times n$ 个小方块构成的方形,每一小方格用拉丁字母(或数字)来表示,共用 n 个字母(数字),若每一行,每一列各个字母均恰好出现一次,这个方就称为一个拉丁方。

如图 5.1 就是三个不同的拉丁方,用来排列拉丁方的字母个数 n 叫拉丁方的阶数,图 5.1 中的三个拉丁方分别为 3, 4, 5 阶。可称为 3×3 , 4×4 , 5×5 拉丁方。

1	2	3
2	3	1
3	1	2

(1)

A	B	C	D
B	C	D	A
C	D	A	B
D	A	B	C

(2)

A	B	C	D	E
B	C	D	E	A
C	D	E	A	B
D	E	A	B	C
E	A	B	C	D

(3)

图 5.1

拉丁方中第一行和第一列均为字母或数字的顺序排列的拉丁方称为标准方,图 5.1 中三个拉丁方均为标准方。

由于拉丁方行和列的数字或字母排列顺序不限,所以 k 阶拉丁方共有 $k!(k-1)!$ 个,如 2×2 拉丁方有 2 个; 3×3 拉丁方有 12 个, 4×4 拉丁方有 576 个, 5×5 拉丁方有 161280 个等等。

拉丁方试验设计是使试验因素的每一水平与拉丁字母对应,每一字母代表一个水平,全部处理均在每一行、每一列上出现一次且仅出现一次。因此不论在行方向还是列方向上出现环境差异时,拉丁方试验都可以象随机区组设计那样克服区组差异干扰,即可以克服二个方向差异带来的干扰。

因此拉丁方设计与随机完全区组设计相似,只是它实行了双向局部控制而已,在排列设计上拉丁方设计的重复数与处理数相等,不能任意更改,试验地必须划分成相等的列数与行数,且这个公共的行数与列数等于各行、列内的小区数,即处理数。

具体排列设计时可随机选用一个标准方,在各行、列配置前对该标准方的行和列分别施行随机调换,然后把各因子的各水平随机配置到各个小区上。

一个 P 阶拉丁方试验的统计模型为

$$x_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ijk} \quad (i, j, k = 1, 2, \dots, P) \quad (5.1)$$

其中 x_{ijk} 为第 i 行、第 j 列、第 k 处理的观察值。 μ 为总平均, α_i 为第 i 行效应, β_j 为第 j 列效应, γ_k 为第 k 处理效应, ε_{ijk} 是服从 $N(0, \sigma)$ 的相互独立随机变量。这一模型是可加的,即要求行、列与处理之间不存在交互作用。拉丁方设计主要检验处理效应,但实际上共有三个因素,除处理因素外,行和列是从两个方向上控制外来因素对试验结果的影响。

作统计分析时,象两向方差分析那样,把 P^2 个观察值的

总离差平方和分解为行、列、处理的离差平方和。

$$L_T = L_{\text{行}} + L_{\text{列}} + L_{\text{处理}} + L_e$$

自由度的分解如下

$$f_T = P^2 - 1 \quad f_{\text{行}} = f_{\text{列}} = f_{\text{处理}} = P - 1$$

$$f_{\text{误差}} = f_e = (p-1)(p-2)$$

$$f_T = f_{\text{行}} + f_{\text{列}} + f_{\text{处理}} + f_e$$

欲检验假设 $H_0: \gamma_k = 0$ 时,

检验统计量为 F

$$F_{\text{处}} = \frac{S_{\text{处理}}^2}{S_e^2}$$

但也可以检验假设 $\alpha_i = 0$ 或 $\beta_j = 0$, 检验统计量分别为

$$F_{\text{行}} = \frac{S_{\text{行}}^2}{S_e^2} \quad F_{\text{列}} = \frac{S_{\text{列}}^2}{S_e^2}$$

例 5.1 采用 5×5 拉丁方设计进行五种油茶品种的比较试验, 小区面积为一亩, 6 年生时测定其亩产油量, 试验结果见表 5.1, 试判断各品种油产量是否有显著差异。

解 由表 5.1 计算出 $x_{...} = 199$

$$x_{..A} = 35, x_{..B} = 32, x_{..C} = 67, x_{..D} = 39, x_{..E} = 26$$

表 5.1

行 \ 列	I	II	III	IV	V	$x_{.j.}$
I	B_5	D_6	E_3	A_{10}	C_{12}	36
II	C_{10}	A_4	B_6	E_5	D_8	30
III	D_8	C_{15}	A_7	B_6	E_5	41
IV	E_7	B_7	C_{13}	D_9	A_5	41
V	A_9	E_6	D_{11}	C_{17}	B_8	51
$x_{.j.}$	39	38	40	47	35	199

现假设各品种产油量无显著差异,

即 $H_0: \gamma_k = 0 (k=1, 2, \dots, 5)$

H_1 : 至少一个 $\gamma_k \neq 0$

$$\text{计算校正项 } C = \frac{x_{...}^2}{p^2} = \frac{199^2}{25} = 1584.04$$

计算各项离差平方和

$$L_T = \sum_i \sum_j x_{ij}^2 - C = 1879 - 1584.04 = 294.96$$

$$L_{\text{行}} = \frac{1}{p} \sum_i x_{i..}^2 - C = \frac{1}{5} (36^2 + 30^2 + \dots + 51^2) - C \\ = 47.76$$

$$L_{\text{列}} = \frac{1}{p} \sum_j x_{.j.}^2 - C = \frac{1}{5} (39^2 + 38^2 + \dots + 35^2) - C \\ = 15.76$$

$$L_{\text{处理}} = \frac{1}{p} \sum_i x_{..k}^2 - C = \frac{1}{5} (x_{..A}^2 + x_{..B}^2 + \dots + x_{..E}^2) - C \\ = \frac{1}{5} (35^2 + 32^2 + \dots + 26^2) - C \\ = 202.96$$

$$L_e = L_T - L_{\text{行}} - L_{\text{列}} - L_{\text{处理}} = 28.48$$

根据以上计算列出方差分析表 5.2

表 5.2

变差来源	离差平方和	自由度	均方	均方比	F 值
行	47.76	4	11.94	5.04*	$F_{0.05}(4, 12) = 3.26$
列	15.76	4	3.94	1.66	
处理(品种)	202.96	4	50.74	21.41**	$F_{0.01}(4, 12) = 5.41$
误差	28.48	12	2.37		
总和	294.96	24			

检验结果认为油茶各品种对产油量有显著影响,行向区组也有显著差异,至于 5 个品种哪一个品种最好,可以进一步

作多重比较后确定,不在此一一列出。

拉丁方设计比随机区组设计精密,因为这种设计实行双向局部控制,减少了误差。但当处理少于4个时,重复次数太少;若处理数多于8个时,其试验小区达64个,施工、调查、计算都会感到太烦,所以这种方法一般应用于处理数为5~8个为宜。如果处理数小于4时,为了克服重复次数少的缺点,可以采用交叉式设计,但不在这里介绍了。

第二节 希腊拉丁方设计

如果把一个用拉丁字母表示的 $P \times P$ 阶拉丁方上再重合上一个由希腊字母表示的 $P \times P$ 阶拉丁方。相重合的两个拉丁方中,每一个希腊字母与每一个拉丁字母共同出现一次且仅出现一次,我们就称这两个拉丁方是正交的。与拉丁方设计相似理解,现在将希腊字母视为另一因素的 P 个水平,这等于完成了一次试验设计,这种设计称为希腊——拉丁方设计见图 5.2

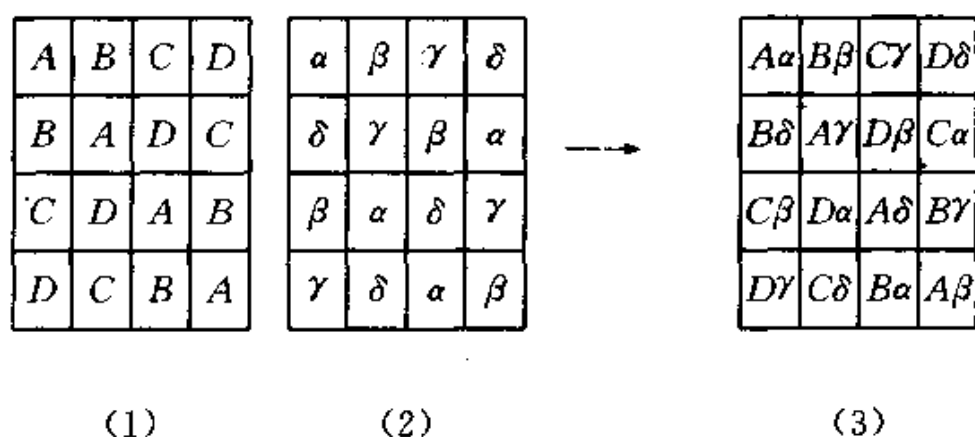


图 5.2

图 5.2 中(1)(2)二个拉丁方互相正交,(3)为(1)(2)重叠的结果。令希腊字母、拉丁字母各代表一种主要试验因素,不同字母代表这二种因素的不同水平,则(3)立即表达了一个试验设计方案,按此试验,试验区的行和列尽管可以有条件上的差异,采用四向分组方差分析即可克服这种差异的干扰使我们对二个主要因子取得明确的比较结果。而如果行、列二个方向上无显著差异的话,一次试验可安排四个因素。

对拉丁方来说 $P=6$ 除外, $P>3$ 的各个拉丁方都存在相应正交拉丁方。附表中列出若干拉丁方和的正交拉丁方可供参考。

希腊拉丁方设计的统计模型为:

$$x_{ijkl} = \mu + \theta_i + \tau_j + \omega_k + \psi_l + \varepsilon_{ijkl} \quad (5.2)$$

其解释基本上与拉丁方相同,不再细述。

例 5.2 为了研究品种(C)及肥料(D)的变化对某种作物收获量的影响,并考虑土壤肥力的差异(行变化与列变化作为土壤肥力 A 与 B),按希腊拉丁方设计试验,并获得产量数据见表 5.3.,试分析品种、肥料的变化对产量是否有显著影响(产量单位公斤)。

解 根据表 5.3,先计算出不同 C 字母(不同品种)的产量之和,不同 D 字母(不同肥料水平)的产量之和。

字 母	C ₁	C ₂	C ₃	C ₄	D ₁	D ₂	D ₃	D ₄
产量和	202	203	223	189	207	287	144	179

表 5.3

		土 壤 肥 力 B				$x_{i...}$
		B_1	B_2	B_3	B_4	
土 壤 肥 力 A	A_1	$C_1D_1=94$	$C_2D_2=62$	$C_3D_3=27$	$C_4D_4=34$	217
	A_2	$C_2D_3=59$	$C_1D_1=44$	$C_1D_1=26$	$C_3D_2=90$	219
	A_3	$C_3D_1=52$	$C_1D_3=33$	$C_1D_2=39$	$C_2D_1=33$	157
	A_4	$C_1D_2=96$	$C_3D_1=54$	$C_2D_1=49$	$C_1D_3=25$	224
$x_{...}$		301	193	141	182	$x_{...}=817$

$$\text{校正数 } C = \frac{x_{...}^2}{p^2} = \frac{817^2}{16} = 41718.06$$

各项离差平方和为

$$L_T = \sum_i \sum_j x_{ij}^2 - C = 50319 - 41718.06 = 8600.94$$

$$L_A = \frac{1}{p} \sum_i x_{i...}^2 - C = \frac{1}{4} (217^2 + \dots + 224^2) - C$$

$$= 42468.75 - 41718.06 = 750.69$$

$$L_B = \frac{1}{p} \sum_i x_{...i}^2 - C = \frac{1}{4} (301^2 + \dots + 182^2) - C$$

$$= 45213.75 - 41718.06 = 3495.69$$

$$L_C = \frac{1}{p} \sum_i x_{...i}^2 - C = \frac{1}{4} (202^2 + \dots + 189^2) - C$$

$$= 41865.75 - 41718.06 = 147.69$$

$$L_D = \frac{1}{p} \sum_i x_{...i}^2 - C = \frac{1}{4} (207^2 + \dots + 179^2) - C$$

$$= 44498.75 - C = 2780.69$$

$$L_e = L_T - L_A - L_B - L_C - L_D = 4572.18$$

列出方差分析表 5.4

表 5.4

变异来源	离差平方和	自由度	均方	均方比	F 值
A	750.69	3	250.23	<1	$F_{0.01}(3,3)=9.3$
B	3495.69	3	1165.23	<1	
C	117.69	3	39.23	<1	
D	2780.69	3	926.90	<1	
E	1572.18	3	524.06		
总和	8600.91	15			

检验结果认为无论品种或肥料对收获均无显著影响,而误差平方和甚大,这说明可能尚有重要因素影响收获量,而未被我们揭示,此时最好根据本试验条件再作进一步分析,找出原因,以便进行试验。

希腊拉丁方的概念还可以象从拉丁方导出希腊拉丁方那样进一步得到扩展,形成中国希腊拉丁方,但我们不准备继续对此作具体的讨论了。

一个 $p \times p$ 拉丁方最多可以有 $p-1$ 个正交拉丁方。如果存在 $p-1$ 个互为正交的 $p \times p$ 拉丁方称为正交拉丁方完全系,将 $p-1$ 个正交拉丁方重叠到一起可以容纳 $p-1$ 个因素。在试验设计时可视具体情况灵活应用。

不难发现,上述试验设计系列,可以容纳的因素愈多而保持试验次数不增加的条件下,则误差项的自由度就愈少,从而降低了试验的灵敏度,误差项自由度是试验设计中应予特别注意的。

第六章 正交试验设计

上面讨论的试验设计多数为单因素或双因素情形,然而在生物科学中,作为试验对象的事物极为复杂,试验结果常常受多种因素影响。例如挖掘松根对水土保持的影响涉及林地植被的破坏、土壤状况、降雨强度、降雨量等等,为了提高嫁接成活率,考虑影响其成活率的因素有药物处理、无性系、操作者、接穗长度等多个因素。在树木栽培方面要找到优质、高产、低耗,生产周期短的生产方案也必须考虑作多因素试验。

正交试验设计是多因素分析的有力工具。通过正交试验不仅可以搞清每个因素对试验结果的影响情况,还可以分清因素的主次以及它们之间交互作用的关系,选出最优水平组合。正交试验形成方案十分简便,可以用较少的实验次数得到较多的信息。

正交试验设计是利用正交表来安排试验,进行数据分析的。

第一节 正交试验设计法

为了对正交试验设计有深入的了解,先通过一个实例说明其基本思想。

为探讨影响湿地松种子催芽的因素,选择了可能影响试验结果的浸种水温 *A*、消毒方式 *B*、催芽时数 *C* 三个因素。每因素的试验水平规定如下:

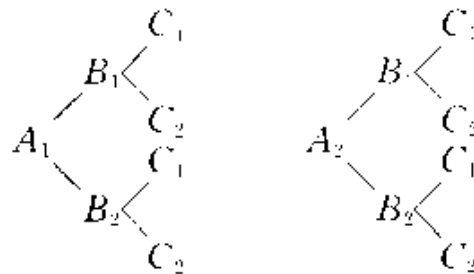
$A_1: A_1 = 25^\circ\text{C}, \quad A_2 = 30^\circ\text{C}$
 $B: B_1 = 20\% \text{石灰水}, \quad B_2 = \text{不消毒}$
 $C: C_1 = 24 \text{小时}, \quad C_2 = 48 \text{小时}$

在正交试验设计中,因素可以是定量的,也可以是定性的,水平间距离可以相等,也可以不相等。

对这个三因素二水平的试验,一般可以有三种方法进行试验

一、全面比较法

即让全部可能的三因素各水平组合都作试验,以进行直接比较,水平组合情形:

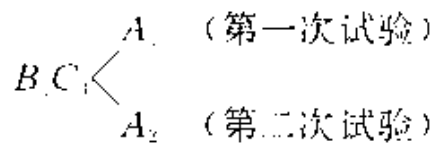


共需作 $2^3 = 8$ 次试验。

但全面比较试验,试验次数随因素数及每因素水平数的增加而急剧增加。如 6 因素,每因素 5 水平的试验需作 $5^6 = 15625$ 次试验,这在实践中几乎是无法完成的。

二、轮换比较法

变化一个因素的水平而固定其它因素的水平。如先固定 B, C 于 B_1, C_1 , 让 A 变化, 而完成二次试验。



如果得出结论以 A_1 为好, 则固定 A 于 A_1 , C 仍固定在 C_1 , 使 B 变化, 作如下二次试验:

$$A_1C_1 \begin{cases} B_1 \text{ (已作过)} \\ B_2 \text{ (第三次试验) (好的结果)} \end{cases}$$

如果得到 B_2 最好, 则固定 A 于 A_1 , B 于 B_2 , 使 C 变化, 做试验。

$$A_2B_2 \begin{cases} C_1 \text{ (已作过试验) (好的结果)} \\ C_2 \text{ (第四次试验)} \end{cases}$$

如果得 C_1 最好, 则最优水平组合为 $A_1B_2C_1$ 。

这种方法, 一般说有一定效果, 但缺点是试验点的代表性差, 分布不均匀, 除 $A_2B_2C_1$ 外 A_2 参与的试验次数太少, 换言之, 可能出现这种情形, 即在很大范围内没有试验点。因此试验所获得的最优水平组合可能并不是所有可能的水平组合中的最佳者。加上试验缺少重复, 所得结论稳定性差。

三、正交试验设计

利用正交表安排试验, 上例中我们选用正交表 $L_4(2^3)$ 来安排试验, 见表 6.1。用 4 次试验得出结论, 表中 1、2 分别代表水平状态 4 次试验的水平组合分别为 $(A_1B_1C_1)$ $(A_1B_2C_2)$ $(A_2B_1C_2)$ 与 $(A_2B_2C_1)$ 。

现在可以将这三种方法所作试验情况, 用图 6.1 来说明。

表 6.1

试验号	因 素		
	A	B	C
1	1	1	1
2	1	2	2
3	2	1	2
4	2	2	1

全面比较法所作 8 次试验, 相当于图中立方体的 8 个顶点。其中 $A_1B_1C_1$ 点放在左下前位置上, 左 A_1 、右 A_2 ; 下 B_1 ; 上

B_2 ; 前 C_1 后 C_2 。轮换法作 4 次试验, 试验点为用 $\triangle$ 标示的顶点。而正交试验法所作 4 次试验用 $\bigcirc$ 标示。这 4 个试验点的特点是: 分散在正方体的不同部位, 各部位选取点的比例相同, 代表性强, 正方体其 6 个面, 每个面有而且都只有二个试验点, 每条棱上有而且只有一个试验点。不仅如此, 这 4 个试验点是全面试验点数 8 的 $1/2$ 。

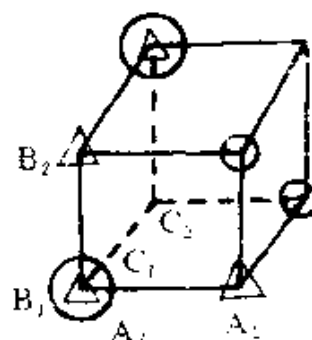
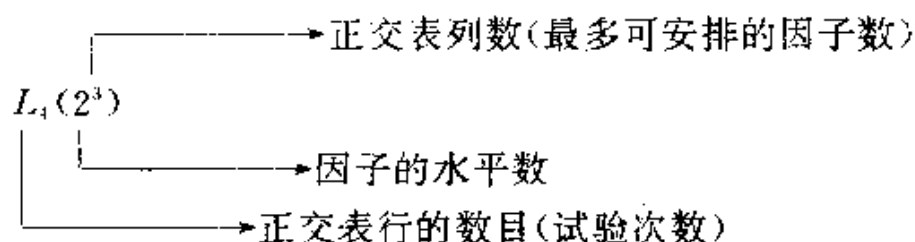


图 6.1

正交试验设计按正交表安排试验, 既能使试验点分布均匀、分散, 又能减少试验次数, 下面还能看到正交试验计算分析简单, 能够清楚地阐明试验条件与试验指标间的关系。

1. 正交表

正交试验设计是利用正交表来安排试验和分析试验结果的。表 6.1 就是一张正交表, 记为 $L_4(2^3)$ 。L 代表正交表, 各数字意义如下:



正交表中各列都由从 1 开始的若干数码 1, 2, 3, ... 组成, 且各列两两正交, 构成矩形表。

一般正交表记为 $L_n(S_1 \times S_2 \times S_3 \times \cdots \times S_m)$ 。L 表示正交

表, n 表示试验次数, m 表示列数, 即最多可能安排的因子数, $S_1, S_2, \dots, S_m$ 分别表示第 1 列, 第 2 列, $\dots$, 第 m 列的水平数。如果有 $S_1 = S_2 = \dots = S_m = S$, 则正交可改记为 $L_n(S^m)$ 。例如 $L_9(3^4)$ 表示用这张正交表安排试验, 需做 9 次试验, 最多可安排 4 个 3 水平的因素。 $L_{18}(2 \times 3^7)$ 需作 18 次试验, 最多可以安排 1 个 2 水平因子和 7 个 3 水平因子。

在行数为 S^m 型的正交表中, 试验次数为 n

$$n = \sum_{i=1}^m (S_i - 1) + 1 = m(S - 1) + 1$$

如 $L_8(2^7)$ 表, $S = 2$ $m = 7$

$$\therefore n = m(S - 1) + 1 = 7 \times 1 + 1 = 8$$

$$L_{27}(3^{13}) S = 3 \quad m = 13$$

$$\therefore n = 13 \times (3 - 1) + 1 = 27$$

在行数为非 S^m 型正交表中,

$$n \geq \sum_{i=1}^m (S_i - 1) + 1$$

如 $L_{18}(2 \times 3^7)$

$$n > (2 - 1) + 7(3 - 1) + 1 = 16$$

利用上述关系式可以从所需考察的因子水平数来决定最低的试验次数, 进而合理选择适当的正交表。

正交表具有如下两条性质:

(1) 每一列中各数字出现次数一样多

(2) 任何两列所构成的有序数对出现的次数都一样多。

以表 6.2 中给出的 $L_9(3^4)$ 为例, 每一列中 1、2、3 三个数字均各出现 3 次, 又 1、2、3 三个数字每二个形成的有序数对可能是 (1,1)(1,2)(1,3)(2,1)(2,2)(2,3)(3,1)(3,2)(3,3) 它们对于 $L_9(3^4)$ 表的任意二列而言, 正好是各出现 1 次。

这两个特性反映了试验点分布的均匀分散性。

表 6.2

1	2	3	4	
1	1	1	1	1
2	1	2	2	2
3	1	3	3	3
4	2	1	2	3
5	2	2	3	1
6	2	3	1	2
7	3	1	3	2
8	3	2	1	3
9	3	3	1	1

2. 试验方案的设计

试验方案的设计首先要根据科学研究或生产实际的需要挑选好试验因素。影响试验结果的每种因素,在正交试验中不可能也没有必要对它们一一加以考察。例如某林场要从各地杉木种源中选取一种种源,并研究适合该种源的栽培方法。这个试验选取了种源、整地挖穴方法、施肥等因素,挑选好因素是试验设计的关键。对选出因素,可凭生产经验及专业知识来确定它们的变化范围,在此范围内选出每个因素的水平。水平间隔要适当,这也是安排试验方案的重要环节,因为它直接关系到试验的效果。

挑选好因素,确定了水平之后,依此选用正交表。

选用正交表的原则是:列数够用,主要因子(包括因素及因素间交互作用)不混杂。

列够用是指选用正交表的列数要大于或等于主要因子的个数,当然也不要大得太多,否则将会使试验次数增加很多

的。在排完所有主要因子后,要留有空列,但空列尽量少些,这样的正交表最适宜。

不混杂是指每一个主因子不能重复安置在相同的列的位置上,这里主要因子仍指因素及因素间的交互作用。

将挑选好的正交表的每列数字变成所对应的因子水平。这样每一行各水平组合就构成了一个试验条件。

例如上面讨论的影响湿地松种子催芽试验中,每因素 2 水平,试验次数要不少于

$$3 \times (2-1) + 1 = 4 (\text{次})$$

可以考虑选用正交表 $L_4(2^3)$ 或 $L_8(2^7)$ 。

又例如为了探讨影响某种苗木嫁接成活率的因素,据分析品种、嫁接者与药物处理可能对成活率影响较大。为此用 3 种品种、选 3 名嫁接者,用 3 种药物处理,而形成三因素三水平试验,它的试验次数要不少于

$$3 \times (3-1) + 1 = 7 (\text{次})$$

因此选用 $L_9(3^4)$ 表(见表 6.2)

为了使各号试验条件(水平组合)一目了然,可将各水平具体内容都注在表中。

3. 试验数据的计算分析

例 6.1 我国东南沿海许多地方大面积发生木麻黄木毒蛾虫害,为探索用核多角体病毒、白僵菌、80%敌敌畏混合使用防治木毒蛾的合理配比,在林间采用正交试验设计进行试验,取 4~6 龄幼虫 100 头套笼观察,喷药后每隔 48 小时检查一次死虫数,最后以 288 小时校正死亡率为试验指标。试验因素及水平排列为 A (核多角体病毒亿/ml), B (白僵菌亿/克), C (80%敌敌畏稀释倍数); $A_1:1, A_2:0.1, A_3:0.01, B_1:2, B_2:1, B_3:0.5, C_1:1 \cdot 1 \text{ 万}, C_2:1 \cdot 1.5 \text{ 万}, C_3:1 \cdot 2 \text{ 万}$ 。

选择正交表 $L_9(3^4)$ 安排试验, 试验结果见表 6.3。

表 6.3

	A	B	C	$P=288$ 小时		
	1	2	3	4	校正死亡率%	$x = \sin^{-1} \sqrt{p}$
1	1	1	1	1	87.8	69.56
2	1	2	2	2	89.3	70.91
3	1	3	3	3	79.6	63.15
4	2	1	2	3	37.5	37.76
5	2	2	3	1	35.2	36.29
6	2	3	1	2	45.7	42.53
7	3	1	3	2	15.0	22.79
8	3	2	1	3	28.8	32.46
9	3	3	2	1	26.3	30.85
K_1	203.62	130.11	144.55	136.70	$T=406.3$	
K_2	116.58	139.66	139.52	136.23		
K_3	86.10	136.53	122.23	133.37		
k_1	67.87	43.37	48.18	45.57		
k_2	38.86	46.55	46.15	45.41		
k_3	28.70	40.74	45.51	44.41		
R	39.17	3.18	7.44	1.14		

我们将 A, B, C 三因素分别按排在正交表的第一、二、三列上。从试验结果看出第二号试验指标值最好, 那末是否水平组合 $A_1B_2C_2$ 就是各因素的最佳水平组合呢? 这尚需依赖于进一步的统计分析。

计算步骤为

(1) 求 K 值。表中第一列的 K_1 表示 A_1 参加三次试验的指标值之和, 亦记为 K_{11} , 其它符号类似。

$$K_{A1} = 69.56 + 70.91 + 63.15 = 203.62$$

$K_{A2} = 116.58, \dots, K_{A3} = 122.23$ 均列于表中

对于空白列各水平的 K 值也一样计算。每一列的三个 K 值之和都应等于指标值总和 T , 这可以用来检验 K 值计算是否正确。

(2) 求 k 值

$$k_{A1} = \frac{K_{A1}}{A_1 \text{ 参加试验次数}} = \frac{203.62}{3} = 67.87$$

其余的 k 值按同样方法求得, 如果各水平参加试验的次数相等, 则此项计算可以略去。 k 值是和 K 值相对应的。

根据 k 值大小可以分别通过 A, B, C 三因素各水平的比较求得最优水平组合 $A_1B_2C_1$, 因为 A 因素以 A_1 为佳, B 因素以 B_2 为最佳, C 因素以 C_1 为佳。而 $A_1B_2C_1$ 这一组合并没有在 9 次试验中出现, 这正体现了正交试验的优点。

(3) 求极差 R

表 6.3 的 R 值在 A 列上的数值表示 A 因素三个水平的平均指标 k_{A1}, k_{A2}, k_{A3} 的极差, 记为 R_A

$$R_A = \max k_{Ai} - \min k_{Ai} = 67.87 - 28.7 = 39.17 (i = 1, 2, 3)$$

$$R_B = 3.18 \quad R_C = 7.44$$

极差 R 愈大的因素对指标的影响愈显著, 本例中 A 为主导因子。从极差值可以看出因素的主次关系为

$$\begin{array}{c} A \longrightarrow C \longrightarrow B \\ \text{主} \longrightarrow \text{次} \end{array}$$

按以上步骤对问题作出分析的过程被称为正交试验的直观分析法。应该说, 这样得到的结论不够严谨、精确, 但由于其简便、直观、实用, 实际工作中还是常常被采用的。

更进一步的,精确的分析应采用方差分析,对上面这个例子的方差分析统计模型为

$$x_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ijk} \quad (i, j, k = 1, 2, 3) \quad (6.1)$$

式中 α, β, γ 分别为因素 A, B, C 的主效应, ε_{ijk} 为随机误差成份, μ 总平均, 其中

$$\sum_{i=1}^3 \alpha_i = \sum_{j=1}^3 \beta_j = \sum_{k=1}^3 \gamma_k = 0$$

$$\varepsilon_{ijk} \sim N(0, \sigma)$$

方差分析的基本做法仍然是

(1) 作统计假设

$$H_{OA}: \alpha_1 = \alpha_2 = \alpha_3 = 0; \quad H_{OB}: \beta_1 = \beta_2 = \beta_3 = 0;$$

$$H_{OC}: \gamma_1 = \gamma_2 = \gamma_3 = 0$$

(2) 计算各项离差平方和

$$\text{校正数 } C = \frac{T^2}{n} = \frac{406.3^2}{9} = 18342.19$$

$$L_T = \sum_{i=1}^9 x_i^2 - C = 20931.09 - 18342.19 = 2588.88$$

$$L_A = \frac{1}{3} \sum_i K_{Ai}^2 - C = 2479.54$$

$$L_B = \frac{1}{3} \sum_j K_{Bj}^2 - C = 15.80$$

$$L_C = \frac{1}{3} \sum_k K_{Ck}^2 - C = 91.38$$

$$L_e = L_T - L_A - L_B - L_C = 2.16$$

但我们指出,实际上可以形式地象 A, B, C 各列一样地计算空白列平方和 $L_{\varnothing}$

$$\begin{aligned}
L_{\text{空}} &= \frac{1}{3} \sum_i K_{\text{空}_i}^2 - C \\
&= \frac{1}{3} (136.70^2 + 136.23^2 + 133.37^2) - C \\
&= 1835.47 - 18342.19 = 2.16
\end{aligned}$$

并有 $L_{\text{空}} = L_e$

换言之,在方案设计时最好留下空白列,并将它视为与误差项对应的因素,这对方差分析是必要的,否则只有通过试验重复才能得到误差项的估计量。

(3) 确定各项自由度

总自由度 = 总试验次数 - 1 = 9 - 1 = 8

各处理自由度 = 水平数 - 1 = 3 - 1 = 2

(4) 列出方差分析表,下结论。

现列出方差分析表 6.4。

表 6.4

变差来源	平方和	自由度	均方	均方比	$F_{0.05}$	$F_{0.01}$
A	2479.54	2	1239.77	1147.94**	19.0	99.0
B	15.80	2	7.9	7.31		
C	91.38	2	45.69	42.31*		
e	2.16	2	1.08			
Σ	2588.88	8	—	—	—	—

分析结果认为核多角体病毒对试验结果的校正死亡率有极为显著的影响,80%的敌敌畏对试验结果有显著的影响,而白僵菌对试验结果没有显著影响。水平的最佳组合同上述的结果。

第二节 交互作用的试验设计与统计分析

在多因素对比试验中,某些因子对试验指标的影响往往有互相制约、互相联系的现象。这样,在处理多因素对比试验时,就不仅需分别研究各因子水平的改变对试验指标的影响,而且需要考察这些因子各水平间如何搭配的问题,这正是因子间的交互作用效应。

在多因素试验中,因素间的交互作用可以很多,全面考虑一切可能的交互作用会使问题大大复杂化,实践中某些因子间的作用根据经验和专业知识可以判断它们是很小的,这时设计试验中如把它们忽略不计会使问题简化,并有利于对主要因素及因素间交互作用的考察,减少试验次数。今后我们将接受这种处理问题的原则。

在正交试验中由于正交表的正交性,考察因素间交互作用可以象考察主因素一样地进行,若因素 A 与因素 B 之间的交互作用视为 $A \times B$,在 A, B 用正交表中确定的列代表的条件下, $A \times B$ 可以象 A, B 一样用正交表的其它的列代表,但列的位置不能任意选择,而用交互作用列表来确定,当交互作用所在列确定之后,分析可以将交互作用与主因素同等对待,并作出方差分析,下面举例说明。

例 6.2 “九二〇”是一种植物生长调节剂,在经济作物生长期间,喷撒一定浓度的“九二〇”可获增产效果。某地用土法生产“九二〇”存在产品效价低(一般在 10000~15000 单位),成本高等问题,(效价是用来衡量产品中某微量物质含量大小的,单位为 10^{-3} 克)为此采用正交试验法进行试验,并采

用以下因素和水平

表 6.5

水平 \ 因子	微量元素的重量 A	玉米粉 B	白糖 C	时间 D
1	0.6%	13%	3%	20 天
2	0.35%	17%	4%	25 天

根据以往经验,微量元素 A、玉米粉 B、白糖 C 之间,可能存在着交互作用(即 $A \times B, B \times C, A \times C$),而时间与它们的交互作用可以忽略。希望用较少的试验,摸清这四个因子和三个交互作用。

一般说来,水平数相同的两个因素,其交互作用所占的列数为水平数减 1,其自由度为:

$$f_{AB} = f_A \times f_B$$

这样该试验一共需要 7 列。

$$\text{试验次数} = \sum_{j=1}^m (\text{每列水平数} - 1) + 1 = 7 (\text{次})$$

查二水平正交表可知,至少要选用 $L_8(2^7)$ 的正交表。因为需要考察 $A \times B$,所以可先把 A 放在第 1 列, B 放在第 2 列。查表 6.6 知,第一列与第 2 列的交互作用列在第 3 列,将 C 放在第 4 列,而第 1 列与第 4 列的交互作用列查表 6.6 后知在第 5 列,而第 2 列与第 4 列的交互作用列在第 6 列,那么将 D 放在第 7 列。具体设计见表 6.7

表 6.6

列 号	1	2	3	4	5	6	7
(1)		3	2	5	4	7	6
(2)			1	6	7	4	5
(3)				7	6	6	4
(4)					1	2	3
(5)						3	2
(6)							1

表 6.7

列 号	1	2	3	4	5	6	7
因 子	A	B	$A \times B$	C	$A \times C$	$B \times C$	D

它常称为表头设计。

从表 6.6 中可以看出上述设计中 $A \times D$ 在第 6 列, $B \times D$ 在第 5 列, $C \times D$ 在第 3 列, 因此实际上有以下表头设计

因 素	A	B	$A \times B$ $C \times D$	C	$A \times C$ $B \times D$	$A \times D$ $B \times C$	D
列 号	1	2	3	4	5	6	7

这个设计的第 3 列出现了 $A \times B$ 和 $C \times D$ 重叠, 第 5 列出现 $A \times C$ 和 $B \times D$ 重叠, 第 6 列出现 $A \times D$ 和 $B \times C$ 重叠的现象, 这种同一列表示二个或二个以上因素的现象称为混杂现象, 但如象上例那样, 如果 $A \times D$, $B \times D$ 和 $C \times D$ 很小而可以忽略, 那么接受表 6.7 那样的表头设计方案, 把它当作没有混杂现象, 以使用较少次数的试验完成试验主要目标。这是正交试验处理问题的一个很重要的思想, 值得吸取。现将上例用表 6.7 作为表头设计后的试验结果列出于表 6.8 中。

考察表中 R 行, 比较 R 值的大小, 得到各因子和各交互作用的重要顺序如下:

$D, A \times C, A, A \times B, B \times C, C, B,$

主—————→次

在选取最优组合时, 由于 $D, A \times C, A$ 比较重要, 它们是主要依据, 其次 $A \times B$ 也有影响, 可作参考。

由于效价愈高愈好, 先考虑最重要因素 D , 因为 $k_{12} > k_{01}$, 故因素 D 应取 D_2 水平。

然后由 $A \times C$ 确定 A, C 的水平, 由于 A 与 C 共有四种搭

配: $A_1C_1, A_1C_2, A_2C_1, A_2C_2$, 每一搭配出现在二次试验中, 取这二次试验的平均指标值来代表之, 它们分别为 2.245, 1.67, 1.38, 1.675。因最高值与 2.245, 即与搭配 A_1C_1 相对应, 故应选择 A 取 A_1 水平, C 取 C_1 水平。

表 6.8

因子 列号 试验号	A	B	A×B	C	A×C	B×C	D	指标效价
1	1	1	1	1	1	1	1	2.05
2	1	1	2	1	2	2	2	2.24
3	1	2	1	2	1	2	2	2.44
4	1	2	2	2	2	1	1	1.10
5	2	1	2	1	2	1	2	1.50
6	2	1	2	2	1	2	1	1.35
7	2	2	1	1	2	2	1	1.26
8	2	2	1	2	1	1	2	2.00
K_1	7.83	7.14	7.55	7.25	7.84	6.65	5.76	$T = 13.94$
K_2	6.11	6.80	6.39	6.69	6.10	7.29	8.18	
k_1	1.9575	1.785	1.8875	1.8125	1.960	1.6625	1.440	
k_2	1.5275	1.700	1.5975	1.6725	1.525	1.8225	2.045	
R	0.43	0.085	0.29	0.14	0.435	0.16	0.605	

再看 A , 因为 $k_{A_1} > k_{A_2}$, 故应取 A_1 水平, 这与从 $A \times C$ 作出的对 A 的水平选取一致, A 的水平选择已确定, 但若关于 A 的单独选择与关于 $A \times C$ 的选择发生矛盾时应服从主要方面, 即以服从更重要的因素 $A \times C$ 作出的选择为准。这也是正交试验中选择最佳水平组合的一条基本原则。

下面选 B , 由于 $A \times B$ 较重要, 而 A 与 B 共有四种搭配, 算出平均效价 $A_1B_1, 2.145; A_1B_2, 1.77; A_2B_1, 1.425; A_2B_2, 1.63$; 以 A_1B_1 最高, 故 B 取 B_1 水平。

综上所述,我们选出了最佳的水平组合应该是 $A_1B_1C_1D_2$ 。这是直观分析的结果。

该问题的方差分析过程为

计算各项离差平方和

$$\text{校正项 } C = \frac{T^2}{n} = \frac{13.94^2}{8} = 24.2905$$

$$L_T = \sum_{i=1}^8 x_i^2 - C = 1.7533$$

$$L_A = \frac{1}{4} \sum_{i=1}^2 K_{A_i}^2 - C = \frac{1}{4} (7.83^2 + 6.11^2) - C = 0.3698$$

$$L_B = \frac{1}{4} \sum_{i=1}^2 K_{B_i}^2 - C = 0.0144$$

$$L_{A \times B} = \frac{1}{4} (7.55^2 + 6.39^2) - C = 0.1682$$

$$L_C = \frac{1}{4} (7.25^2 + 6.69^2) - C = 0.0392$$

$$L_{A \times C} = \frac{1}{4} (7.84^2 + 6.10^2) - C = 0.3784$$

$$L_{B \times C} = \frac{1}{4} (6.65^2 + 7.29^2) - C = 0.0512$$

$$L_D = \frac{1}{4} (5.76^2 + 8.18^2) - C = 0.7321$$

列出方差分析表,见表 6.9。我们看到仅仅在显著水平 $\alpha=0.10$ 上,因素 D 显著,其它因素,包括交互作用在内均未达到显著水平。这里我们将离差平方和最小的 B 因素所在列当成误差项 e 的列,而作出上述分析的。实践中,可以将所有经检验不显著的因素合并——离差平方和与自由度分别求和,并视为误差项后,进一步作方差分析。

表 6.9

变差来源	离差平方和	自由度	均方	均方比	F
A	0.3698	1	0.3698	25.6806	
B	0.0144	1	0.0144		$F_{0.05}(1,1)$
C	0.0392	1	0.0392	2.7222	=161
D	0.7321	1	0.7321	50.8403	$F_{0.01}(1,1)$
A×C	0.3784	1	0.3784	26.2778	=39.9
B×C	0.0512	1	0.0512	3.5556	
A×B	0.1682	1	0.1682	11.6806	
合计	1.7533	8	—	—	—

第三节 有重复试验的正交试验

在使用正交表安排试验时,会遇到这样的情形,正交表各列已被因子和交互作用占满,没有留下空白列,这时原则上方差分析将无法进行,因为没有办法对误差项作出估计,这时或者应选用更大的正交表来解决,或者可以作重复试验,从重复中找到误差项的估计。有时表头设计中虽留有空白列,但由于实验本身的需要等原因也有做重复试验的,即让同样的水平组合重复二次或二次以上试验,得到多个试验指标值。(请注意:重复试验不是重复测量或重复取样)。

做重复试验的正交设计,与无重复试验的正交设计,从试验的实施上并无重大差别,做统计分析时要用各试验号下重复试验的数据和 K 值进行计算,离差平方和的计算则需除以重复的次数。

在无重复的试验中,把空白列的离差平方和作为误差的

离差平方和,其中既包含有试验误差,也包含有其他各种干扰的误差,它是整体误差,我们称为第一类误差离差平方和,记为 L_{e_1} 。在有重复的试验中,还有第二类误差离差平方和 L_{e_2} ,它定义为:

$$L_{e_2} = \sum_{i=1}^n \sum_{k=1}^L (x_{ik} - \bar{x}_i)^2$$

上式中 n 为试验次数, L 为重复数。 L_{e_2} 为整个试验的组内离差平方和,它反映了试验误差大小,是局部误差。在计算时则作下式

$$L_{e_2} = \sum_{i=1}^n \sum_{k=1}^L x_{ik}^2 - \frac{1}{L} \sum_{i=1}^n T_i^2$$

重复试验中,第一类误差比第二类误差大。原则上第二类误差不能用来检验因子和交互作用各水平间显著差异。如果二类误差间没有显著差异,可以将它们合并,其检验方法:假设第一类误差与第二类误差没有显著差异,可以建立 F 统计量。

$$F = \frac{L_{e_1}/(f-1)}{L_{e_2}/[n(L-1)]} \sim F[(f-1), n(L-1)]$$

上式中 f 为空白列的水平数。如果 $F \geq F_\alpha$, 接受假设,从而不能将第二类误差并入第一类误差。如果 $F < F_\alpha$, 拒绝假设,认为二类误差没有显著差异,从而将它们合并后作为“试验误差”,这样有

$$L_e = L_{e_1} + L_{e_2}$$

其自由度 $dfe = dfe_1 + dfe_2$ 。请注意,当第一类误差与第二类误差有显著差异时,应当用第一类误差作为误差项,用它去检验各因子以及交互作用各水平间是否存在显著差异,否则得到的结论是错误的。

下面用实例说明其分析过程。

例 6.3 福建省毛竹介壳虫危害十分严重。介壳虫的腹面的口针刺入竹肉组织吸取汁液,被害竹株长势衰弱、出笋率下降,严重者整株死亡,为此作防治试验。供试药剂(因素)以及浓度(水平)见表 6.10。并制成混合液进行防治,表中各水平排列经随机化处理,各条件重复试验二次,每隔 12 小时观察记录一次死虫数,72 小时的死亡率经校正并作反正弦变换后的数值作为试验指标,结果见表 6.11

表 6.10

水 平 \ 因 素	A:20%氧化乐果	B:水胺硫磷	C:快灭磷
1	原液	原液	原液
2	1:10	1:10	1:10
3	1:20	1:20	1:20

1. 计算各项离差平方和

$$\text{校正项 } C = \frac{T^2}{N \cdot L} = \frac{1221.64^2}{9 \times 2} = 82911.35$$

$$\begin{aligned} L_A &= \frac{1}{6} \sum_i K_{Ai}^2 - C \\ &= \frac{1}{6} (416.09^2 + 399.26^2 + 406.29^2) - C = 23.82 \end{aligned}$$

$$L_B = \frac{1}{6} \sum_i K_{Bi}^2 - C = 171.95$$

$$L_C = \frac{1}{6} \sum_i K_{Ci}^2 - C = 368.70$$

$$L_{e_1} = \frac{1}{6} \sum_{i=1}^3 K_{\frac{1}{2}i}^2 - C = 179.40$$

$$\begin{aligned} L_{e_2} &= \sum_i \sum_{j=1}^2 x_{ij}^2 - \frac{1}{2} \sum_{i=1}^9 T_i^2 \\ &= 83854.75 - 83655.21 = 199.54 \end{aligned}$$

第一类误差自由度为 2, 第二类误差自由度为 $9 \times (2-1) = 2$.

表 6.11

试验号	因 素				指 标 x_{ij}		合计 T
	A	B	C	D(空白)	I	II	
1	1	1	1	1	64.30	63.87	128.17
2	1	2	2	2	70.18	77.89	148.07
3	1	3	3	3	71.57	68.28	139.85
4	2	1	2	3	65.57	69.64	135.21
5	2	2	3	1	77.08	66.89	143.97
6	2	3	1	2	55.92	64.16	120.08
7	3	1	3	2	53.61	64.01	117.62
8	3	2	1	3	62.31	66.66	128.97
9	3	3	2	1	78.03	81.67	159.70
K_1	416.09	381	377.22	431.84	$T = 1221.64$		
K_2	399.26	421.01	442.98	385.77			
K_3	406.29	419.63	401.44	404.03			
k_1	69.35	63.50	62.87	71.97			
k_2	66.54	70.17	73.83	64.29			
k_3	67.72	69.94	66.91	67.35			
R	2.81	6.67	10.96				

2. 形成统计量 F

$$F = \frac{L_{r_1}/2}{L_{r_2}/9} = \frac{179.40/2}{199.54/9} = 4.046 < F_{\alpha, 05}(2, 9) = 4.26$$

$F < F_{\alpha}$ 认为第一类误差基本上只包含试验误差, 从而我们将第一、第二类误差合并为试验误差。合并后即可按通常方差分析方法得出应有的结论。

3. 列出方差分析表, 见表 6.12

表 6.12

变差来源	离差平方和	自由度	均方	均方比	F 值
A	23.82	2	11.91	0.346	$F_{0.05} = 3.98$
B	171.95	2	85.97	2.496	$F_{0.01} = 7.21$
C	368.69	2	184.35	5.35	
e_1	179.40	2	34.45		
e_2	199.54				
总和	943.4	17			

从表可以作出推断:快灭磷对试验结果有显著影响,三个因子中,以快灭磷最佳,其次为氧化乐果,水胺硫磷最差,这三种药物的最佳水平组合为水胺硫磷 1:10,氧化乐果原液,快灭磷 1:10。

第四节 水平数不等的试验设计

前面我们着重讨论了水平数相同的试验,但在林业生产和科研中常需进行各因子水平数不完全一样的试验。这就是水平数不等的试验设计问题。在此仍以实例说明水平数不同的试验如何直接应用混合水平正交表进行试验设计与统计分析。

混合水平正交表是通过水平数相等的正交表进行并列,改造而得到的,它保持了基本的正交性条件,其使用方法大致与前同。

例 6.4 1986 年在福建省沙县 屏州伐木场进行采挖松根后对水土保持影响的试验研究。采用水平数不等的试验设计,并选用混合型正交表 $L_3(4 \times 2^4)$,具体因素和水平设置见

表 6.13

表 6.13

水 平 \ 因 素	A: 坡度	B: 炼山	C: 挖根	D: 整地
1	25°	不炼山	不挖	不整地
2	30°	炼山	挖	整地
3	35°			
4	40°			

试验指标为年径流量($\text{m}^3/\text{公顷}$)

现直接将试验结果连同表 $L_8(4 \times 2^4)$ 一起放在表 6.14 中。

先作直观分析,仍由各列 k 值大小决定水平选择,但关于因子主次关系的确定则需作调整,因为当两因子对指标有同等影响时,水平多的因子自然会有较大的极差,因此对不同水平数可以查阅以下折算系数表,查得相应的折算系数,乘以原极差值得到修正的极差值 R' ,作为可比较的极差值再观察各因子主次关系。

表 6.14

试验号	因 素				空白	年径流量
	A	B	C	D		
1	1	1	1	1	1	221.2858
2	1	2	2	2	2	217.9006
3	2	1	1	2	2	294.4601
4	2	2	2	1	1	177.5870
5	3	1	2	1	2	187.1011
6	3	2	1	2	1	262.5204
7	4	1	2	2	1	157.9443
8	4	2	1	1	2	130.7602
K_1	439.186	860.7913	909.0265	716.7341	$T = 1649.5595$	
K_2	472.0471	788.7682	740.5330	932.8254		
K_3	449.6215					
K_4	288.7045					
k_1	219.59	215.20	227.26	179.18	$T = 1649.5595$	
k_2	236.02	197.19	185.13	233.21		
k_3	224.81					
k_4	144.35					
R	183.3426	72.0231	168.4935	216.0913		
R'	184.0928	64.8208	151.6442	194.4822		

折算公式为

$$R' = R \cdot d \sqrt{n} \quad (6.2)$$

R' —— 折算后的修正极差

R —— 原来的极差值

d —— 折算系数

n —— 每一水平的试验次数

表 6.15 极差折算系数表

水平数	2	3	4	5	6	7	8	9	10
折算系数 d	0.71	0.52	0.45	0.40	0.37	0.35	0.34	0.32	0.31

对例 6.4 数据,我们有

$$R'_A = R_A \cdot d_A \sqrt{2} = 183.3426 \times 0.71 \times \sqrt{2} = 184.0928$$

$$R'_B = R_B \cdot d_B \sqrt{4} = 72.0231 \times 0.45 \times \sqrt{4} = 64.8208$$

$$R'_C = R_C \cdot d_C \sqrt{4} = 168.4935 \times 0.45 \times \sqrt{4} = 151.6442$$

$$R'_D = R_D \cdot d_D \sqrt{4} = 216.0913 \times 0.45 \times \sqrt{4} = 194.4822$$

并将它列出在表 6.14 的最后一行里,由此看出各因子的主次关系为

整地→坡度→挖根→炼山

主————→次

下面再进行方差分析

$$\begin{aligned} \text{校正项 } C &= \frac{(\sum_i x_i)^2}{n} = \frac{T^2}{n} = \frac{1649.5595^2}{8} \\ &= 340130.818 \end{aligned}$$

$$L_T = \sum x_i^2 - C = 20529.5657$$

$$L_A = \frac{1}{2} \sum_{i=1}^4 K_{Ai}^2 - C = 10480.6521$$

$$L_B = \frac{1}{4} \sum_{i=1}^2 K_{Bi}^2 - C = 648.4159$$

$$L_C = \frac{1}{4} \sum_{i=1}^2 K_{Ci}^2 - C = 3548.7574$$

$$L_D = \frac{1}{4} \sum_{i=1}^2 K_{Di}^2 - C = 5836.9312$$

$$L_e = L_T - L_A - L_B - L_C - L_D = 14.8091$$

由此列出方差分析表 6.16

表 6.16

变差来源	离差平方和	自由度	均方	均方比	F 值
A	10480.6521	3	3493.5507	235.91*	$F_{0.01}(3,1)=216$
B	648.4159	1	648.4159	43.78	$F_{0.05}(3,1)=540$
C	3548.7574	1	3548.7574	239.63*	$F_{0.01}(1,1)=405$
D	5836.9312	1	5836.9312	394.14*	$F_{0.05}(1,1)=161$
e	14.8091	1	14.8091		
合计	20529.5627	7	—	—	

试验结果认为除炼山外,坡度、挖掘松根、整地这三个因子对年径流量影响都是显著的。换言之,采挖松根过程中,不同坡度,整不整地,挖不挖根都会造成年径流量的显著差异。进一步多重比较后可对采挖松根对水土保持的影响作出评价。

第五节 单纯形最优化方法的应用

正交试验设计能寻找被试验因素中最主要的影响因子以及最优因素水平组合。但是当因素水平间隔较大时,为了寻找更好更精确的最优水平组合,就必须分批进行试验,象捕鱼撒网式逐渐缩小寻找最优区域,以求得最优值。这种处理方法试验次数也较多,特别当实验因素增加,所需的实验次数将按指数规律增加。因此在应用正交试验设计方法解决存在交互作用的多因素最优化时,也存在缺点。

Spendley (1962 年)首先提出一种高效率的连续最优化的单纯形法。该法在国外称 Simplex 法,其基本思想是从一个

初始的单纯形开始,按照一定推移法则“运动”,不断产生新的单纯形,迫使单纯形不断地向最优区域移动,并按预定精度充分地接近最优点。

所谓单纯形是一个几何图形,它由 m 维空间中的 $m+1$ 个点 $p_0, p_1, \dots, p_m$ 构成,且 $p_0 - p_1, \dots, p_0 - p_m$ 线性无关。因此在二维空间单纯形是一个三角形,在三维空间是一个四面体,在高维空间则是一个超多面体。这些几何图形的每个顶点相当各个实验点,其坐标值就是每个实验点相应的各个实验参数的值。基本单纯形法(Basic Simplex Method, 简称为 BSM 法)是通过单纯形中最坏响应点的“反射”来实现其运动的,其运动遵循下法则

I、每观察响应值一次,单纯形就向前推移一步。

II、找出当前单纯形中最坏响应点 P_w , 求出 P_w 相对剩余 m 个点的形心点 P_c 的镜反射点 P_r , 用 P_r 代替 P_w , 构成一个相邻的新单纯形,从而实现单纯形的推移。

对于一个任意维的单纯形,假如其顶点用坐标矢量 $P_1, P_2, \dots, P_w, \dots, P_m, P_{m+1}$ 来表示,放弃 P_w , 剩余点 $P_1, P_2, \dots, P_{w-1}, P_{w+1}, \dots, P_m, P_{m+1}$ 的形心点 P_c 可用下式计算

$$P_c = \frac{1}{m} (P_1 + P_2 + \dots + P_{w-1} + P_{w+1} + \dots + P_m + P_{m+1}) \quad (6.3)$$

那么 P_w 关于 P_c 的反射点 P_r 可用下式计算:

$$P_r = P_c + (P_c - P_w) \quad (6.4)$$

III、如果反射点在新的单纯形中是最坏响应点,按 I 法则,单纯形经再次反射后会重复先前的单纯形,这样会引起单纯形的振荡并陷入困境,这时则不要执行法则 II,而应用次坏响应点 P_s 代替 P_w 进行反射,并继续下去。

尽管实验存在偶然误差,但一般情况下重测各点的响应是不必要的,因为错误的推移必定会得到低响应的结果,这种结果必定会被法则Ⅰ和Ⅲ所修正。

Ⅳ、为了排除偶然的假高响应点,如果某一顶点保留在 $m+1$ 个单纯形中,在应用法则Ⅰ以前,重复测定这一顶点的响应。

Ⅴ、如果某一新点落在某变量的定义域以外,或对目标函数的表达式是无意义的,则不必进行实验观测,只要令此点为很坏的响应即可。

单纯形由 m 维空间的 $m+1$ 个点所构成的几何图形。单纯形法具体实施方法,首先根据一个无量纲变量构成矩阵 P :

$$P = \begin{bmatrix} P_1 & P_2 & \cdots & P_j & \cdots & P_m \\ -P_1 & P_2 & \cdots & P_j & \cdots & P_m \\ 0 & -2P_2 & \cdots & P_j & \cdots & P_m \\ 0 & 0 & & \vdots & & \vdots \\ \vdots & 0 & & -jP_j & & \vdots \\ \vdots & \vdots & & 0 & & \vdots \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & -mP_m \end{bmatrix} \quad (6.5)$$

$$P_j = \sqrt{\frac{1}{2j(j+1)}} \quad (6.6)$$

构成一个 m 维的正单纯形,它由 m 列, $m+1$ 行构成,并且以无量纲形式表示 m 个因子的实验设计。首先根据经验或初步实验确定初始单纯形,根据初始单纯形的参数进行实验,然后分析实验结果,找出最坏点并依上述法则组织新的单纯形再进行试验。

上述 BSM 法存在精度与收敛速度间矛盾。若采用尺寸

较大的初始单纯形,可以很快就推移到最优区域,但求出的最优点精度不够,可能是局部的,若采用尺寸较小的初始单纯形,则可以推移到较精确的最优点上,但所需推移次数增加。

为了克服上述矛盾 Nelder 和 Mead 于 1965 年提出了改进单纯形法。该法在 BSM 法基础上增加了“扩张”和“压缩”两功能。这两个功能既能加速单纯形的前进,又能按预定精度充分地接近最优点。

假设 P_b, P_n, P_w 分别为一个单纯形中最好、次好和最坏响应点,其对应的响应值分别为 f_b, f_n, f_w 。反射点 P_r 的响应值可能出现下列三种情况:

I、假如 f_r 比 f_b 还要好,则说明反射方向正确,应沿 $P_w P_r$ 方向进一步探讨,并按下式求扩张点 P_c 。

$$P_c = P_r + r(P_r - P_w) \quad (6.7)$$

此处 r 是扩张系数($r > 1$)

测定 P_c 的响应 f_c ,若 f_c 比 f_b 还好,则新的单纯形为 $P_b P_n P_c$ 否则就是 P_b, P_n, P_r 。

II、假如 f_r 比 f_b 差,但又比 f_n 好,此时既不扩张也不压缩,新的单纯形为 $P_b P_n P_r$ 。

假如 f_r 比 f_n 差,则说明反射方向不正确,应予压缩,此时又有两种情况,如果 f_r 比 f_w 还差,压缩点 P_i 应靠近 P_w 。

$$P_i = P_r - \beta(P_r - P_w) \quad (6.8)$$

此处 β 为压缩系数($0 < \beta < 1$), $P_b P_n P_i$ 构成新的单纯形。

如果 f_r 比 f_w 好,压缩 P_n 应靠近 P_r 。

$$P_u = P_n + \beta(P_r - P_n) \quad (6.9)$$

$P_b P_u P_r$ 构成新的单纯形。

III、假如 P_r 和 P_u 的响应 f_r 和 f_u 比 f_w 还差,还可以将现的单纯形 $P_b P_n P_w$ 向 P_b 收缩二分之一,构成新单纯形。

例 6.5 吴保国等(1991 年)为了改善木素基乳化剂的乳化性能,考虑四种添加剂和一种乳化剂 A, B, C, D, E 五个因子。观察指标为表面张力。首先应用 $L_{25}(5^5)$ 正交表安排实验,并进行直观分析。各因子水平表见表 6.17

表 6.17

单位:浓度(%)

因 子	水 平 数				
	1	2	3	4	5
A	60	80	100	40	20
B	30	20	10	5	0
C	40	20	10	5	0
D	40	10	0	20	5
E	20	0	5	10	40

* 表中数据为作者推算

实验安排与结果见表 6.18。

实验结果因子 c 的水平变化对表面张力影响很小,所以用单纯形优化时,去掉此因子,对正交试验所得最佳参数进行筛选,现以 A_2, B_3, D_4, E_4 四个因子作为单纯形的零水平。根据经验 A 因子的步长为 0.2, B 因子的步长为 0.1, D 因子的步长为 0.1, E 因子步长为 0.1,这样组成了一个 4×5 的矩阵, P_i 按下式求出:

$$P_1 = \frac{B-0.1}{0.1}, P_2 = \frac{E-0.1}{0.1}, P_3 = \frac{A-0.8}{0.2}, P_4 = \frac{D-0.2}{0.1}$$

表 6.18

	A	B	C	D	E	$\sigma(10^{-3}N/cm)$
1	1	1	1	1	1	45.9
2	1	2	2	2	2	47.8
3	1	3	3	3	3	42.8
4	1	4	4	4	4	46.3
5	1	5	5	5	5	50.2
6	2	1	2	4	5	36.8
7	2	2	3	5	1	42.6
8	2	3	4	1	2	40.7
9	2	4	5	2	3	44.7
10	2	5	1	3	4	47.7
11	3	1	3	2	4	42.5
12	3	2	4	3	5	56.8
13	3	3	5	4	1	40.7
14	3	4	1	5	2	45.5
15	3	5	2	1	3	45.6
16	4	1	4	5	3	50.7
17	4	2	5	1	4	41.3
18	4	3	1	2	5	41.3
19	4	4	2	3	1	52.5
20	4	5	3	4	2	54.9
21	5	1	5	3	2	59.1
22	5	2	1	4	3	42.5
23	5	3	2	5	4	42.5
24	5	4	3	1	5	49.4
25	5	5	4	2	1	50.0
I	46.6	47.0	44.6	44.6	46.3	
平均 I	42.5	46.2	45.0	45.3	49.6	
σ II	46.2	41.6	46.4	51.8	45.3	
IV	48.1	47.6	48.9	44.2	44.1	
V	48.7	49.6	47.2	46.3	46.9	

用此值构成初始单纯形, 仍用表面张力作为考察指标。初始单纯形数据和实验结果列于表

表 6.19

N	B	E	A	D	$\sigma(10^{-5}\text{N/cm})$
1	0.15	0.13	0.82	0.22	36.92
2	0.05	0.13	0.82	0.22	36.38
3	0.10	0.04	0.82	0.22	37.69
4	0.10	0.10	0.74	0.22	36.30
5	0.10	0.10	0.80	0.14	37.23

由表 6.19 数据可以看出最坏点为 3, 因为此点的表面张力最大, 剔除此点, 算出新点 6。

形心点 $B^c=0.1$ $E^c=0.115$ $A^c=0.795$ $D^c=0.2$

点 6: $B_6=0.1$ $E_6=0.19$ $A_6=0.77$ $D_6=0.18$

用 1、2、5、6 组成新的单纯形列于表 6.20

表 6.20

N	B	E	A	D	$\sigma(10^{-5}\text{N/cm})$
1	0.15	0.13	0.82	0.22	36.92
2	0.05	0.13	0.82	0.22	36.38
4	0.10	0.10	0.74	0.22	36.30
5	0.10	0.10	0.80	0.14	37.23
6	0.10	0.19	0.77	0.18	37.03

表 6.20 所列情况属于 I 情况, 即新点响应值介于最优值和最坏值之间, 不需要扩张和压缩, 故选 $\beta=1$, 计算点 7 的参数, 点 7 实验的响应值为 35.91, 属于情况 I, 故选 $r=2$, 进一步得到更优点响应值 35.83, 但做点 9 时, 响应值又回升到 36.53, 故又选 $\beta=1$ 继续优化。如此反复优化, 共用 12 次实验, 得到一个最佳响应值 35.50。各因子浓度值分别为

$A:0.81$, $B:0.1$, $D:0.29$, $E:0.07$

从实例中看到, 把正交实验设计与改进单纯形法结合起来, 可以改变初始单纯形选点的盲目性, 较一般单因子实验设

计大大减少实验次数,并且提高了实验精度。单纯形法突出优点是不需要知道各因子参数间的函数表达式,只是寻找函数最陡变化方向。不管变量有多少,向前移动一步只做一次实验,故方法显得简单快速,并且特别适合于具有交互作用的多因子优化实验。

第七章 平衡不完全区组设计

第一节 平衡不完全区组设计

平衡不完全区组设计(Balanced Incomplete Blocks)简称 BIB 设计。

在随机区组设计中,每区组内的小区与处理数相等,从而每一处理恰好配置在一个小区内,各区组之间尽管有差异,但在每一区组内,由于环境条件大致相同,全部处理可以直接比较。当因素水平数(处理数)较多时,象林木的种源试验,水平数多到上百个,此时能满足局部控制条件的区组却容纳不下这样多的处理数,也就是说,每一区组只能容纳一部分处理,这时可采用平衡不完全区组设计。

平衡不完全区组设计是应用平衡不完全区组设计表来设计试验方案的一种试验设计法。所谓不完全是指一个区组内不是把所有处理都配置进来,例如试验有 V 个处理,而每个区组若只能容纳 $K(K < V)$ 个处理,这样,一个区组内的处理是不完全的,即不完全区组之意。而这时我们希望将 V 个处理合理地安排到全部区组中,原则是使每种处理在一个区组内最多出现一次,并且任何一对处理在试验中出现的次数都相同,这种设计就是平衡的。

例 7.1 某林业局苗圃,为比较 4 种药物防治杉木猝倒病的效果,由于药物喷撒要求范围稍大,以垄为区组时,每个区组只能容纳 2 个处理(药剂),因此采用平衡不完全区组设计。利用现存的 BIB 试验设计表对药物防治这个具体问题,

安排了 12 次试验,得表 7.1,其中假定已直接取得了试验结果——猝倒病发病率(%)

表 7.1

处理	区 组						$x_{i.}$
	1	2	3	4	5	6	
1	64.6		63.8		59.8		188.2
2	40.9			39.9		42.6	123.4
3		82.2	83.2			85.2	250.6
4		61.1		38.5	-11.3		88.3
$\cdot x_{.j}$	105.5	143.3	147	78.4	48.5	127.8	650.5

这个设计中每种处理在每一区组内最多只出现一次,每一对处理组合在全部试验中均出现一次,每一处理在全部试验中都出现三次。

平衡不完全区组设计的特点:

(1)各处理的重复数是相同的,记重复数为 r ,则上例中 $r=3$ 。

(2)任意一对处理在同一区组进行比较的次数相同,即每一对处理在全部试验中出现次数相同,记其为 λ ,则上例 $\lambda=1$ 。

(3)每一区组包含 K 个处理, $K < V$, 这里 $K=2, V=4, K$ 也就是每区组小区数。

由于这些特点,处理仍然可以合理比较

假设区组数为 $b, b \geq V$, 每区组的小区数为 K , 则总的小区数,亦即总的观察数 N 为 $N=b \cdot K$, 又由于每一处理重复 r 次,故 $N=b \cdot K=V \cdot r$ (7.1)

因为每一处理在全部区组内重复 r 次,在每一区组内最多出现一次,故任一处理 j ,可以在 r 个区组中出现,每个区组

内有 k 个小区,因而第 j 号处理在 r 个区组里将有 $r(k-1)$ 个小区被除 j 之外的 $(V-1)$ 个处理所占有。显然,每一对处理 (j,i) 在同一区组内出现的次数 λ 应为:

$$\lambda = \frac{r(K-1)}{V-1} \quad (7.2)$$

$$\text{即 } \lambda(V-1) = r(K-1) \quad (7.3)$$

由此还可得到 $\lambda < r$ 。

综上所述,可知 BIB 设计的必要条件是 b, K, V, r, λ 五个参数都应为正整数,并且满足

$$(1) Vr = bK$$

$$(2) \lambda < K \quad K < V \quad V \leq b$$

$$(3) \lambda(V-1) = r(K-1)$$

对任意一组满足上面条件的五个参数,可以形成一张 BIB 设计表,类似表 7.1,利用这种表与正交设计一样,即可完成一个 BIB 设计。

平衡不完全区组设计的统计模型为

$$x_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij} \quad (i=1,2,\dots,V \quad j=1,2,\dots,b) \quad (7.4)$$

这个模型与两向分组方差分析模型相似。

但现在,在 BIB 设计中,平方和分解式为

$$L_T = L_{\text{处理(调整后)}} + L_{\text{区组}} + L_e \quad (7.5)$$

式中调整后的处理项平方和相当于随机区组设计中的处理项平方和,由于 BIB 中每一个处理出现在 r 个不同的区组中,不同处理出现在不同的 r 个区组中,我们从样本资料中直接能计算的总和处理 $x_{1.}, x_{2.}, \dots, x_{r.}$ 之间的差异,不仅包含处理间的差异,实际上还包含由于不同处理所在区组不同带来的区组差异,由此计算出的处理平方和为未调整的,对处理进

行调整的目的就是要消除不同区组差异对处理的影响,以真实反映处理的效应。 $L_{\text{处理(调整后)}}$ 可按 7.6 式计算

$$L_{\text{处理(调整后)}} = \frac{K \sum_{i=1}^V Q_i^2}{\lambda V} \quad (7.6)$$

其中 Q_i 是调整的第 i 处理总和

$$Q_i = x_i - \frac{1}{K} \sum_{j=1}^b n_{ij} x_{.j} \quad (i=1, 2, \dots, V) \quad (7.7)$$

其中 $n_{ij} = \begin{cases} 1 & \text{当第 } i \text{ 处理出现在第 } j \text{ 区组中时} \\ 0 & \text{否则} \end{cases}$

$\frac{1}{K} \sum_{j=1}^b n_{ij} x_{.j}$ 是包含第 i 处理的区组总平均

调整处理项自由度为 $V-1$ 。

区组的离差平方和 $L_{\text{区组}}$ 的计算是常规的,自由度为 $b-1$ 。

这样由总自由度 $N-1$, 可以得出误差项自由度为

$$N-1-(b-1)-(V-1)=(N-V-b+1)$$

检验处理是否显著的统计量为

$$F = \frac{S_{\text{处理(调整后)}}^2}{S_e^2}$$

在 H_0 : 处理差异不显著条件下, F 应该服从 $F(V-1, N-V-b+1)$ 分布

BIB 设计的方差分析可归结为表 7.2

表 7.2

变差来源	离差平方和	自由度	均 方	均方比
处 理 (调整后)	$\frac{K}{\lambda V} \sum_{i=1}^v Q_i^2$	$V-1$	$\frac{L_{\text{处理(调整后)}}}{V-1} = S_{\text{处}}^2$	$\frac{S_{\text{处}}^2}{S_{\text{误}}^2}$
区 组	$\frac{1}{K} \sum_{j=1}^b x_{.j}^2 - C$	$b-1$	$\frac{L_{\text{区组}}}{b-1} = S_{\text{区}}^2$	$\frac{S_{\text{区}}^2}{S_{\text{误}}^2}$
误 差	$L_T - L_{\text{处理(调整后)}} - L_{\text{区组}}$	$\frac{N-v}{-b+1}$	$\frac{L_{\text{误差}}}{N-V-b+1} = S_{\text{误}}^2$	
总 和	$\sum_i \sum_j x_{ij}^2 - C$	$N-1$		

各处理效应估计量为

$$\hat{a}_i = \frac{K}{\lambda V} Q_i$$

调整后处理平均数为

$$\bar{x}_i(\text{调整后}) = \hat{a}_i + \bar{x} \quad \bar{x} = \frac{x_{..}}{N}$$

关于已提出的例子,我们有

$$V=4, \quad b=6, \quad K=2, \quad r=3,$$

由 7.2 式

$$\lambda = \frac{r(K-1)}{V-1} = \frac{3(2-1)}{4-1} = 1$$

$$N=b \quad K=V \quad r=12$$

$$C = \frac{x_{..}^2}{N} = \frac{650.5^2}{12} = 35262.52$$

$$L_T = \sum_{i=1}^4 \sum_{j=1}^6 x_{ij}^2 - C = 7627.97$$

$$L_{\text{区组}} = \frac{1}{K} \sum_{j=1}^6 x_{.j}^2 - C = \frac{1}{2} (105.5^2 + \cdots + 127.8^2) - C \\ = 3500.38$$

$$Q_1 = x_{1.} - \frac{1}{K} \sum_{j=1}^6 n_{1j} x_{.j} = 188.2 - \frac{1}{2} (105.5 + 147 + 48.5)$$

$$=188.2-150.5=37.7$$

$$Q_2 = x_2 - \frac{1}{K} \sum n_{2j} x_{.j} = -32.45$$

$$Q_3 = x_3 - \frac{1}{K} \sum n_{3j} x_{.j} = 41.55$$

$$Q_4 = x_4 - \frac{1}{2} (143.3 + 78.4 + 48.3) = 88.3 - 135.1 \\ = -46.8$$

$$L_{\text{处理(调整后)}} = \frac{K \sum Q_i^2}{\lambda V} = \frac{2 \times 6390.935}{1 \times 4} = 3195.47$$

$$L_e = L_T - L_{\text{区组}} - L_{\text{处理(调整后)}} = 932.12$$

$$f_T = N - 1 = 11 \quad f_{\text{区组}} = b - 1 = 5 \quad f_{\text{处理(调整后)}} = V - 1 = 3$$

$$f_e = f_T - f_{\text{区组}} - f_{\text{处理(调整后)}} = 3$$

列出方差分析表 7.3

表 7.3

变差来源	离差平方和	自由度	均方	均方比	F 值
区组	3500.38	5	700.076	2.25	
处理(调整)	3195.47	3	1065.157	3.43	$F_{0.05}(3,3)=9.28$
误差	932.12	3	310.71		
总和	7627.97	11			

结论是四种药物对防治杉木猝倒病效果均没有达到显著水平。

下面计算各处理的调整后均数

$$\bar{x} = \frac{650.5}{12} = 54.21$$

$$\bar{x}_1 = \hat{a}_1 + \bar{x} = \frac{K}{\lambda V} Q_1 + \bar{x} = \frac{2}{1 \times 4} \times 37.7 + 54.21 = 73.06$$

$$\bar{x}_2 = 37.985 \quad \bar{x}_3 = 74.985 \quad \bar{x}_4 = 30.81$$

BIB 设计在林木遗传育种,种源试验中均属常用设计方

法,请再看应用实例。

例 7.2 某地作 8 个杉木单亲子代与一个对照家系共 9 个处理进行子代测定,试用 BIB 设计。试验地有 I、II、III、IV 四个条件不相似的大范围,每个范围内分划为三个不同区组,每区组可划分为三个小区。

问题给出参数如下:

处理数 $V=9$; 区组数 $b=12$; 重复数 $r=4$; 区组内小区数 $K=3$ 。

$$\lambda = \frac{r(K-1)}{(V-1)} = \frac{4 \times (3-1)}{8} = 1$$

查阅 BIB 设计表,我们找到了与此对应的模式,其配置图如下:

I				II				III				IV			
(1)	1	2	3	(4)	1	4	7	(7)	1	5	9	(10)	1	6	8
(2)	4	5	6	(5)	2	5	8	(8)	2	6	7	(11)	2	4	9
(3)	7	8	9	(6)	3	6	9	(9)	3	4	8	(12)	3	5	7

() 内数字为区组号; I、II、III、IV 为重复数

每试验小区内株数为 4, 以小区树高平均值为指标。

根据模式配置图,我们立即可以完成试验的设计,并得到试验结果一并列于表 7.4。

$$N = b \cdot K = 12 \times 3 = 36$$

$$C = \frac{x_{..}^2}{N} = \frac{294.84^2}{36} = 2414.7396$$

$$L_T = \sum \sum x_{ij}^2 - C = 23.844$$

表 7.4

家系	区 组												x_i
(处理)	1	2	3	4	5	6	7	8	9	10	11	12	
V_1	9.40			7.38			8.62			8.54			33.94
V_2	8.68				9.52			8.52			8.00		34.72
V_3	9.36					9.24			8.42			9.08	36.10
V_4		9.10		7.40					8.14		8.20		32.84
V_5		6.70			7.14		7.32			8.14		6.86	28.02
V_6		8.72				9.06		9.32					35.24
V_7			6.46	7.30				8.60				8.56	30.92
V_8			8.20		7.90				7.26	7.86			31.22
$V_9(CK)$			8.52			8.26	7.22				7.84		31.84
$x_{..}$	27.44		23.18		24.56		23.16		23.82		24.04		294.84
		24.52		22.08		26.56		26.44		24.54		24.50	

$$L_{\text{区组}} = \frac{1}{K} \sum_{j=1}^{12} x_{.j}^2 - C = \frac{1}{3} (27.44^2 + \cdots + 24.50^2) - C$$

$$= 8.889$$

$$L_{\text{处理(调整后)}} = \frac{K \sum_{i=1}^v Q_i^2}{\lambda V}$$

$$Q_1 = x_{1.} - \frac{1}{K} \sum_{j=1}^{12} n_{1j} x_{.j} = 33.94 - \frac{1}{3} (27.44 + 22.08 + 23.16 + 24.54)$$

$$= 33.94 - 32.41 = 1.53$$

$$Q_2 = x_{2.} - \frac{1}{K} \sum_{j=1}^{12} n_{2j} x_{.j} = 34.72 - 34.16 = 0.56$$

为方便计, 令 $V_i = \sum_{j=1}^{12} n_{ij} \cdot x_{.j}$, 全部 Q_i 的计算列在表 7.5 中

表 7.5

项目	1	2	3	4	5	6	7	8	9
V_i	97.22	102.48	102.32	94.46	96.74	102.06	96.20	96.10	96.94
$\frac{1}{K}V_i$	32.41	34.16	34.11	31.49	32.25	34.02	32.07	32.03	32.31
x_i	33.94	34.72	36.10	32.84	28.02	35.24	30.92	31.22	31.84
Q_i	1.53	0.56	1.99	1.35	-4.23	1.22	-1.15	-0.81	-0.47

于是 $\sum Q_i^2 = 30.0179$

$$L_{\text{处理(调整后)}} = \frac{3 \times 30.0179}{1 \times 9} = 10.006$$

$$L_e = L_T - L_{\text{区组}} - L_{\text{处理(调整后)}} \\ = 23.844 - 10.006 - 8.889 = 4.949$$

$$f_T = N - 1 = 35, \quad f_{\text{区组}} = 6 - 1 = 12 - 1 = 11$$

$$f_{\text{处理(调整后)}} = V - 1 = 9 - 1 = 8 \quad f_e = 35 - 11 - 8 = 16$$

由此可列出方差分析表 7.6

表 7.6

变差来源	自由度	离差平方和	均方	均方比	F 值
区组	11	8.889	0.808	2.615*	$F_{0.01}(11,16) = 3.62$
家系	8	10.006	1.251	4.049*	$F_{0.05}(11,16) = 2.45$
误差	16	4.949	0.309		$F_{0.01}(8,16) = 3.89$ $F_{0.05}(8,16) = 2.59$
总和	35	23.844	—	—	—

结论是：区组间有显著差异，各家系间差异极显著。进而按 $\bar{x}_{i(\text{调整后})} = \hat{a}_i + \bar{x} = \frac{K}{\lambda V} - Q_i + \bar{x}$ 计算每一处理的调整后平均数

$$\bar{X} = \frac{X_{..}}{N} = \frac{294.84}{36} = 8.19$$

由表 7.5 最后一列,可以立即得到 $\bar{x}_i$,我们将它们列在表 7.7 中。

表 7.7

处理	1	2	3	4	5	6	7	8	9
$\bar{x}_i$ (调整后)	8.70	8.38	8.85	8.64	6.78	8.60	7.81	7.92	8.03

可以用传统方法对这些调整后处理平均数作出多重比较。但不在此一一列出了。

第二节 平衡格子设计

平衡格子设计是不完全区组设计的特殊形式。当处理 V 可以开平方时可使用。

一、设计要求

(1)处理数 V 是某个数的平方。如 $V=16$,它是 4 的平方。

(2)每个区组内的小区数(k)是处理数的平方根。

$$k = \sqrt{V}, \text{如 } V=16, \quad k = \sqrt{V} = \sqrt{16} = 4.$$

(3)重复数是 $r=k+1$

(4)区组数为 $b=k(k+1)$ 。如 $v=16, k=4$, 则
 $b=k(k+1)=4(4+1)=20$ 。

可列出表 7.8

表 7.8

处理数	每一区组内小区数	重复数	区组数
9	3	4	12
16	4	5	20
25	5	6	30
49	7	8	56
64	8	9	72
81	9	10	90

注：处理数为 36 时，安排不出这样设计

二、具体作法

例如，设有 9 个杉木种源参加试验，每个区组安排 3 个种源，重复 4 次，共有 12 个区组。

具体步骤是将种源按以下方法分组。

(1) 把种源按顺序排成三行，每行有 3 个种源，见图 7.1 中(I)作为第一群。

(2) 把 I 中在同一列的种源改成一列，见图 7.1 中(II)，作为第二群。

(3) 按两个方向把在同一对角线的种源，排成一列，见图 7.1 中(III)(IV)，这样把 9 个种源分成 4 群 12 组。

1 组	1 2 3	4 组	1 4 7	7 组	1 5 9	10 组	7 5 3
2 组	4 5 6	5 组	2 5 8	8 组	2 6 7	11 组	4 2 9
3 组	7 8 9	6 组	3 6 9	9 组	4 8 3	12 组	8 6 1
(I)		(II)		(III)		(IV)	

图 7.1

每群作为一个重复。在每重复内随机地把重复内区组排好，然后再把每一区组内的处理(种源)随机地设置在小区，见图 7.2。

2	5 4 6	4	7 4 1	7	1 9 5	12	7 5 3
1	1 2 3	5	5 2 8	9	2 6 7	10	6 8 1
3	9 7 8	6	6 3 9	8	3 8 4	11	2 4 9
(I)		(II)		(III)		(IV)	

图7.2 3×3 平衡格子设计配置图

($v=9, k=3, r=4, \lambda=1, b=12$)

这种设计限制条件:

(1)处理数必须是某数的平方。若处理数不是某数的平方,但与之接近,可用增加处理(或对照)或减少处理数来满足这个限制。

(2)重复数必须是 $K+1$,如认为重复次数太多或不够时,则需改用其他设计。

由图 7.1 可以看出,每对品种在同一区组只同时出现一次,这样排列编组可以保证所有品种能在相同的精度下相互比较,故这种设计叫做平衡格子设计。另外,当同重复内区组间变异不大,或对受环境影响不大的观察值,每一重复可看作是一个随机区组进行统计分析。

$4 \times 4, 5 \times 5$, 平衡格子设计分组见下图 7.3, 图 7.4,

(1)	1	2	3	4	(5)	1	5	9	13	(9)	1	6	11	16
(2)	5	6	7	8	(6)	2	6	10	14	(10)	5	2	15	12
(3)	9	10	11	12	(7)	3	7	11	15	(11)	9	14	3	8
(4)	13	14	15	16	(8)	4	8	12	16	(12)	13	10	7	4
区组 (I)					区组 (II)					区组 (III)				

(13)	1	14	7	12
(14)	13	2	11	8
(15)	5	10	3	16
(16)	9	6	15	4

区组 (N)

(17)	1	10	15	8
(18)	9	2	7	16
(19)	13	6	3	12
(20)	5	14	11	4

区组 (V)

图 7.3 4×4 平衡格子设计
($v=16, k=4, r=5, b=20, \lambda=1$)

(1)	1	2	3	4	5
(2)	6	7	8	9	10
(3)	11	12	13	14	15
(4)	16	17	18	19	20
(5)	21	22	23	24	25

区组 (I)

(6)	1	6	11	16	21
(7)	2	7	12	17	22
(8)	3	8	13	18	23
(9)	4	9	14	19	24
(10)	5	10	15	20	25

区组 (II)

(11)	1	7	13	19	25
(12)	21	2	8	14	20
(13)	16	22	3	9	15
(14)	11	17	23	4	10
(15)	7	12	18	24	5

区组 (III)

(16)	1	12	23	9	20
(17)	16	2	13	24	10
(18)	6	17	3	14	25
(19)	21	7	18	4	15
(20)	11	22	8	19	5

区组 (IV)

(21)	1	17	8	24	15
(22)	11	2	18	9	25
(23)	21	12	3	19	10
(24)	6	22	13	4	20
(25)	16	7	23	14	5

区组 (V)

(26)	1	22	18	14	10
(27)	6	2	23	19	15
(28)	11	7	3	24	20
(29)	16	12	8	4	25
(30)	21	17	13	9	5

区组 (VI)

图 7.4 5×5 平衡格子设计
($v=25, k=5, r=16, b=30, \lambda=1$)

第三节 部分平衡格子设计

一、简单格子设计及三重格子设计

如果试验林地不够大,就以平衡格子设计的前两个重复 (I、II) 作为两个重复,如需有 4 次重复则每个重复可以再重复一次,即平衡格子设计的重复 I 作为简单格子设计的重复

I 及 II, 平衡格子设计的重复 II 及 IV, 如此类推。故简单格子设计的重复次数是 2 的倍数。

三重格子设计是采用上述平衡格子设计的前三个重复作为三重格子的 3 次重复。如需要 6 次重复, 则把每个重复再重复一次, 所以重复次数是 3 或者是 3 的倍数, 排列见图 7.5。

1	2	3
4	5	6
7	8	9

重复 I

1	4	7
2	5	8
3	6	9

重复 II

1	5	9
2	6	7
3	8	4

重复 III

图 7.5 三重格子设计

二、矩形格子设计

这种设计可用于处理(或品种) $v = n(n+1)$, 如 $3 \times 4 = 12$, $4 \times 5 = 20$ 的情况设计处理分组以下述方法进行。

(1) 先用拉丁字母排成 $n+1$ 的拉丁方

例如 $n=3$, 则 $n+1=4$ 拉丁方, 即同一字母在行和列上都只出现一次。

(2) 然后把处理顺序排在拉丁方内, 只留拉丁方对角线上的字母不排处理。

(3) 排好后把同一行的处理分为一组, 作为 X 重复。

(4) 再把同一列的处理分为一区作为 Y 重复。

(5) 最后把同一拉丁方字母上的处理分为一组作为 Z 重复。

矩形格子设计可以有 3、6、9、次重复。

如 3×4 矩形格子设计, 见图 7.6。

	A	B ₁	C ₂	D ₃
	D ₄	C	B ₅	A ₆
	B ₇	A ₈	D	C ₉
	C ₁₀	D ₁₁	A ₁₂	B
X ₁	1	2	3	
X ₂	4	5	6	
X ₃	7	8	9	
X ₄	10	11	12	
区组 重复 X				
Y ₁	1	8	11	
Y ₂	2	5	12	
Y ₃	3	6	9	
Y ₄	4	7	10	
区组 重复 Y				
Z ₁	1	5	7	
Z ₂	2	9	10	
Z ₃	3	4	11	
Z ₄	6	8	12	
区组 重复 Z				

图 7.6 3×4 矩形格子设计

X ₁	1	2	3	4
X ₂	5	6	7	8
X ₃	9	10	11	12
X ₄	13	14	15	16
X ₅	17	18	19	20
区组 重复 X				
Y ₁	5	9	13	17
Y ₂	1	10	14	18
Y ₃	2	6	15	19
Y ₄	3	7	11	20
Y ₅	4	8	12	16
区组 重复 Y				
Z ₁	8	11	15	18
Z ₂	2	9	16	20
Z ₃	4	7	14	17
Z ₄	1	5	12	19
Z ₅	3	6	10	13
区组 重复 Z				

图 7.7 4×5 矩形格子设计

X ₁	1	2	3	4	5
X ₂	6	7	8	9	10
X ₃	11	12	13	14	15
X ₄	16	17	18	19	20
X ₅	21	22	23	24	25
X ₆	26	27	28	29	30
区组 重复 X					
Y ₁	6	11	16	21	26
Y ₂	1	12	17	22	27
Y ₃	2	7	18	23	28
Y ₄	3	8	13	24	29
Y ₅	4	9	14	19	30
Y ₆	5	10	15	20	25
区组 重复 Y					
Z ₁	7	13	19	25	27
Z ₂	5	14	16	23	29
Z ₃	1	8	20	21	30
Z ₄	2	9	15	22	26
Z ₅	3	10	11	17	28
Z ₆	4	6	12	18	24
区组 重复 Z					

图 7.8 5×6 矩形格子设计

设置试验时,同样是先在一重复内随机地将区组排好,然后把每一区组内处理随机的安置在小区。

部分平衡不完全区组,统计计算比较麻烦,如果区组间差别较大会影响各处理间比较时的精度。但在树木育种中大量品种,种源以及杂交组合,尤其是要求小区面积大的林间试验乃是很有用的设计。

第八章 裂区设计

第一节 裂区设计

在多因素的许多林业试验中常常会遇到这样一种情形：根据专业知识知道有些因子各水平间容易表现出较大差异，而另一些因子各水平间差异会小些。例如在化学采脂试验中，马尾松割制与刺激剂型的复因子试验，马尾松割制会表现出较大差异。又例如用化学药剂防治某种森林害虫的试验，不同药剂会表现较大差异，而药剂浓度间的差异较小，对于表现较大差异的因素，虽有较大试验误差仍可发现不同水平间的差异，而对表现较小差异的因素，则要求对试验误差更精密的控制，否则将难于鉴别该因素各水平间的差异。当同一试验中出现这样的不同类型的因子时裂区设计是有效的。

裂区设计首先将整个试验区分成几个大区，在每个大区内安排比较容易表现出差异的因素的几种处理，它们常称为主处理，然后在主处理所在各区内引进第二类因素的各个处理，它们称为副处理。

实践中，某些因素需要占用较大范围，而另一些则占较小范围即可，通常需要占大范围的因素恰恰是容易表现出差异的因素。显然裂区设计在这种情形下也是很适用的。

称主处理所在区叫主区或整区，在主区内配置各副处理的小区叫副区或裂区。由于副区范围小，不同副区之间比不同主区之间更为接近，因此对副处理可以使局部控制原则实现

得更好些,从而精确程度会比主区更好。

一、裂区试验的数学模型

裂区设计的统计模型为

$$x_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \nu_k + \epsilon_{ijk} \quad (8.1)$$

$$(i=1, 2, \dots, r, j=1, 2, \dots, p, k=1, 2, \dots, q)$$

x_{ijk} 表示第 i 区组, 第 j 主处理, 第 k 副处理的观察值, μ 总平均, α_i 为第 i 区组(因素 A)的效应值, β_j 为第 j 主处理(因素 B)的效应值, $(\alpha\beta)_{ij}$ 为主区的试验误差, ν_k 为第 k 副处理(因素 C)的效应值, ϵ_{ijk} 为副区的试验误差。并假设

$$\epsilon_{ijk} \sim N(\sigma_e)$$

$$\text{主区试验误差} \sim N(0, \sigma_b)$$

处理效应模型可以是固定模型,也可以是随机模型。

二、裂区设计的统计分析过程

各因素离差平方和计算类似于无重复的三因素方差分析。

下面分别以 $x_{...}$, $x_{..k}$, x_{ijk} 表示

$$\sum_{i=1}^r \sum_{j=1}^p \sum_{k=1}^q x_{ijk} \quad \sum_{i=1}^r \sum_{j=1}^p x_{ij.} \quad \text{及} \quad \sum_{i=1}^r x_{i..}$$

$$\text{校正项 } C = \frac{x_{...}^2}{r \cdot p \cdot q} = \frac{T^2}{rpq}$$

于是有:

$$L_T = \sum_i \sum_j \sum_k x_{ijk}^2 - C$$

主区总离差平方和为

$$L_T = \frac{1}{q} \sum_{i=1}^r \sum_{j=1}^p x_{ij.}^2 - C$$

区组离差平方和

$$L_A = \frac{1}{pq} \sum_{i=1}^r x_{i..}^2 - C$$

主处理离差平方和

$$L_B = \frac{1}{rq} \sum_{j=1}^p x_{.j.}^2 - C$$

主区误差项离差平方和

$$L_{AB} = L_T - L_A - L_B$$

副处理离差平方和

$$L_C = \frac{1}{rp} \sum_{k=1}^q x_{..k}^2 - C$$

主副交互作用离差平方和

$$L_{BC} = \frac{1}{r} \sum_{j=1}^p \sum_{k=1}^q x_{.jk}^2 - C - L_B - L_C$$

副区误差项离差平方和

$$L_e = L_T - L_A - L_B - L_C - L_{BC}$$

它们的自由度分别为

主区项自由度: $rp-1$

区组项(A)自由度: $r-1$

主处理项(B)自由度: $p-1$

主区误差项(AB)自由度: $(r-1)(p-1)$

副处理项(C)自由度: $q-1$

主副交互作用项($B \times C$)自由度: $(p-1)(q-1)$

副区误差项自由度: $p(q-1)(r-1)$

裂区试验分析后,如果主处理、副处理、主副交互作用显著时,则对它们作平均数比较时必须计算四种不同的试验误差,以满足不同平均数的比较:

(1) 主处理平均数比较,例如 $\bar{b}_1 - \bar{b}_2$, 其差数标准差为 $2S_{AB}^2$

(2) 副处理平均数比较, 例如 $\bar{c}_1 - \bar{c}_2$, 其差数标准差为

$$\sqrt{\frac{2S_e^2}{rp}}$$

(3) 同一主处理的两个副处理平均数比较, 例如 $(\bar{b}_1\bar{c}_1 - \bar{b}_1\bar{c}_2)$ 其差数标准差为 $\sqrt{\frac{2S_e^2}{r}}$

(4) 同一副处理或不同副处理的两个主处理平均数的比较, 如 $(\bar{b}_1\bar{c}_1 - \bar{b}_2\bar{c}_1)$ 或 $(\bar{b}_1\bar{c}_2 - \bar{b}_2\bar{c}_1)$, 其差数标准差为

$$2[(q-1)S_e^2 + S_{AB}^2]$$

该试验误差包括两种误差, L_{AB} 与 L_e , 而采用其加权平均。由于这种误差并不相等, 因此, 检验时需考虑两个总体方差不同的情况, t 值计算公式如下

$$t = \frac{(q-1)S_e^2 t_e + S_{AB}^2 t_{AB}}{(q-1)S_e^2 + S_{AB}^2}$$

上式中 t_e 表示副区误差项自由度下的 t 值, t_{AB} 表示主区误差项下的 t 值。

三、裂区试验的统计分析实例

福建省大力推广种植经济作物猕猴桃。1985 年开始, 在主要种植区发生猕猴桃透翅蛾蛀杆害虫的危害。为了筛选出防治该害虫的化学药剂及浓度水平, 考虑药剂差异较易辨别, 又要求试验面积较大, 而药剂浓度水平对试验准确性要求更高, 因此决定将防治试验设计为裂区试验。以药剂为主处理, 药剂浓度水平为副处理。药剂为溴氢菊脂 (B_1)、DDV (B_2)、敌百虫 (B_3), 每种药剂设置三种浓度水平, (C_1 、 C_2 、 C_3), 将主处理和副处理均排列为随机区组, 林间排列布置见图 8.1

A_1	A_2	A_3	A_1	A_2	A_3	A_2	A_3	A_1	A_3	A_1	A_2
C_1	C_2	C_3	C_2	C_2	C_3	C_1	C_1	C_2	C_2	C_2	C_3
C_2	C_1	C_1	C_3	C_1	C_1	C_2	C_1	C_3	C_2	C_1	C_1
C_3	C_1	C_2	C_1	C_1	C_2	C_1	C_1	C_1	C_1	C_1	C_2
I			II			III			IV		

图 8.1

试验结果将观察值列在表 8.1 上

表 8.1

区组 (A)	药剂 (B)	浓度水平 C			x_{ij}	$x_{..}$
		C_1	C_2	C_3		
I	B_1	0.99	0.95	0.75	2.69	7.7
	B_2	0.93	0.75	0.70	2.38	
	B_3	0.98	0.87	0.82	2.67	
II	B_1	0.93	0.93	0.88	2.74	7.6
	B_2	0.90	0.65	0.70	2.25	
	B_3	0.94	0.88	0.80	2.62	
III	B_1	0.94	0.96	0.80	2.73	7.83
	B_2	0.92	0.85	0.65	2.42	
	B_3	0.96	0.87	0.85	2.68	
IV	B_1	0.97	0.93	0.80	2.70	7.85
	B_2	0.93	0.85	0.60	2.33	
	B_3	0.90	0.85	0.80	2.55	
$x_{..k}$		11.32	10.29	9.15	$x_{...} = 30.76$	

为了计算各因素离差平方和,尚需列出 $B \times C$ 的两向表

8.2

表 8.2

主处理 B	副处理 C			x_{ij}	$\bar{b}$
	C_1	C_2	C_3		
B_1	3.86	3.77	3.23	10.86	0.91
B_2	3.68	3.05	2.65	9.38	0.78
B_3	3.78	3.47	3.27	10.52	0.88
$x_{...}$	11.32	10.29	9.15	$x_{...}$ $=30.76$	--
$\bar{C}_i$	0.94	0.86	0.76	—	—

$$\text{校正项 } C = \frac{x_{...}^2}{r p q} = \frac{30.76^2}{4 \times 3 \times 3} = 26.283$$

$$L_T = \sum_i \sum_j \sum_k x_{ijk}^2 - C = (0.99^2 + 0.95^2 + \dots + 0.8^2) - C = 0.372$$

主区总离差平方和

$$L_T = \frac{1}{q} \sum_i \sum_j x_{ij.}^2 - C = \frac{1}{3} (2.69^2 + \dots + 2.55^2) - C = 0.109$$

区组离差平方和

$$L_A = \frac{1}{pq} \sum_{i=1}^r x_{i..}^2 - C = \frac{1}{9} (7.74^2 + \dots + 7.58^2) - C = 0.004$$

主处理平方和

$$L_B = \frac{1}{r q} \sum_{j=1}^p x_{.j.}^2 - C$$

$$= \frac{1}{12}(10.86^2 + 9.38^2 + 10.52^2) - C = 0.10$$

主区误差的离差平方和

$$L_{AB} = L_T - L_A - L_B = 0.005$$

副处理离差平方和

$$\begin{aligned} L_c &= \frac{1}{rp} \sum_{b=1}^q x_{..k}^2 - C \\ &= \frac{1}{12}(11.32^2 + 10.29^2 + 9.15^2) - C = 0.196 \end{aligned}$$

主副处理交互作用($B \times C$)离差平方和

$$\begin{aligned} L_{AB} &= \frac{1}{r} \sum_j \sum_k x_{jk}^2 - C - L_B - L_C \\ &= \frac{1}{4}(3.86^2 + \dots + 3.27^2) - 26.283 - 0.100 \\ &\quad - 0.196 = 0.030 \end{aligned}$$

副区误差项离差平方和 L_e

$$L_e = L_T - T - L_C - L_{BC} = 0.037$$

然后计算各部分均方,列出方差分析表 8.3,注意,在主区中,用主区试验误差(AB)对主处理(B),区组(A)作检验,在副区中用副区误差(e)对副处理(C)和主副交互作用($B \times C$)作检验,从而得以下各 F 值

$$\text{检验区组差异 } F = \frac{L_A/3}{L_{AB}/6} = \frac{0.0013}{0.0008} = 1.625$$

$$\text{检验主处理差异 } F = \frac{L_B/2}{L_{AB}/6} = \frac{0.0500}{0.0008} = 62.5$$

$$\text{检验付处理差异 } F = \frac{L_c/2}{L_e/18} = \frac{0.0980}{0.0021} = 46.67$$

$$\text{检验主副交互作用差异 } F = \frac{L_{BC}/4}{L_e/18} = \frac{0.0075}{0.0021} = 3.57$$

表 8.3

变差来源	自由度	离差平方和	均方	均方比	F 值
区组(A)	3	0.004	0.0013	1.625	$F_{3,6}^{0.05}=4.76$
主处理(B)	2	0.100	0.0500	62.5**	$F_{2,6}^{0.01}=10.9$
主区误差(AB)	6	0.005	0.0008		
主区	11	0.109			
付处理(C)	2	0.196	0.0980	46.67**	$F_{2,18}^{0.01}=6.01$
主付处理	4	0.030	0.0075	3.57*	$F_{4,18}^{0.05}=2.93$
交互(BC)					
付区误差(e)	18	0.037	0.0021		
总	35	0.372			

方差分析结果表明主处理(药剂)副处理(药剂浓度)以及它们的交互作用均达到显著水平。

进而对平均数作检验

1. 各处理平均数间的检验

主处理(主区)平均数差数的标准差为

$$\sqrt{\frac{2S_{AB}^2}{rq}} = \sqrt{\frac{2 \times 0.0008}{4 \times 3}} = 0.01154$$

在主区误差自由度等于 6 情况下,取 $\alpha=0.05$,

$t_{\alpha}(6)=2.447$,故最小显著差数为

$$LSD_{0.05}=0.01154 \times 2.447=0.028$$

类似有

$$LSD_{0.01}=0.01154 \times 3.707=0.043$$

关于主处理平均数的差数列出下表

主处理(药剂)	指标	$\bar{B}_1 - \bar{B}_j$	$B_j B_i$
溴氰菊脂 B_1	0.91		
敌百虫 B_2	0.88	0.03 ⁺	
DIN B_3	0.78	0.13 ⁺ *	0.10

2. 副处理的平均数之差的检验

副处理平均数之间差数的标准差为

$$\sqrt{\frac{2S_e^2}{rp}} = \sqrt{\frac{2 \times 0.0021}{4 \times 3}} = 0.0187$$

$$LSD_{0.05} = t_{0.05}(18) \times 0.0187 = 2.101 \times 0.0187 = 0.039$$

$$LSD_{0.01} = t_{0.01}(18) \times 0.0187 = 2.878 \times 0.0187 = 0.0538$$

易于得出副处理两两平均数之间差异均达到极显著水平。

3. 主、副处理交互作用各水平组合间的平均数差数检验。

①比较同一主处理的两个副处理平均数,如在主处理 B_1 下, C_1 和 C_2 之比较:它们的差数为 $0.965 - 0.943 = 0.022$, 它们的差数标准差为

$$\sqrt{\frac{S_e^2}{r}} = \sqrt{\frac{2 \times 0.0021}{4}} = 0.0324$$

$$LSD_{0.05} = 0.0324 \times t_{0.05}(18) = 0.068$$

$$LSD_{0.01} = 0.0324 \times t_{0.01}(18) = 0.0934$$

故在 B_1 处理下的 C_1, C_2 之间差异不显著。

其它检验不再一一列出。

②比较同一副处理的两个主处理平均数,例如在 C_1 条件下,比较 B_1 和 B_2

$$\text{因为 } \bar{B}_1 - \bar{B}_2 = 0.965 - 0.92 = 0.045$$

差数的标准差为

$$\begin{aligned} & \sqrt{\frac{2[(q-1)S_e^2 + S_{AB}^2]}{rq}} \\ &= \sqrt{\frac{2[2 \times 0.0021 + 0.0008]}{12}} \\ &= 0.0289 \end{aligned}$$

主区误差自由度为 6, $t_{0.05}(6) = 2.447$

副区误差自由度为 18, $t_{0.05}(18) = 2.101$

$$\begin{aligned} t &= \frac{(q-1)S_e^2 t_1 + S_{AB}^2 t_{AB}}{(q-1)S_e^2 + S_{AB}^2} \\ &= \frac{2 \times 0.0021 \times 2.101 + 0.0008 \times 2.447}{2 \times 0.0021 + 0.0008} \\ &= \frac{0.0108}{0.005} = 2.16 \end{aligned}$$

观察之 t 值为 $t = \frac{0.045}{0.0289} = 1.557$

因为 $1.557 < 2.16$ 故 B_1 与 B_2 (在 C_1 条件下) 差异不显著, 其它检验类似不一一列出。

第二节 林木多点多年试验设计

一、林木多年试验方法

林木是多年生的植物, 许多试验往往必须连续数年或更长时间, 才能较明显地得结果。例如林木的种源试验, 尽管众多专家寻找许多早期选择的办法, 而林木的种源试验一般也得 5 至 10 年, 著名专家俞新妥教授对马尾松的种源试验就进行了二十多年。这也是森林的重要特点所决定的。苗木试验一般 1 至 2 年, 但在研究中往往必须探讨其月生长量。因此这

类试验的最大特点是每试验区所接受处理不变,而每试验区分成若干个不同年份采集数据。现主要介绍林木多年试验数据的分析方法。

多年生植物试验数据分析方法,类似于裂区设计试验数据的分析方法,可将各处理看成主区(整区)处理,而把年份看成副区(裂区)处理,因而主区误差(AB)可以用来测定处理对试验结果差异显著性,副区误差(e)可以用来测定年份和处理与年份交互作用对试验结果差异显著性。

例 8.1 1973 年华南热带作物研究院栽培组进行橡胶林中化学除草,测定枯草对橡胶产量影响的试验,6 个处理(包括对照),重复 2 次,连续两年的干胶产量见表 8.4。

分析步骤如下

(1) 计算各项平方和

$$\text{校正项 } C = \frac{x_{...}^2}{rpq} = \frac{(455.09)^2}{6 \times 2 \times 2} = 8629.4545$$

主区总离差平方和

$$\begin{aligned} L_T &= \frac{1}{q} \sum_{i=1}^6 \sum_{j=1}^2 x_{ij}^2 - C \\ &= \frac{1}{2} (41.19^2 + 33.82^2 + \cdots + 30.43^2) - C \\ &= 8724.3296 - 8629.4545 \\ &= 94.8751 \end{aligned}$$

区组离差平方和

$$\begin{aligned} L_A &= \frac{1}{pq} \sum_{k=1}^2 x_{.k}^2 - C \\ &= \frac{1}{12} (240.46^2 + 214.63^2) - C \\ &= 8657.2540 - 8629.4545 \\ &= 27.7995 \end{aligned}$$

表 8.4

单位(克/株次)

处理	测定时间	区组		处理和
		I	II	
B_1 (不喷药)	1973 年 8 月	21.37	16.18	37.55
	1971 年 8 月	19.82	17.61	37.46
	和	41.19	33.82	75.01
B_2	1973 年 8 月	24.37	20.76	45.13
	1974 年 8 月	17.79	20.55	38.34
	和	42.16	41.31	83.47
B_3	1973 年 8 月	22.24	21.09	43.33
	1974 年 8 月	17.09	20.13	37.22
	和	39.33	41.22	80.55
B_4	1973 年 8 月	23.18	16.98	40.16
	1974 年 8 月	17.82	15.38	33.20
	和	41.00	32.36	73.36
B_5	1973 年 8 月	20.53	19.08	39.61
	1974 年 8 月	14.94	16.41	31.35
	和	35.47	35.49	70.96
B_6	1973 年 8 月	25.85	17.65	43.50
	1974 年 8 月	15.46	12.78	28.24
	和	41.31	30.43	71.74
区组总和		240.46	214.63	455.09
年区组总和	1973 年	137.54	111.74	249.28
	1974 年	102.92	102.89	205.81

主处理(施药)离差平方和

$$\begin{aligned}
 L_B &= \frac{1}{rq} \sum_{j=1}^6 x_{.j}^2 - C \\
 &= \frac{1}{4} (75.01^2 + 83.47^2 + 80.55^2 + \cdots + 71.74^2) - C \\
 &= 8661.4206 - 8629.4545
 \end{aligned}$$

$$= 31.9661$$

主区误差的离差平方和

$$\begin{aligned} L_{AB} &= L_T - L_A - L_B \\ &= 94.8751 - 31.9661 - 27.7995 \\ &= 35.1095 \end{aligned}$$

总离差平方和

$$\begin{aligned} L_T &= \sum_{i=1}^2 \sum_{j=1}^6 \sum_{k=1}^2 x_{ijk}^2 - C \\ &= (21.37^2 + 16.18^2 + \cdots + 12.78^2) - C \\ &= 8862.6567 - 8629.4545 \\ &= 233.2022 \end{aligned}$$

副处理(年份)离差平方和

$$\begin{aligned} L_C &= \frac{1}{rp} \sum_{k=1}^2 x_{..k}^2 - C \\ &= \frac{1}{12} (249.28^2 + 205.81^2) - C \\ &= 8708.1895 - 8629.4545 \\ &= 78.7350 \end{aligned}$$

主副处理交互作用($B \times C$)离差平方和

$$\begin{aligned} L_{AB} &= \frac{1}{r} \sum_{j=1}^6 \sum_{k=1}^2 x_{.jk}^2 - C - L_B - L_C \\ &= \frac{1}{2} (37.55^2 + 45.13^2 + \cdots + 28.24^2) \\ &\quad - C - L_B - L_C \\ &= 140.2114 - 78.7350 - 31.9661 \\ &= 29.5103 \end{aligned}$$

副处理与区组的交互作用

$$L_{AC} = \frac{1}{p} \sum_{i=1}^2 \sum_{k=1}^2 x_{ik}^2 - C - L_A - L_C$$

$$\begin{aligned}
&= \frac{1}{6} (137.54^2 + 111.74^2 + 102.92^2 + 102.89^2) \\
&\quad - C - L_A - L_C \\
&= 134.2051 - 78.7350 - 27.7995 \\
&= 27.6706
\end{aligned}$$

副区误差离差平方和

$$\begin{aligned}
L_e &= L_T - L_{T'} - L_{AC} - L_{BC} - L_C \\
&= 233.2022 - 94.8751 - 78.7350 - 29.5103 \\
&\quad - 27.6705 \\
&= 2.4113
\end{aligned}$$

(2) 列出方差分析表 8.5

表 8.5

变差来源	自由度	离差平方和	均方	均方比	F 值
区组 A	1	27.7995	27.7995	3.96	$F_{0.05} = 4.84$ $F_{0.01} = 9.65$
主处理 B	5	31.9661	6.3932	<1	
主区误差 AB	5	35.1095	7.0219		
主区	11	94.8751			
副处理 (年份) C	1	78.7350	78.7350	163.25	$F_{0.05} = 6.61$ $F_{0.01} = 16.26$
主副处理交互作用 BC	5	29.5103	5.9021	12.24	$F_{0.05} = 5.05$ $F_{0.01} = 10.97$
区组与副处理交互作用 AC	1	27.6705	27.6705	57.37	$F_{0.05} = 6.61$ $F_{0.01} = 16.26$
副区误差 e	5	2.4113	0.4823		
总	23	233.2022			

从表 8.5 可以看出:

1、从三年合并产量来看,处理间差异不显著。

I、年份间差异极显著,故需对各年份间产量继续作 q 检验。由于本例仅二年产量,其年份间显著性完全由 F 检验所决定。查表 8.4 知 1974 年产量总值为 205.81,1973 年产量总值为 249.28,由此可知 1974 年比 1973 年减产是极显著的。

II、年份 $\times$ 处理极显著,说明不同年份,处理的反应有所不同,对此很值得继续用 q 检验深入分析。

IV、年份 $\times$ 区组极显著,说明不同年份各区组的处理反应有所不同,但这不是我们了解的主要目的,故可不必继续作 q 检验。

(3) 年份 $\times$ 处理的 q 检验

$$\begin{aligned}
 D_{.1.}(a=6) &= q_{.1.}(a=6, f=5) \cdot \sqrt{\frac{\text{误差(副区)均方}}{4}} \\
 &= 0.03 \times \sqrt{\frac{0.4823}{4}} \\
 &= 6.03 \times 0.35 \\
 &= 2.11 \\
 D_{.01.}(a=6) &= q_{.01.}(a=6, f=5) \times 0.35 \\
 &= 8.91 \times 0.35 \\
 &= 3.12
 \end{aligned}$$

表 8.6

处理	78 年产量	减 74 年产量	差数 $\div 4$			
B_1	15.26	3.79	2.29	2.12	2.08	1.75
B_2	8.26	2.04	0.51	0.37	0.33	
B_3	6.96	1.72	0.21	0.04		
B_4	6.79	1.68	0.17			
B_5	6.11	1.51				
B_6	0.09					

结论：处理 B_0 与 B_1 、 B_2 、 B_3 相比，1974 年的产量显著或极显著下降，其余处理间 1974 年产量下降的差异不显著。

二、多点多年试验分析方法

林木引种、栽培措施、种源试验以及森林的病虫害防治试验常要求试验点布置在多个地方，以考察试验结果的稳定性，提高试验的可靠性。又由于林木的许多试验是长周期的，这样就形成了一种多点、多年的试验设计种类。

例如 某林木的种源试验，参试种源为 m 种，分别在 n 个试验点进行试验，观察年份为 t 年，其观察的数据可以依三因素的方差分析进行。该方法在每二章方差分析中已经叙述过，这里不再重复。

在变差来源项目和自由度计算时可参考表 8.7

表 8.7

变差来源	自由度
种源	$m-1$
试验点	$n-1$
年份	$t-1$
种源×试验点	$(m-1) \times (n-1)$
种源×年份	$(m-1) \times (t-1)$
试验点×年份	$(n-1) \times (t-1)$
种源×试验点×年份(误差)	$(m-1)(n-1)(t-1)$
总变差	$mnt-1$

第九章 协方差分析

试验设计的主要目的是要控制试验误差,提高试验准确性。要做到这一点就必须控制试验条件的均匀性,使各处理变量尽可能处在一致的试验条件下。在林业的试验设计中,一般说来,处理多数是可以人为控制的,但也常遇到一些处理变量很难或不可能人为控制的情形。例如杉木优树无性系嫁接试验研究,希望各砧木的接口粗细完全一样是不可能,但砧木接口粗细对试验影响很大。又如病虫害防治试验,要求各小区发病率或危害程度完全一致很不容易。又如林木的地理种源试验,由于调种有前后,所以它们不能同一天播种,这样一年后观察其苗高时,各种种源生长天数不一样。对于这些处理不能人为控制的情况,其试验结果的数据就不能应用方差分析方法来分析、检验各处理之间的差异,而应当应用协方差分析方法。

协方差分析又称为相关变量分析或互变量分析。它是将回归分析与方差分析相结合而产生的一种统计分析方法。

在协方差分析中,通常象杉木优树无性嫁接试验研究,嫁接成活率称为反应变量 Y ,而与 Y 有线性回归关系的另一变量(砧木接口宽度)称为伴随变量或协变量 X 。协方差分析就是通过反应变量与协变量之间的线性关系,调整观察的反应变量。如果不做这种调整,由于协变量的存在会使误差平方和增大,其结果不能真实地检验出反应变量在不同处理之间所存在的真正差异。因此,可以说协方差分析是一种调整无法控

制又影响反应变量的一种方差分析方法,而这种调整是通过回归分析来实现的,这叫做统计控制。统计控制作为试验控制的一种辅助手段,通过它减少试验误差,提高试验准确性,使结果更加可靠。

第一节 协方差分析方法

一、协方差分析方法

在协方差分析中最简单的一种,就是具有一个协变量的单因素试验设计的协方差分析方法。

单因素协方差分析的统计模型为:

$$y_{ij} = \mu + \alpha_i + \beta(x_{ij} - \bar{x}_{..}) + \epsilon_{ij} \quad \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, n \end{cases} \quad (9.1)$$

式中 y_{ij} 表示第 i 次处理所得反应变量的第 j 次观察值。 x_{ij} 表示第 i 次处理,第 j 次观察值 y_{ij} 的相应的协变量值, $\bar{x}_{..}$ 是 x_{ij} 的平均数, μ 是总平均数, α_i 表示第 i 次处理效应值, β 是 y_{ij} 在 x_{ij} 上的线性回归系数, ϵ_{ij} 是随机误差变量。

使用协方差分析,需要满足以下几个条件:

(1) 随机误差变量 ϵ_{ij} 必须是遵从正态分布 $N(0, \sigma)$ 的独立随机变量。

(2) y_{ij} 与 x_{ij} 之间存在线性关系、同时协变量 x_{ij} 不受处理效应 α_i 的影响。

(3) 处理效应之和等于零即 $\sum_{i=1}^a \alpha_i = 0$ 。

从协方差分析的统计模型,可以看出它是方差分析模型和回归分析线性模型的结合。 α_i 是单因素方差分析中处理效

应值, β 是回归分析中回归系数。如今 $\mu' = \mu - \bar{\beta}x_{..}$, 那么协方差分析的统计模型改写为:

$$y_{ij} = \mu' + \alpha_i + \beta x_{ij} + \varepsilon_{ij} \quad \begin{cases} i=1, 2, \dots, a \\ j=1, 2, \dots, n \end{cases} \quad (9.2)$$

假定进行品种比较试验, 包括 a 个品种, 每品种重复 n 次, 在试验过程中同时记录每次试验的产量以及每小区的植株数目, 以 y 代表产量, 以 x 代表株数。那么 x 是协变量, y 是反应变量, 因为我们的目标是研究不同品种的产量是否有显著差异, 但株数不同产量是不同的。

协方差分析需要计算总离差平方和、品种间(处理的)、品种内(误差的)离差平方和以及交叉乘积和

总的:

$$\begin{cases} LS_y = \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2 = \sum_i \sum_j y_{ij}^2 - \frac{y_{..}^2}{a \cdot n} \\ LS_x = \sum_i \sum_j (x_{ij} - \bar{x}_{..})^2 = \sum_i \sum_j x_{ij}^2 - \frac{x_{..}^2}{a \cdot n} \\ LS_{xy} = \sum_i \sum_j (x_{ij} - \bar{x}_{..})(y_{ij} - \bar{y}_{..}) \\ \quad = \sum_i \sum_j x_{ij}y_{ij} - \frac{(x_{..})(y_{..})}{a \cdot n} \end{cases} \quad (9.3)$$

品种间(处理的):

$$\begin{cases} LT_{yy} = \sum_{i=1}^a \sum_{j=1}^n (\bar{y}_{i.} - \bar{y}_{..})^2 = \sum_{i=1}^a \frac{y_{i.}^2}{a} - \frac{y_{..}^2}{a \cdot n} \\ LT_{xx} = \sum_i \sum_j (\bar{x}_{i.} - \bar{x}_{..})^2 = \sum_{i=1}^a \frac{x_{i.}^2}{a} - \frac{x_{..}^2}{a \cdot n} \\ LT_{xy} = \sum_i \sum_j (\bar{x}_{i.} - \bar{x}_{..})(\bar{y}_{i.} - \bar{y}_{..}) \\ = \sum_{i=1}^a \frac{(x_{i.})(y_{i.})}{n} - \frac{(x_{..})(y_{..})}{a^2 \cdot n} \end{cases} \quad (9.4)$$

品种内(误差的)

$$\begin{cases} LE_{yy} = \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2 = LS_{yy} - LT_{yy} \\ LE_{xx} = \sum_i \sum_j (x_{ij} - \bar{x}_{i.})^2 = LS_{xx} - LT_{xx} \\ LE_{xy} = \sum_i \sum_j (x_{ij} - \bar{x}_{i.})(y_{ij} - \bar{y}_{i.}) = LS_{xy} - LT_{xy} \end{cases} \quad (9.5)$$

从 x, y 两个变量的离差平方和, 乘积和的分析中可以计算总的、处理的误差的各部分回归系数。

总的回归系数。

$$b_e = \frac{\sum_i \sum_j (x_{ij} - \bar{x}_{i.})(y_{ij} - \bar{y}_{i.})}{\sum_i \sum_j (\bar{x}_{i.} - \bar{x}_{..})^2} = \frac{LS_{xy}}{LS_{xx}} \quad (9.6)$$

处理的(品种间的)回归系数

$$b_T = \frac{\sum_i \sum_j (\bar{x}_{i.} - \bar{x}_{..})(\bar{y}_{i.} - \bar{y}_{..})}{\sum_i \sum_j (\bar{x}_{i.} - \bar{x}_{..})^2} = \frac{LT_{xy}}{LT_{xx}} \quad (9.7)$$

误差的(品种内的)回归系数

$$b_e = \frac{\sum_i \sum_j (x_{ij} - \bar{x}_{i.})(y_{ij} - \bar{y}_{i.})}{\sum_i \sum_j (x_{ij} - \bar{x}_{i.})^2} = \frac{LE_{xy}}{LE_{xx}} \quad (9.8)$$

协方差分析的基本点是通过协变量 x 调整反应变量 y 。从统计模型(9.1)知道, μ 的估计量为 $\bar{y}_{..}$, β 的估计量为 b^* ,

$$b^* = \frac{LE_{xy}}{LE_{xx}} \quad (9.9)$$

处理(品种的)效应 α_i 的估计量为

$$\bar{y}_{i.} - \bar{y}_{..} - b^*(\bar{x}_{i.} - \bar{x}_{..}) \quad (9.10)$$

在这个模型中,误差平方和 L_e

$$L_e = LE_{yy} - b^* LE_{xy} \quad (9.11)$$

该项自由度为 $a(n-1)-1$, 它的均方为

$$S_e^2 = \frac{L_e}{a(n-1)-1} \quad (9.12)$$

如果试验不存在处理效应,则原统计模型将成为

$$y_{ij} = \mu + \beta(x_{ij} - \bar{x}_{i.}) + \epsilon_{ij}$$

这是个普通的一元回归问题,这时 μ 和 β 估计量分别为 $\bar{y}_{..}$ 和

$b = \frac{LS_{xy}}{LS_{xx}}$, 其误差平方和为

$$LS_e = LS_{yy} - \frac{LS_{xy}^2}{LS_{xx}} \quad (9.13)$$

具有 $an-2$ 个自由度。

如果试验本身存在处理效应,但若不按存在处理效应对待,那么此时计算出来的误差平方和 LS_e 会大于按存在处理效应计算所得误差平方和 LS_e , 而它们之差正是由处理效应 α_i 所产生的平方和,且具有 $a-1$ 自由度,这个差值是 $LS_e - LS_e$ 。于是可以用 F 检验来检验是否存在处理效应的假设。

由于协变量的存在,它们的处理和误差项平方和都需经过调整:

$$\text{处理平方和 } LS_e - LS_e = (LS_{yy} - \frac{LS_{xy}^2}{LS_{xx}}) - (LE_{yy} - \frac{LE_{xy}^2}{LE_{xx}})$$

误差平方和 $LS_e = LE_{yy} - \frac{LE_{xy}^2}{LE_{xx}}$

如果协变量不存在,即

$$LS_{xy} = LS_{yx} = LE_{xy} = LE_{yx} = 0$$

则

处理平方和成为 $LS_{yy} - LE_{yy} = LT_{yy}$

误差平方和成为 $LS_e = LE_{yy}$

在协方差分析中,总认为存在协变量,于是必须通过 y 在 x 上的回归调整 LS_{yy} 和 LE_{yy} 。调整后的总平方和为

$$LS_e = LS_{yy} - \frac{LS_{xy}^2}{LS_{xx}}$$

调整后的误差平方和为

$$LS_e = LE_{yy} - \frac{LE_{xy}^2}{LE_{xx}}$$

由于在调整平方和时,用了另一个参数 β ,所以调整后误差平方和的自由度不是 $a(n-1)$ 而是 $a(n-1)-1$ 。

由于 $\bar{y}_i$ 包括处理效应和在协变量上的回归效应,因此对平均数也要作相应调整。

调整前,根据统计模型 9.1 必须检验是否满足上述协方差分析的前提

(1) 检验各处理的剩余方差是否齐性

可由原始数据计算出各处理剩余平方和,确定自由度,用 Bartlett 检验法检验。

(2) 检验回归直线是否平行($\beta_i = \beta$)

因为模型中的 β 是综合反应 y 对 x 的线性关系的,对各处理应该是公共的回归系数,它是由 b^* 估计的,如果各处理加归系数与 b^* 差异显著那么协方差分析将失去理论基础。

为此应先计算各处理的剩余平方和 LS^w (关于各处理求

和), 和用公共回归系数 b^* 计算得到的误差平方和 LS_e , 进而计算回归系数平方和

$$LS_{\text{回归系数}} = LS_e - LS_e^w \quad (9.14)$$

相应地 $f_{\text{回归系数}} = f_e - f_e^w$

于是

$$S_{\text{回归系数}}^2 = \frac{LS_{\text{回归系数}}}{f_{\text{回归系数}}} \quad (9.15)$$

然后用 S_e^2 对 $S_{\text{回归系数}}^2$ 做检验, 两者差异不显著时, 表明各处理形成的回归线有公共斜率, 当然也可以作图而做出直观判断。

(3) 检验回归是否显著

归属于回归的平方和为

$$LS_R = \frac{LE_{xy}^2}{LE_{xx}} \quad (9.16)$$

剩余平方和为

$$LS_e = LE_{yy} - \frac{LE_{xy}^2}{LE_{xx}} \quad (9.17)$$

建立统计量 F

$$F = \frac{LS_R/1}{LS_e/[a(n-1)-1]} \quad (9.18)$$

$F > F_\alpha$ 时回归显著, 显著水平 α ; 反之不显著。

在满足上述条件的情况下

$$\text{调整后的 } \bar{y}_i = \bar{y}_{i(\text{调整})} - b^* (\bar{x}_i - \bar{x}_{..}) \quad (9.19)$$

其中

$$b^* = \frac{LE_{xy}}{LE_{xx}}$$

调整后处理平均数的标准误差为

$$S_{\text{调整的}\bar{y}_i} = \sqrt{S_e^2 \left[\frac{1}{n} + \frac{(\bar{x}_i - \bar{x}_{..})^2}{LE_{xx}} \right]} \quad (9.20)$$

二、协方差分析的计算过程

例 9.1 为了比较 3 种施肥方案对某经济树种产量的影响,测得每株树的产量 y (斤)和胸径 x (cm)数据如表 9.1,试问 3 种施肥方案对产量是否有显著影响?

由于树木的不同胸径将说明试验林木的大小,而不同大小的林木产量是不同的,对产量自然会造成影响,现在 3 种施肥方案所用林木胸径不同,为了推断施肥的效应,必须排除胸径的效应,否则得到的结论将是不正确的。因此我们采用协方差分析。

表 9.1

A_1	x	15	13	11	12	12	16	14	17	$\bar{x}_1 = 13.750$
	y	85	83	65	76	80	91	84	90	$\bar{y}_1 = 81.750$
A_2	x	17	16	18	18	21	22	19	18	$\bar{x}_2 = 18.625$
	y	97	90	100	95	103	106	99	94	$\bar{y}_2 = 98.000$
A_3	x	22	24	20	23	25	27	30	32	$\bar{x}_3 = 25.375$
	y	89	91	83	95	100	102	105	110	$\bar{y}_3 = 96.875$

首先检验 $\varepsilon_{ij} \sim N(0, \sigma^2)$, $\beta_i = \beta$, 以及 $\beta \neq 0$ 是否可以得到满足。

(1) 分别计算用 3 种施肥方案(A_1, A_2, A_3)的产量与胸径的回归方程

$$\hat{y}_1 = 33.516 + 3.508x_1$$

$$\hat{y}_2 = 54.570 + 2.232x_2$$

$$\hat{y}_3 = 143.141 + 2.118x_3$$

将 3 组数据合并,即合并各项离差平方和和交叉乘积和,计算公共的回归系数 b^*

$$b^* = 2.400$$

用公共回归系数代替 $b_i (i=1, 2, 3)$, 这 3 条直线在 y 上截距值, 由 $\bar{y}_i - b^* \bar{x}_i$ 确定, 这样可以得到 3 条平行回归直线

$$\hat{y}_1 = 48.750 + 2.400x_1$$

$$\hat{y}_2 = 53.300 + 2.400x_2$$

$$\hat{y}_3 = 35.975 + 2.400x_3$$

(2) 用每一种处理各自的回归系数 b_i , 计算剩余平方和

$$\text{处理 1 剩余平方和 } LS_{e1} = 99.86 \quad f_1 = 6$$

$$\text{处理 2 剩余平方和 } LS_{e2} = 32.420 \quad f_2 = 6$$

$$\text{处理 3 剩余平方和 } LS_{e3} = 47.171 \quad f_3 = 6$$

则组 3 剩余平方和 LS_e^w 为

$$LS_e^w = LS_{e1} + LS_{e2} + LS_{e3} = 179.457$$

相应自由度为 $18 = a(n-2)$

再用公共回归系数 b^* 计算 3 条平行直线的剩余平方和,

$$\text{因为 } LS_e = LE_{xy} - \frac{LE_{xy}^2}{LE_{xx}} = 227.615$$

相应自由度为 $a(n-1)-1=20$

(3) 检验方差是否齐性

$$S_{e1}^{2w} = \frac{99.866}{6} = 16.644$$

$$S_{e2}^{2w} = \frac{32.420}{6} = 5.403$$

$$S_{e3}^{2w} = \frac{47.171}{6} = 7.862$$

采用二个差异最大的均方做检验,

$$F = \frac{S_{e1}^{2w}}{S_{e2}^{2w}} = \frac{16.644}{5.403} = 3.08$$

$$F_{0.05}(6, 6) = 4.82 \quad F < F_{0.05} \quad \text{认为方差齐性。}$$

(4) 检验回归线是否平行

$$LS_{\text{回归系数}} = LS_r - LS_r^* = 227.615 - 179.457 = 48.158$$

$$f_{\text{回归系数}} = 20 - 18 = 2$$

$$S_{\text{回归系数}}^2 = \frac{48.158}{2} = 24.079$$

$$F = \frac{\frac{48.158}{2}}{\frac{179.457}{18}} = 2.415 < F_{0.05}(2, 18) = 3.55$$

因此三条回归线可认为平行。

(5) 检验回归是否显著

$$LS_R = \frac{LE_{xy}^2}{LE_{xx}} = 1010.76 \quad f_R = 1$$

$$LS_r = 227.615 \quad f_r = 20$$

$$F = \frac{S_R^2}{S_r^2} = \frac{1010.76/1}{227.615/20} = 88.8$$

$$\text{而 } F_{0.01}(1, 20) = 8.1 \quad F > F_{0.01}$$

回归关系显著存在

(6) 检验处理效应是否显著 $H_0: \alpha_i = 0$

处理效应平方和为

$$\begin{aligned} LS_r - LS &= (LS_{xy} - \frac{LS_{xy}^2}{LS_{xx}}) - (LE_{yy} - \frac{LE_{xy}^2}{LE_{xx}}) \\ &= (2555.958 - \frac{1080.750^2}{720.500}) - (1258.375 \\ &\quad - \frac{420.875^2}{175.250}) \\ &= 934.833 - 227.615 = 707.218 \end{aligned}$$

$$F = \frac{(LS_e - LS_f) / (a - 1)}{LS_e / [a(n - 1) - 1]} = \frac{\frac{707.218}{2}}{\frac{227.615}{20}} = 31.07$$

$$F_{0.01}(2, 20) = 5.72$$

$F > F_{0.01}$ 认为不同施肥方案对产量影响显著。

实践中,常常直接认为上条件满足,或只是根据 b_i 判断回归线平行后即进入下一阶段

(7) 计算出调整的平均数

$$\text{调整后的 } \bar{y}_i = \text{调整前 } \bar{y}_i - b_i(x_i - x_{..})$$

我们有

$$\text{调整后 } \bar{y}_1 = 81.750 - 2.4(13.750 - 19.25) = 94.950$$

$$\text{调整后 } \bar{y}_2 = 98.000 - 2.4(18.625 - 19.25) = 99.500$$

$$\text{调整后 } \bar{y}_3 = 96.875 - 2.4(25.375 - 19.25) = 82.175$$

$$\text{调整前产量平均数 } \bar{y}_2 > \bar{y}_1 > \bar{y}_3$$

调整后产量平均数 $\bar{y}_2 > \bar{y}_1 > \bar{y}_3$, 如不调整将会得出不正确的结论, 这是因为施肥方案 3 的试验林木胸径较方案 1 的试验林木胸径大, 从而造成假象, 排除了协变量, 即林木胸径影响后 A_3 不如 A_1 , 这个结论是真实的。

第二节 随机区组设计的协方差分析

当试验设计为随机区组试验时, 协方差分析的统计模型为

$$y_{ij} = \mu + \alpha_i + \nu_j + \beta(x_{ij} - \bar{x}_{..}) + \varepsilon_{ij} \quad \left. \begin{array}{l} i = 1, \dots, a \\ j = 1, \dots, m \end{array} \right\} \quad (9.21)$$

i 表示第 i 处理, j 表示第 j 区组, 若令 $\mu' = \mu - \beta\bar{x}_{..}$, 上述模型可改写为

$$y_{ij} = \mu' + \alpha_i + v_j + \beta x_{ij} + \varepsilon_{ij} \quad (9.22)$$

其中 v_j 指区组(重复)效应, $\sum v_j = 0$

随机区组设计的协方差分析其原理和方法与前大致相同, 下面结合实例说明具体方法

例 9.2 某一农场作橡胶树肥料试验, 采用正交表 $L_9(3^4)$ 安排试验, 试验为随机区组设计, 区组数 $m=4$, 处理数为 $a=9$, 各小区处理前平均基围为 x , 处理后平均基围为 y , 如表 9.2。

表 9.2

处理	区组 I		区组 II		区组 III		区组 IV		总 数	
	x	y	x	y	x	y	x	y	x	y
1	30.3	32.9	30.5	31.9	28.2	31.3	27.4	30.7	116.4	126.5
2	32.8	35.4	30.7	33.1	29.1	31.8	24.0	27.4	116.6	127.7
3	31.7	34.7	25.6	28.4	25.5	28.1	23.8	26.4	106.4	117.6
4	26.7	29.7	31.6	34.1	27.1	30.1	25.4	28.3	110.8	122.2
5	32.9	35.9	33.0	36.1	26.1	28.7	25.7	28.8	117.7	129.5
6	30.0	31.9	32.3	35.1	28.1	30.9	28.4	31.7	118.8	129.6
7	34.3	37.5	31.8	35.0	27.1	30.8	28.1	31.8	121.3	135.1
8	31.5	34.8	30.9	33.1	29.5	31.7	27.8	31.2	119.7	130.8
9	29.6	32.3	30.1	33.2	26.2	29.4	24.0	26.9	109.9	121.8
Σ	279.8	305.1	276.5	299.7	246.9	272.8	234.6	263.2	1037.8	1140.8

首先计算各部分(总的、处理的、区组的、误差的)的离差平方和

总的:

$$\begin{aligned}
LS_{yy} &= \sum_{i=1}^9 \sum_{j=1}^4 (y_{ij} - \bar{y}_{..})^2 = \sum_{i=1}^9 \sum_{j=1}^4 y_{ij}^2 - \frac{y_{..}^2}{am} \\
&= 32.9^2 + 35.4^2 + \cdots + 26.9^2 - \frac{(1140.8)^2}{36} \\
&= 276.34 \\
LS_{xx} &= \sum_{i=1}^9 \sum_{j=1}^4 (x_{ij} - \bar{x}_{..})^2 = \sum_{i=1}^9 \sum_{j=1}^4 x_{ij}^2 - \frac{x_{..}^2}{am} \\
&= 30.3^2 + 32.8^2 + \cdots + 31.7^2 + \cdots + 24.0^2 \\
&\quad - \frac{(1037.8)^2}{36} \\
&= 285.05 \\
LS_{xy} &= \sum_{i=1}^9 \sum_{j=1}^4 (x_{ij} - \bar{x}_{..})(y_{ij} - \bar{y}_{..}) \\
&= \sum_{i=1}^9 \sum_{j=1}^4 x_{ij}y_{ij} - \frac{(x_{..})(y_{..})}{am} \\
&= 30.3 \times 32.9 + \cdots + 24.0 \times 26.9 \\
&\quad - \frac{1037.8 \times 1140.8}{36} = 276.26
\end{aligned}$$

处理间的

$$\begin{aligned}
LT_{yy} &= \sum_{i=1}^4 \sum_{j=1}^9 (\bar{y}_{i.} - \bar{y}_{..})^2 = \sum_{i=1}^4 \frac{y_{i.}^2}{m} - \frac{y_{..}^2}{am} \\
&= \frac{1}{4} (126.5^2 + 127.7^2 + \cdots + 121.8^2) - 36150.68 \\
&= 57.93 \\
LT_{xx} &= \sum_{i=1}^4 \sum_{j=1}^9 (\bar{x}_{i.} - \bar{x}_{..})^2 = \sum_{i=1}^4 \frac{x_{i.}^2}{m} - \frac{x_{..}^2}{am} \\
&= \frac{1}{4} (116.4^2 + 116.6^2 + \cdots + 109.9^2) - 29917.47 \\
&= 50.34 \\
LT_{xy} &= \sum_{i=1}^4 \sum_{j=1}^9 (\bar{x}_{i.} - \bar{x}_{..})(\bar{y}_{i.} - \bar{y}_{..})
\end{aligned}$$

$$\begin{aligned}
&= \sum_i \sum_j \frac{(x_{ij})(y_{ij})}{m} - \frac{(x_{..})(y_{..})}{am} \\
&= \frac{1}{4} (116.4 \times 126.5 + 116.6 \times 127.7 + \dots \\
&\quad + 109.9 \times 121.8) - 32886.73 \\
&= 53.07
\end{aligned}$$

区组的(区组的离差平方和用 LQ 表示)

$$\begin{aligned}
LQ_{yy} &= \sum_i \sum_j (\bar{y}_{.j} - \bar{y}_{..})^2 = \sum_{j=1}^4 \frac{y_{.j}^2}{9} - \frac{y_{..}^2}{9 \times 4} \\
&= \frac{1}{9} (305.1^2 + \dots + 263.2^2) - 36150.68 = 138.23
\end{aligned}$$

$$\begin{aligned}
LQ_{xx} &= \sum_i \sum_j (\bar{x}_{.j} - \bar{x}_{..})^2 = \sum_{j=1}^4 \frac{x_{.j}^2}{9} - \frac{x_{..}^2}{9 \times 4} \\
&= \frac{1}{9} (279.8^2 + \dots + 234.6^2) - 29917.47 = 164.43
\end{aligned}$$

$$\begin{aligned}
LQ_{xy} &= \sum_i \sum_j (\bar{x}_{.j} - \bar{x}_{..})(\bar{y}_{.j} - \bar{y}_{..}) \\
&= \sum_i \sum_j \frac{(x_{.j})(y_{.j})}{9} - \frac{(x_{..})(y_{..})}{9 \times 4} \\
&= \frac{1}{9} (279.8 \times 305.1 + \dots \\
&\quad + 234.6 \times 263.2) - 32886.73 \\
&= 150.50
\end{aligned}$$

由 $S = T + Q + E$

$$LE_{yy} = LS_{yy} - LT_{yy} - LQ_{yy} = 80.18$$

$$LE_{xx} = LS_{xx} - LT_{xx} - LQ_{xx} = 70.28$$

$$LE_{xy} = LS_{xy} - LT_{xy} - LQ_{xy} = 72.69$$

以上各项的离差平方和计算结果汇总列出在表 9.3 中

表 9.3

变差来源	自由度	xx 项	xy 项	yy 项
总变异(S)	35	285.05	276.26	276.34
区组(Q)	3	164.43	150.50	138.23
处理(T)	8	50.34	53.07	57.93
误差(E)	24	70.28	72.69	80.18
处理+误差(V)	32	120.62	125.76	138.11

下面检验回归关系的显著性。检验应从 y 的总变异中除去区组和处理的变异,即以 y 的误差项变异作为测定 y 对于 x 的回归关系的总变异。

$$L_{\text{总}} = LE_{yy} \quad f_{\text{总}} = 24$$

$$L_{\text{回归}} = \frac{LE_{xy}^2}{LE_{xx}} \quad f_{\text{回归}} = 1$$

$$L_{\text{剩余}} = LE_{yy} - \frac{LE_{xy}^2}{LE_{xx}} \quad f_{\text{剩余}} = 23$$

$$\text{这样, } L = \frac{L_{\text{回归}}/1}{L_{\text{剩余}}/23} = \frac{75.18/1}{[80.18 - \frac{72.69^2}{70.28}]/23} = 341.73$$

而 $F_{0.01}(1, 23) = 7.88$, 因 $F > F_{0.01}$, y 对于 x 的回归是极显著的, 换言之, 值得用回归来矫正每一处理的平均基围, 如果回归不显著则不必作协方差分析, 直接方差分析即可。

除去回归关系后的误差平方和为

$$L_e = LE_{yy} - \frac{LE_{xy}^2}{LE_{xx}} = 5.0 \quad f_e = 23$$

$$L_{e'} = \frac{LV_{xy}^2}{LV_{xx}} = 138.11 - \frac{125.76^2}{120.62} = 6.99$$

$$L_{e'} - L_e = 1.99 \quad \text{自由度} = 31 - 23 = 8$$

根据以上计算可列出除去回归关系后的处理的方差分析见表 9.4

表 9.4

变异来源	自由度	平方和	均方	均方比	$F_{0.05}$
处理	8	1.99	0.25	1.44	2.38
误差	23	5.00	0.22		
处理+误差	31	6.99	—		

从上表中看出检验结果,除去回归关系后各处理差异不显著,分析到此为止。

如果检验结果除去回归关系后处理间差异显著,这时矫正每处理的平均基围并作 t 检验

$$b^* = \frac{LE_{xy}}{LE_{xx}}$$

$$b^* = \frac{72.69}{70.28} = 1.03$$

$$\text{调整后 } \bar{y}_i = \text{调整前 } \bar{y}_i - b^* (\bar{x}_i - \bar{x}_{..})$$

例如处理 1

$$\text{调整后 } \bar{y}_i = \text{调整前 } \bar{y}_i - b^* (\bar{x}_i - \bar{x}_{..})$$

$$= 31.6 - 1.03 \left(\frac{116.4}{4} - 28.8 \right) = 31.3$$

第三节 多个协变量的协方差分析

上面介绍的是单个协变量的协方差分析方法。在林业科学研究中,多个协变量的协方差分析亦常能遇见。为此,以一个实例说明多个协变量协方差分析的基本方法,限于篇幅,不在此作很详细的说明,但其原理与单个协变量的协方差分析原理完全一致,只不过把协方差分析中的回归系数改为偏回归系数,回归平方和以及回归剩余平方和相应地改用偏回归平方和与偏回归剩余平方和,如果是两个协变量的话,它们的计算公式如下:

误差项的:

$$b_{yx_1 \cdot x_2} = \frac{Lx_1y \cdot Lx_2x_2 - Lx_2y \cdot Lx_1x_2}{Lx_1x_1 \cdot Lx_2x_2 - L^2x_1x_2}$$

$$b_{yx_2 \cdot x_1} = \frac{Lx_2y \cdot Lx_1x_1 - Lx_1y \cdot Lx_1x_2}{Lx_1x_1 \cdot Lx_2x_2 - L^2x_1x_2}$$

$$\text{偏回归平方和} = b_{yx_1 \cdot x_2} \cdot Lx_1y + b_{yx_2 \cdot x_1} \cdot Lx_2y$$

$$\text{偏回归剩余平方和} = Lyy - b_{yx_1 \cdot x_2} \cdot Lx_1y - b_{yx_2 \cdot x_1} \cdot Lx_2y$$

例 9.3 为了探讨不同年份某林木害虫的自然死亡率有无显著差异,从实践中知道空气温度(x_1)和空气湿度(x_2)对害虫的自然死亡率有影响。因此,采用二个协变量的协方差分析之,具体调查数据见下表 9.5。

表 9.5

年份	变量	重 复						年份总和
		1	2	3	4	5	6	
1980	x_1	25.6	25.4	30.8	33.0	28.5	28.0	171.3
	x_2	14.9	13.3	4.6	14.7	12.8	7.5	67.8
	y	19.0	22.2	35.3	32.8	25.3	35.8	170.4
1981	x_1	25.4	28.3	35.3	32.4	25.9	24.2	171.5
	x_2	7.2	9.5	6.8	9.7	9.2	7.5	49.9
	y	32.4	32.2	43.7	35.7	28.3	35.2	207.5
1982	x_1	27.9	34.4	32.5	27.5	23.7	32.9	178.9
	x_2	18.6	22.2	10.0	17.6	14.4	7.9	90.7
	y	26.2	34.7	40.0	29.6	20.6	47.2	198.3
重 复 总 和	x_1	78.9	88.1	98.6	92.9	78.1	85.1	521.7
	x_2	40.7	45.0	21.4	42.0	36.4	22.9	208.4
	y	77.6	89.1	119.0	98.1	74.2	118.2	576.2

先计算各个平方和和乘积和,其结果列于表 9.6。

然后计算偏回归系数和偏回归平方和、偏回归剩余平方和;

表 9.6

变异原因	自 由 度	平 方 和			乘 积 和		
		Lx_1x_1	Lx_2x_2	L_{yy}	Lx_1y	Lx_2y	Lx_1x_2
年 份	2	106.34	171.46	629.22	109.83	-257.03	-47.06
重 复	5	6.26	139.41	124.42	8.41	-22.26	26.24
误 差	10	117.93	74.20	288.66	142.01	-21.46	20.17
总 和	17	230.53	385.07	982.30	341.25	-300.75	-0.65
年份+误差	12	224.27	245.66	857.88	332.84	-278.49	-26.89

$$b_{yx_1 \cdot x_2} = \frac{142.01 \times 74.20 - (-21.46) \times 20.17}{117.93 \times 74.20 - 20.17^2}$$

$$= \frac{10969.9902}{8343.5771} = 1.3148$$

$$b_{yx_2 \cdot x_1} = \frac{(-21.46) \times 17.93 - 142.01 \times 20.17}{117.93 \times 74.20 - 20.17^2}$$

$$= \frac{-5395.1195}{8343.5771} = -0.6466$$

$$\text{偏回归平方和} = 1.3148 \times 142.01 + (-0.6466) \times (-21.46) = 200.5907$$

$$\text{偏回归剩余平方和} = 228.66 - 200.5907 = 28.069$$

检验回归是否显著,列出方差分析表 9.7。从表 9.7 中看出偏回归显著,说明必须用偏回归的方法来校正死亡率 y ,以除去加 x_1 和 x_2 不同所受的影响,再进行 y 的校正方差分析。

表 9.7

变差来源	自由度	平方和	均方	均方比	$F_{0.05}$	$F_{0.01}$
偏回归	2	56.292	28.146	8.022	4.71	9.55
剩余	8	28.096	3.5086			
总的(误差)	10	228.66				

进行 y 的校正方差分析时,先求出 y 校正误差平方和及校正的“处理+误差”的平方和,然后由校正的“处理+误差”平方和减去误差平方和,便得到校正的处理间平方和

y 的校正误差平方和

$$= L_{yy} - b_{yx_1 \cdot x_2} L_{x_1 y} - b_{yx_2 \cdot x_1} L_{x_2 y} = 28.06$$

y 的校正“处理+误差”平方和

$$= L_{yy} - \frac{L^2_{x_1 y}}{L_{x_1 x_1}} - \frac{L^2_{x_2 y}}{L_{x_2 x_2}}$$

$$= 857.88 - \frac{332.84^2}{224.27} - \frac{(-278.49)^2}{245.66} = 48.20$$

上式中 L_{yy} 、 $L_{x_1x_1}$ 、 $L_{x_2x_2}$ 、 L_{x_1y} 、 L_{x_2y} 均匀“年份+误差”的各项平离差平方和

y 的校正处理(年份)间平方和

$=y$ 的校正“处理+误差”平方和 $-y$ 的校正误差平方和

$$=48.20-28.07=20.13$$

校正后的方差分析见表 9.8

表 9.8

变差来源	自由度	平方和	均方	均方比	$F_{0.05}$
年份(处理)	2	20.13	10.07	2.87	4.46
试验误差	8	28.07	3.51		
年份+误差	10*	48.20			

* 年份+误差原来自由度为 12, 因回归系数有两个, 所以其自由度为 10。

从表中看出, 校正后的死亡率无显著差异。

对 y_{ij} ($i=1, 2, 3, j=1, 2, \dots, 6$) 的校正值记为 y'

$$y' = \bar{y} + byx_1 \cdot x_2(x_1 - \bar{x}_1) + byx_2 \cdot x_1(x_2 - \bar{x}_2)$$

例如, 第一协变量 x_1 的 1980 年第一个重复的观察值为 25.6, 第二个协变量 x_2 的 1980 年第一个重复的观察值为 14.9。

$$\bar{x}_1 = 28.98 \quad \bar{x}_2 = 11.58 \quad \bar{y} = 32.01$$

则相应的校正值为

$$\begin{aligned} y' &= \bar{y} + byx_1 \cdot x_2(x_1 - \bar{x}_1) + byx_2 \cdot x_1(x_2 - \bar{x}_2) \\ &= 32.01 + 1.3148(25.6 - 28.98) + (-0.6466) \\ &\quad (14.9 - 11.58) \\ &= 25.42 \end{aligned}$$

其它校正值可依此类推。还可以得到不在样本资料内其它观测死亡率的校正值, 只要已知 x_1 与 x_2 的值。

第十章 回归正交设计

林业科学研究中,对多因素试验一般都采用正交试验设计。在这方面取得许多成果。例如对毛竹施用N(氮)、P(磷)、K(钾)三要素的肥料,采用正交设计,对施肥处理后调查新竹产量,分析结果得到N、P、K合理配比。但人们还希望利用已有试验数据,能给出整个区域上因素(N、P、K)与指标(新竹产量)之间关系,明确其函数表达式,即建立数学模型:

$$W = f(N, P, K)$$

式中W为指标。以便应用数学模型来控制施肥,达到指导生产的目的。

要建立数学模型,人们也愈来愈要求以较少试验建立较高精度的回归方程。这样就得摆脱古典回归分析的被动局面,主动地把试验的安排、数据的处理和回归方程的精确度统一起来加以考虑。根据试验目的和数据分析的要求来选择试验点,不仅使得在每个试验点上获得的数据具有最大信息,从而减少试验次数,而且使数据的统计分析具有一些较好的性质。这就是将“回归分析法”与“正交试验法”有机结合起来,用较少的试验而得到精度较高回归方程的试验设计方法,称为“回归正交设计”。

在毛竹施肥、营林措施最优化的问题中,对不同的施肥方案的物理方面、化学方面、生物工程方面、生理生化方面不是完全了解它们过程的机制的,此时用回归分析这种方法来解决施肥方案的最优问题,是一种有效办法。这种方法类似于控

制论中的“黑箱”理论。系统的输入和输出都是知道的,但系统内部结构不清楚。在施肥过程中,毛竹生长过程是“黑箱”,输入的是N、P、K。输出的就是新竹产量。尽管我们不知道它的内部结构,但是通过回归正交设计方法可以得到:

$$W=f(N,P,K)$$

的函数表达式,并由此得到实现最优化的措施。因此回归正交设计成为回归分析中一个重要方面。

回归正交设计产生于50年代,发展到今天内容已相当丰富。按照回归模型的次数分为一次回归设计、二次回归设计。本章主要介绍一次和二次的回归设计。

第一节 一次回归正交设计

一次回归正交设计主要是应用二水平正交表。例如二水平正交表 $L_4(2^3)$ 、 $L_8(2^7)$ 、 $L_{12}(2^{11})$ 、 $L_{16}(2^{15})$ 、 $L_{64}(2^{63})$ 等。上述毛竹施肥试验中有三个因素N、P、K。选用正交表 $L_8(2^7)$,并把正交表中的“1”和“2”二个水平改为+1与-1(或改为-1和+1),然后把三个因素分别放在第1、2、4列上。这时正交表中的“-1”和“+1”不仅表示因素的状态,而且还表示变量的取值。若三个因素N、P、K还存在着交互作用,这些交互效应在回归分析中可用变量的非线性项表示。因为二水平的因素,其交互作用只占一列,所以交互效应也只占改造后二水平正交表的一列,这一列可以从交互效应表中查得。交互作用列加入试验计划并不影响正交性。究竟用哪一张二水平正交表呢?这要根据因子的个数以及交互作用所占列数来确定,依是否容纳得下因素及交互作用(最后留有空白列),而且试验次数

较少原则来选择二水平正交表。毛竹施肥试验就可选用 $L_8(2^7)$ 表,例如用 x_1, x_2, x_3 表示三个因素 N、P、K,并把它们分别放在正交表 $L_8(2^7)$ 的 1、2、4 列上就得到如表 10.1 的试验计划表。

这里没有把 $L_8(2^7)$ 表的其它各列都取来是因为这些

列对安排三因素试验而言,不起作用。我们得到了 8 个试验点,它们在三维空间中正好是一个正方体的 8 个顶点。若以 x_{ij} 表示第 i 次试验中第 j 变量(因素)的改造后的取值,则有

(1)任一列之和 $\sum_i x_{ij} = 0$

(2)任二列内积 $\sum_i x_{ik} x_{ij} = 0$

因此这种试验设计是具有正交性的。表 10.2 列出经过上述代换后所得的几张常用二水平正交表,由于代换前后的两种正交表没有本质差别,故还用 $L_4(2^3), L_8(2^7)$ 等符号表示。

表 10.1

试验号	x_1	x_2	x_3
1	1	1	1
2	1	1	-1
3	1	-1	1
4	1	-1	-1
5	-1	1	1
6	-1	1	-1
7	-1	-1	1
8	-1	-1	-1

表 10.2 $L_4(2^3)$

试验号	x_1	x_2	x_3
1	1	1	1
2	1	-1	-1
3	-1	1	-1
4	-1	-1	1

试验号	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$
1	1	1	1	1	1	1	1
2	1	1	-1	1	-1	-1	-1
3	1	-1	1	-1	1	-1	-1
4	1	-1	-1	-1	-1	1	1
5	-1	1	1	-1	-1	1	1
6	-1	1	-1	-1	1	-1	1
7	-1	-1	1	1	-1	-1	1
8	-1	-1	-1	1	1	1	-1

这种表在指定 x_1, x_2, x_3 所在列位置之后,正好可以用 $x_i x_j$ 表示 x_i 与 x_j 二因素的交互作用列, $x_i x_j x_k$ 表示三因素的高阶交互作用,这一点可以立即由正交性加以验证(与交互作用表对照)。

下面我们来说明回归正交设计、分析的过程。

一、确定因子的变化范围

例如 有 i 个因子 $Z_1, Z_2, \dots, Z_i$ 与某项指标 y , 设计中首先要确定每个因子 Z_k 的变化范围。设 Z_{1k} 和 Z_{2k} 分别表示因子 Z_k 变化的下界和上界。假如试验在水平 Z_{1k} 和 Z_{2k} 上进行,那么分别称 Z_{1k} 和 Z_{2k} 为 Z_k 的下水平和上水平,称它们的算术平均为 Z_{0k}

$$Z_{0k} = \frac{Z_{1k} + Z_{2k}}{2}$$

它们差的一半为其因素 Z_k 的变区间 Δ_k

$$\Delta_k = \frac{Z_{2k} - Z_{1k}}{2}$$

二、对 Z_k 的水平进行编码

所谓编码,就是对因子的取值作如下的线性变换:

$$x_k = \frac{Z_k - Z_{0k}}{\Delta_k}$$

这样建立了因素 Z_k 和 x_k 取值变换关系即:

下水平 $Z_{1k} \leftrightarrow -1$

0 水平 $Z_{0k} \leftrightarrow 0$

上水平 $Z_{2k} \leftrightarrow +1$

从变换关系式中可以知道当因子 Z_k 在 $[Z_{1k}, Z_{2k}]$ 内变化时,它的编码 x_k 值就在区间 $[-1, 1]$ 内变化。例如有二个因素时,因素的研究区域在平面上是一个矩形,通过变换后研究区域就变成了一个正方形。

因子 $Z_k (k=1, \dots, p)$ 进行编码之后,原先要求算 y 对 $Z_1, Z_2, \dots, Z_p$ 的回归方程,现在变成求算 y 对 $x_1, x_2, \dots, x_p$ 的回归方程。不管在一次回归设计中还是在二次回归设计中,都应当先对因素 $Z_k (k=1, 2, \dots, p)$ 进行编码。

三、回归系数的计算与统计检验

根据改造后正交表进行试验,如果试验次数为 N 次,可以得到下列数据 $(x_{11}, x_{12}, x_{13}, \dots, x_{1p}, y_1), (x_{21}, x_{22}, x_{23}, x_{2p}, y_2), \dots, (x_{i1}, x_{i2}, x_{i3}, \dots, x_{ip}, y_i), (x_{N1}, x_{N2}, x_{N3}, \dots, x_{Np}, y_N)$ 。

由于一次回归的数学模型为

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \dots + \beta_p x_{ip} + \epsilon_i \quad (10.1)$$

式中 i 表示第 i 次试验 $\epsilon_i \sim N(0, \sigma_\epsilon)$, 其中 $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ 是待估计系数,

假定通过 N 次试验所得试验数据计算出来的回归方程为

$$\hat{y}_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i3} + \dots + b_p x_{ip} \quad (10.2)$$

其中 $b_0, b_1, b_2, \dots, b_p$ 是参数 $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ 的最小二乘估计

如果令

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{N1} & x_{N2} & \cdots & x_{Np} \end{bmatrix} \quad (10.3)$$

X 称为结构矩阵。

由正交性

$$\begin{aligned} A = X'X &= \begin{bmatrix} N & \sum_i x_{i1}^2 & 0 \\ 0 & \sum_i x_{i2}^2 & \cdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \sum_i x_{ip}^2 \end{bmatrix} \\ &= \begin{bmatrix} N & 0 \\ 0 & N \\ \vdots & \vdots \\ 0 & \cdots & N \end{bmatrix} \end{aligned} \quad (10.4)$$

这里 A 矩阵称为信息矩阵。

A 的逆阵恰为相关矩阵 C

$$C = A^{-1} = \begin{bmatrix} \frac{1}{N} & 0 \\ 0 & \frac{1}{N} \\ \vdots & \vdots \\ 0 & \cdots & \frac{1}{N} \end{bmatrix} \quad (10.5)$$

由多元线性回归分析方法中可以知道

$$B=XY=\begin{bmatrix} \sum_i Y_i \\ \sum_i x_{i1}y_i \\ \vdots \\ \sum_i x_{ip} \cdot y_i \end{bmatrix}=\begin{bmatrix} B_0 \\ B_1 \\ \vdots \\ B_p \end{bmatrix} \quad (10.6)$$

而 $b=A^{-1}$ 即

$$\begin{cases} b_0 = \frac{B_0}{N} = \frac{1}{N} \sum_i y_i \\ b_j = \frac{B_j}{N} = \frac{1}{N} \sum_i x_{ij} y_i \end{cases} \quad j=1,2,\dots,p \quad (10.7)$$

从上面可以看到,正是由于正交设计计划的正交性,因而消除了回归系数间相关性,所以一次回归正交设计的计算与显著性检验同正交多项式回归完全类似。

从上面推导中可知,各变量的偏回归平方和 U_j

$$U_j = b_j B_j = b_j \sum_i x_{ij} \cdot y_i = N b_j^2 \quad (j=1,2,\dots,p) \quad (10.8)$$

U_j 与 b_j 的平方成正比, b_j 的绝对值越大, U_j 也越大,这就说明在正交设计所求得的回归方程中,每一个回归系数 b_j 的绝对值大小,刻画了对变量 x_j 在过程中的作用。

回归平方和记为 U

$$U = \sum_{j=1}^p U_j$$

总平方和记为 $L_{\text{总}}$

$$L_{\text{总}} = \sum_i y_i^2 - \frac{(\sum_i y_i)^2}{N} = \sum_i y_i^2 - \frac{B_0^2}{N} \quad (10.9)$$

剩余平方和记为 $L_{\text{剩余}}$

$$L_{\text{剩余}} = L_{\text{总}} - U = L_{\text{总}} - \sum_{j=1}^p U_j \quad (10.10)$$

那么对于因子 x_1 的回归显著性检验, 需要求出 F 值

$$F_1 = \frac{U_1/1}{L_{\text{剩}}/(N-p-1)} \quad (10.11)$$

如果 $F > F_{0.05}(1, N-p-1)$, 则认为因素 x_1 对试验指标 y 的影响达到显著水平, 反之则不显著。对其它变量的回归显著性检验也是类似的。

还是由于正交性, 对整个回归(y 对全部 P 个变量 $x_1 \cdots x_p$) 的显著性检验, 需求 F 值

$$F = \frac{(U_1 + U_2 + \cdots + U_p)/p}{L_{\text{剩}}/(N-p-1)} \quad (10.12)$$

如果 $F > F_{0.05}(p, N-p-1)$, 认为整个回归关系显著, 显著水平达到 $\alpha=0.05$, 否则认为回归不显著。

对回归显著性检验的具体计算过程, 可参看表 10.3 和表 10.4

表 10.3

试验号	x_0	x_1	x_2	x_3	...	x_p	y
1	1	x_{11}	x_{12}	x_{13}	...	x_{1p}	y_1
2	1	x_{21}	x_{22}	x_{23}	...	x_{2p}	y_2
3	1	x_{31}	x_{32}	x_{33}	...	x_{3p}	y_3
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
N	1	x_{N1}	x_{N2}	x_{N3}	...	x_{Np}	y_N

记 B_i 为 $\sum y_i, \sum x_{i1}y_i, \sum x_{i2}y_i, \dots, \sum x_{ip}y_i$
 ($i=0, 1, 2, \dots, p$) 则 $b_i = \frac{B_i}{N}$ 可与上面这一行对应地列为

$$\frac{B_0}{N} \quad \frac{B_1}{N} \quad \frac{\hat{B}_2}{N} \quad \cdots \quad \frac{B_p}{N} \quad (j=0,1,2,\cdots p)$$

$U_j = b_j B_j$ 即

$$U_0 \quad U_1 \quad U_2 \quad \cdots \quad U_p$$

若计算出 $\sum y_i^2$ 则 $L_{\text{总}} = \sum y_i^2 - \frac{B_0^2}{N}$ $L_{\text{剩}} = L_{\text{总}} - (U_1 U_2 + \cdots + U_p)$

表 10.4

变差来源	离差平方和	自由度	均方	F 值
x_1	$U_1 = B_1^2/N$	1	U_1	$U_1/S_{\text{剩}}^2$
x_2	$U_2 = B_2^2/N$	1	U_2	$U_2/S_{\text{剩}}^2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_p	$U_p = B_p^2/N$	1	U_p	$U_p/S_{\text{剩}}^2$
回归	$L_{\text{回}} = u_1 + \cdots + u_p$	p	$\sum_{i=1}^p U_i/p$	$\frac{1}{p} \sum u_i/S_{\text{剩}}^2$
剩余	$L_{\text{剩}} = L_{\text{总}} - \sum_{i=1}^p U_i$		$S_{\text{剩}}^2 = L_{\text{剩}}/(N-p-1)$	
总和	$L_{\text{总}} = \sum y_i^2 - \frac{B_0^2}{N}$	$N-1$		

四、零水平的重复试验

上面所叙述的回归方程的检验,只是说明相对于平均剩余平方和来说,变量(因素)的影响显著与否,如果检验结果回归方程显著,也只是说明在被研究的区域内部,回归方程预测的理论与实测值拟合较好。但对于研究区域的内部回归方程的预测理论值与实测值是否也拟合得较好还不知道。因此有必要在零水平($Z_{01}, Z_{02}, \cdots, Z_{0p}$)处再安排一些重复试验,如果重复数为 r ,其试验结果分别为 $y_{01}, y_{02}, \cdots, y_{0r}$ 。设零水平的

实际试验结果的平均数为 $\bar{y}_0$,它与回归方程中常数项 $b_0 = \frac{1}{N} \sum y_i$,是否有显著差异,可作 t 检验。

$$t = \frac{|b_0 - \bar{y}_0| \sqrt{f_{\text{剩余}} + f_0}}{\sqrt{L_{\text{剩余}} + L_0} \cdot \sqrt{\frac{1}{N} + \frac{1}{r}}} \quad (10.13)$$

当按 10.13 式计算的 t 值 $< t_{\alpha}(f_{\text{剩余}} + f_0)$ 时,认为 b_0 与 $\bar{y}_0$ 无显著差异,即在区域中心,一次回归方程与实测值还是拟合较好的,否则认为用一次回归方程来描述 y 对 $x_1 x_2 \cdots x_p$ 的关系还不够确切,可以考虑建立高次的回归方程。

式 10.13 中 $f_{\text{剩余}} = N - p - 1$, $f_0 = r - 1$, r 为重复数,而 $L_0 = \sum_{i=1}^r (y_{0i} - \bar{y}_0)^2$ 。

五、交互作用的处理

交互作用是指多个因素的联合作用,当引入交互作用项以后,原来的回归不再是线性的了。交互效应项可直接完成 x_i 与 x_j 的乘积,并占据 x_i 与 x_j 交互作用列所在位置,因此回归系数 b_{ij} 的计算完全相似于线性项,即有

$$b_{ij} = B_{ij} / N \quad (10.14)$$

$$\text{其中 } B_{ij} = \sum_k x_{ik} x_{jk} y_k \quad i, j = 1, 2, \cdots, p, i \neq j \quad (10.15)$$

检验它们的显著性用偏回归平方和

$$U_{ij} = b_{ij} B_{ij} \quad i, j = 1, 2, \cdots, p, i \neq j \quad (10.16)$$

这正好是充分体现了正交性的优越性。

试验中增加了交互作用项后,减少了正交表中的空白列,或者没有空白列存在,如果空白列不存在则剩余项(误差项)就没有自由度了,此时试验必须安排重复试验。同时在试验设计中要注意混杂现象的发生。一定要弄清楚哪些交互效应是

可以忽略的,哪些是不可忽略的,在可以忽略交互作用而原来的交互作用所在的列上配置其它变量可以减少试验计划所需试验次数,多放一个因子相当于做全因子试验的预定试验次数的一半,统计上称为 $1/2$ 实施,类似地多放二个因子,相当于 $1/4$ 实施…等,实践中人们常常采用这类部分实施法,用较少的试验次数,求取比较满意的试验结果。

下面举例说明一次回归设计具体的统计分析过程。

例 10.1 近年来我国南方毛竹的经营水平不断提高,采用各种复合肥施肥技术,来提高毛竹的生产力。为了得到合理配方,进行下列试验,配方施肥就是 N 、 P 、 K 三要素按不同比例配成肥料。三个因素水平的编码表如表 10.5

表 10.5

因子	$N:Z_1(kg)$	$P:Z_2(kg)$	$K:Z_3(kg)$
下水平(-1)	10	10	5
上水平(1)	20	30	15
变动区间 Δ_i	5	10	5
零区间(0)	15	20	15

试验采用正交表 $L_8(2^7)$ 。每一试验小区为 1 亩。试验控制为区组,试验指标为第二年的新竹产量(吨/亩)。试验结果及计算过程见表 10.6。试验根据专业知识判断 N 和 P 可能存在交互作用,因此试验只安排 N 、 P 、 K 和 N 与 P 交互项。从表 10.6 中可以看到 $X_1(N)$ 和 $X_3(K)$,以及 X_1X_2 项有显著影响。

表 10.6

试验号	X_0	X_1	X_2	X_3	X_{12}	空白	空白	空白	新竹产量 (吨/亩)
1	1	1	1	1	1	1	1	1	1.51
2	1	1	1	-1	1	-1	-1	-1	1.49
3	1	1	-1	1	-1	1	-1	1	1.45
4	1	1	-1	-1	-1	-1	1	1	1.61
5	1	-1	1	1	-1	-1	1	-1	0.85
6	1	-1	1	-1	-1	1	-1	1	0.96
7	1	-1	-1	1	1	-1	-1	1	0.78
8	1	-1	-1	-1	1	1	1	-1	0.81
$B_j =$ $\sum x_j y_i$	9.46	2.66	0.16	-0.28	-0.28	0	0.1	0.2	$L_{总} = 0.0098$ $f_{总} = 3$
$b_j = \frac{B_j}{N}$	1.1825	0.3325	0.0200	-0.035	-0.035	0	0.0125	0.0325	$F = \frac{L_{总}/4}{L_{剩余}/3}$
$U_j = b_j B_j$		0.8845	0.0032	0.0098	0.0098	0	0.0013	0.0085	$= \frac{0.9073/4}{0.0098/3}$
$F = \frac{U_j}{L_j/3}$		270.77	0.9796	3.0000	3.0000				$= 69.44$
显著性		$\alpha = 0.01$		$\alpha = 0.25$	$\alpha = 0.25$				

对所得的回归方程进行 F 检验。

$$F = \frac{(U_1 + U_2 + U_3 + U_4)/4}{L_{误差}/3} = 69.44$$

$F > F_{0.01}(4, 3)$, 因此认为回归方程高度显著, 即用一次回归模型来描述过程, 拟合情况是好的。又由于 x_2 对 y 影响不显著, 因此将 x_2 剔除, 于是得到的回归方程是

$$\hat{y} = 1.1825 + 0.3325x_1 - 0.035x_3 - 0.035x_1x_3$$

其中 $x_1 = \frac{z_1 - 15}{5}$, $x_3 = \frac{z_3 - 15}{5}$

第二节 二次回归正交设计

如果用一次回归正交设计法描述某一个具体过程,用零水平的重复试验发现 t 值研究值达到显著的水平,说明该过程用一次回归方程来拟合因子与试验指标关系不太合适,需要用二次或高次回归方程来描述。这里我们将介绍二次回归正交设计。

如果有 p 个变量,二次回归方程的一般形式为:

$$\hat{y} = b_0 + \sum_{j=1}^p b_j x_j + \sum_{i < j} b_{ij} x_i x_j + \sum_{j=1}^p b_{jj} x_j^2 \quad (10.17)$$

共有回归系数 $q = 1 + p + c_p^2 + p = 1 + 2p + \frac{p(p-1)}{2} = c_{p+1}^2$ 这说明,要得到 p 个变量的二次回归方程,试验次数 N 应不小于 q 。但事实上,为了计算二次回归的系数,每个变量所取的水平数应大于等于 3,因而所要做的试验次数往往比较多。

一、二次回归正交设计的原理

为了获得二次回归方程,但又必须以较少试验次数为条件。人们研究了这个问题以后提出一种“组合设计”来解决,即在因子空间中选择几类有不同特征的点,把它们适当组合起来而形成试验计划。如果 $p=2$,组合设计中试验点的分布具体如下:

在二变量 x_1, x_2 情况下,组合设计由 $N=9$ 组成,具体见图 10.1

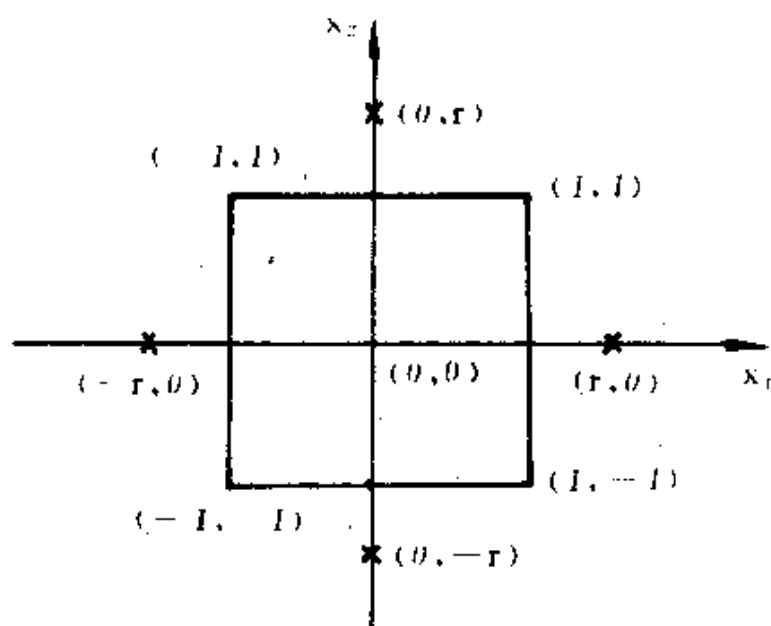
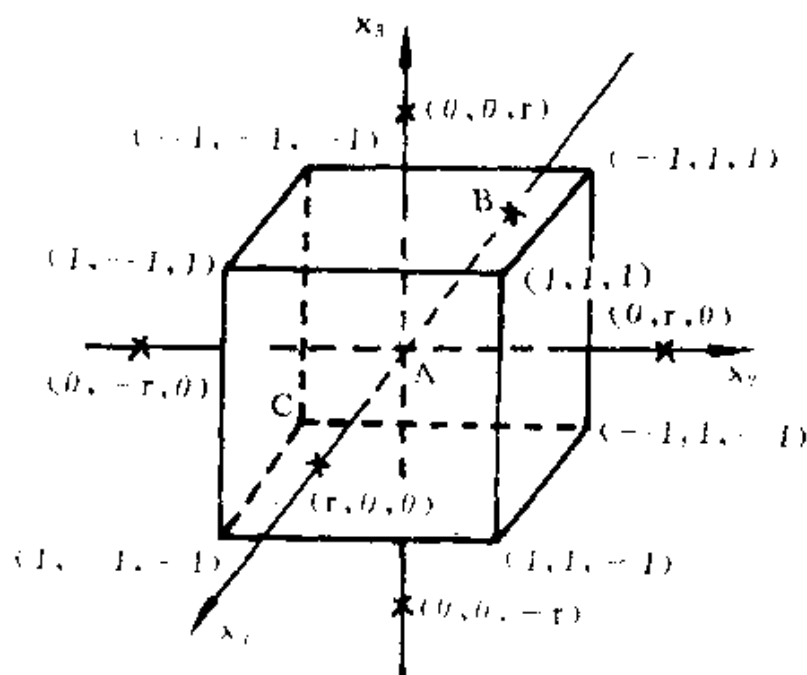


图 10.1

	x_1	x_2	
1	(1	1)	这四个试验点组成二水平(+1,和-1)的全因子试验 2^2
2	(1	-1)	
3	(-1	1)	
4	(-1	-1)	
5	(r	0)	这四个试验点分布在 x_1 和 x_2 轴上星号位置。
6	(-r	0)	
7	(0	r)	
8	(0	-r)	
9	(0	0)	由 x_1 和 x_2 的零水平组成的中心试验点。

在三变量 x_1, x_2, x_3 的情况下,组合设计由 $N=15$ 个点组成,具体见图 10.2



注：图中 A、B、C 三个星号点的星号位置是
 $A = (0, 0, 0)$ $B = (-r, 0, 0)$ $C = (-1, -1, 1)$

图 10.2

这些试验点可以分成三类,描述见表 10.7

表 10.7

	I	II	III
试验次数	8	6	1
试验点	$(1, 1, 1)$ $(-1, 1, 1)$ $(r, 0, 0)$ $(0, 0, r)$ $(0, 0, 0)$		
分布	$(1, 1, -1)$ $(-1, 1, -1)$ $(-r, 0, 0)$ $(0, 0, -r)$		
	$(1, -1, 1)$ $(-1, -1, 1)$ $(0, r, 0)$		
	$(1, -1, -1)$ $(-1, -1, -1)$ $(0, -r, 0)$		

试验点特征：由 x_1, x_2, x_3 的二水平 $(+1)$ 和 (-1) 分布在 x_1, x_2, x_3 的零水平组成的全因子试验 2^3 轴上的星号位置的中心试验点

根据 $p=2, p=3$ 时试验点的分布,可以归纳出,一般情

况下, p 个变量的组合设计由下列 N 个点组成。

如果以 m_c 表示二水平(+1 和 -1)的全因子试验点的个数 2^p 或它的部分实施试验点个数 $2^{p-1}, 2^{p-2}, \dots$ 等。

$2p$ 为分布在 p 个坐标轴上的星号点, 它们与中心点的距离 r 称为星号臂, r 是待定参数。根据一定的要求(如正交性)调节 r , 就可以得到各种具有很好性质的设计(如正交设计)。

m_c 表示为在各变量都取零水平的中心点的重复试验次数, 它可以用一次或多次。

以 N 表示试验次数, 有

$$N = m_c + 2p + m \quad (10.18)$$

用组合设计安排试验计划有一系列优点, 首先它的试验点比全因子试验要少得多, 但仍有足够的剩余自由度。例如 $p=4$ 是参数个数 $q=15$ 全因子试验次数为 $3^4=81$, 那么剩余自由度为 66, 而组合设计的试验次数为 25, 试验次数会减少将近 70%, 而且组合设计的剩余自由度为 10, 这一般也能满足试验的要求。

其次, 二次回归正交设计是在一次回归基础上获得的, 这对试验者来说是很方便的。因为如果一次回归不显著, 那么只要在一次回归试验的基础上, 再在星号点和中心补充做一些试验, 就可求得二次回归方程。

为了使组合设计成为正交设计(即满足正交性), 还要确定适当的星号臂 r 。以 $p=3$ 情况为例来说明如何确定一个适当的 r 。

在三个变量 x_1, x_2, x_3 情况下, 二次回归设计的结构矩阵见表 10.8

表 10.8

试验号	x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	x_1^2	x_2^2	x_3^2
1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	-1	1	-1	-1	1	1	1
3	1	1	-1	1	-1	1	-1	1	1	1
4	1	1	-1	-1	-1	-1	1	1	1	1
5	1	-1	1	1	-1	-1	1	1	1	1
6	1	-1	1	-1	-1	1	-1	1	1	1
7	1	-1	-1	1	1	-1	-1	1	1	1
8	1	-1	-1	-1	1	1	1	1	1	1
9	1	r	0	0	0	0	0	r^2	0	0
10	1	$-r$	0	0	0	0	0	r^2	0	0
11	1	0	r	0	0	0	0	0	r^2	0
12	1	0	$-r$	0	0	0	0	0	r^2	0
13	1	0	0	r	0	0	0	0	0	r^2
14	1	0	0	$-r$	0	0	0	0	0	r^2
15	1	0	0	0	0	0	0	0	0	0

从表 10.8 可以看出,对一次回归设计,在试验计划中加入星号点(试验点)后,并没有破坏一次变量列和交互效应列的正交性,表的正交性是被 x_0 和 $x_j^2 (j=1,2,3)$ 所破坏的,因此我们可以计算得到:

$$\sum_{i=1}^N x_{ij}^2 = m_c + 2r^2 \neq 0 \quad (10.19)$$

$$\sum_{i=1}^N x_{i0}x_{ij}^2 = m_c + 2r^2 \neq 0 \quad (10.20)$$

$$\sum_{i=1}^N x_{ij}^2 \cdot x_{ik}^2 = m_c \neq 0 (j, k=1, 2 \cdots p) j \neq k \quad (10.21)$$

为了使组合设计具有正交性,必须在使得相关矩阵 $C = (X'X)^{-1}$ 为对角阵的条件下定出 r 值。为说明问题的方便,将结构矩阵 X 中的列重新加以排列,把平方列 $x_j^2 (j=1,2,3)$ 放

在 x_0 列之后, x_j 列之前, 并且令

$$e = m_c + 2r^2$$

$$f = m_c + 2r^4$$

于是有

$$X'X = \begin{bmatrix} N & e & e & e & & & \\ e & f & m_c & m_c & & & \\ e & m_c & f & m_c & & & \\ e & m_c & m_c & f & & & \\ & & & & e & & \\ & & & & & e & \\ & & & & & & e \\ & & & & & & & m_c \\ & & & & & & & & m_c \\ & & & & & & & & & m_c \\ & & & & & & & & & & m_c \end{bmatrix} \quad (10.22)$$

其中空白片皆为 0, 这是 $p=3$ 的情况, 类似地, 对任意 p 个变量可求得:

$$X'X = \begin{bmatrix} N & e & e & \cdots & e \\ e & f & m_c & \cdots & m_c \\ e & m_c & f & \cdots & m_c \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ e & m_c & m_c & \cdots & f \\ & & & & e & & \\ & & & & & e & \\ & & & & & & e \\ & & & & & & & m_c \\ & & & & & & & & m_c \\ & & & & & & & & & m_c \\ & & & & & & & & & & m_c \end{bmatrix} \quad (10.23)$$

它的逆矩阵 $C = (X'X)^{-1}$ 也具有同样形式, 其元素用大写字母表示。

$$(X'X)^{-1} = \begin{bmatrix} K & E & E & \cdots & E \\ E & F & G & \cdots & G \\ E & G & F & \cdots & G \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ E & G & G & \cdots & F \\ & & & e^{-1} & \\ & & & e^{-1} & \\ & & & \cdots & e^{-1} \\ & & & & m_c^{-1} \\ & & & & m_c^{-1} \\ & & & & \cdots & m_c^{-1} \end{bmatrix} \quad (10.24)$$

其中

$$\left. \begin{aligned} H &= 2r^2[Nf + (p-1)Nm_c - Pe^2] \\ K &= 2r^2H^{-1}[f + (p-1)]m_c \\ F &= H^{-1}[Nf + (p-2)Nm_c - (p-1)e^2] \\ E &= -2H^{-1}er^4 \\ G &= H^{-1}(e^2 - Nm_c) \end{aligned} \right\} \quad (10.25)$$

从矩阵 10.21 可以看出,要使组合设计是正交的,首先要令 $G=0$,其次 $E=0$

要令 $G=0$,即有 $e^2 - Nm_c = 0$

又因为 $e = m_c + 2r^2$, $N = m_c + 2p + m_0$,将 e 和 N 代入上式有

$$\begin{aligned} (m_c + 2r^2) - (m_c + 2p + m_0)m_0 &= 0 \\ r^4 + m_cr^2 - \frac{m_c}{2}(p + \frac{m_0}{2}) &= 0 \end{aligned} \quad (10.26)$$

当 $m_c = 2^p$ 时(全因子试验情况)有

$$r^4 + 2^p r^2 - 2^{p-1}(p + 0.5m_c) = 0 \quad (10.27)$$

当 $m_c = 2^{p-1}$ 时(部分实施法)有

$$r^4 + 2^{p-1} - 2^{p-2}(p + 0.5m_c) = 0 \quad (10.28)$$

当给定 p 和 m_0 , 就可以根据方程 10.27 和 10.28 计算出 r^2 的值来。

常用的 r^2 值见表 10.9, 如 $p=4, m_c=2^4, m_0=1$ 时, $r^2=2.00$ 。

为满足正交性, 还要求 $E=0$, 但从 10.25 式看到 $E = -2H^{-1}er^4 \neq 0$, 所以还必须对结构矩阵中 x_j^2 列的元素作变换。即对平方项 x_j^2 实施中心化。

$$\text{令 } x'_{ij} = x_{ij} - \frac{1}{N} \sum_i x_{ij}^2 \quad (10.29)$$

代替结构矩阵中变量平方和列, 即可使系数矩阵中除第一行和第一列元素之外, 其余元素皆为零(对称), 从而使相关矩阵中有

$$E=0$$

例如上述 $p=3, m_c=2^3, m_0=1$ 的情况下,

$$r=1.215 \quad x'_{ij} = x_{ij}^2 - \frac{1}{N} \sum_i x_{ij}^2 = x_{ij}^2 - 0.73$$

这样, 二次回归正交设计的结构矩阵成为表 10.10 的形式, 同样理由可以作出四因子的二次回归正交设计的结构矩阵 $X(m_0=1)$, 见表 10.11。

表 10.9

m_{ij}	p			
	2	3	4	5 (1/2 实施)
1	1.00	1.475	2.00	2.39
2	1.160	1.650	2.198	2.58
3	1.317	1.831	2.390	2.77
4	1.475	2.006	2.580	2.95
5	1.606	2.164	2.770	3.14
6	1.742	2.325	2.950	3.31
7	1.873	2.481	3.140	3.49
8	2.000	2.633	3.310	3.66
9	2.123	2.782	3.490	3.83
10	2.243	2.928	3.660	4.00

表 10.10

试验号	x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	x'_1	x'_2	x'_3
1	1	1	1	1	1	1	1	0.27	0.27	0.27
2	1	1	1	-1	1	-1	-1	0.27	0.27	0.27
3	1	1	-1	1	-1	1	-1	0.27	0.27	0.27
4	1	1	-1	-1	-1	-1	1	0.27	0.27	0.27
5	1	-1	1	1	-1	-1	1	0.27	0.27	0.27
6	1	-1	1	-1	-1	1	-1	0.27	0.27	0.27
7	1	-1	-1	1	1	-1	-1	0.27	0.27	0.27
8	1	-1	-1	-1	1	1	1	0.27	0.27	0.27
9	1	1.215	0	0	0	0	0	0.746	-0.73	-0.73
10	1	-1.215	0	0	0	0	0	0.746	-0.73	-0.73
11	1	0	1.215	0	0	0	0	-0.73	0.746	-0.73
12	1	0	-1.215	0	0	0	0	-0.73	0.746	-0.73
13	1	0	0	1.215	0	0	0	-0.73	-0.73	0.746
14	1	0	0	-1.215	0	0	0	-0.73	-0.73	0.746
15	1	0	0	0	0	0	0	-0.73	-0.73	-0.73

表 10.11

试验号	x_0	x_1	x_2	x_3	x_4	x_1x_2	x_1x_3	x_1x_4	x_2x_3
1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	-1	1	1	-1	1
3	1	1	1	-1	1	1	-1	1	-1
4	1	1	1	-1	-1	1	-1	-1	-1
5	1	1	-1	1	1	-1	1	1	-1
6	1	1	-1	1	-1	-1	1	-1	-1
7	1	1	-1	-1	1	-1	-1	1	1
8	1	1	-1	-1	-1	-1	-1	-1	1
9	1	-1	1	1	1	-1	-1	-1	1
10	1	-1	1	1	-1	-1	-1	1	1
11	1	-1	1	-1	1	-1	1	-1	-1
12	1	-1	1	-1	-1	-1	1	1	-1
13	1	-1	-1	1	1	1	-1	-1	-1
14	1	-1	-1	1	-1	1	-1	1	-1
15	1	-1	-1	-1	1	1	1	-1	1
16	1	-1	-1	-1	-1	1	1	1	1
17	1	1.414	0	0	0	0	0	0	0
18	1	-1.414	0	0	0	0	0	0	0
19	1	0	1.414	0	0	0	0	0	0
20	1	0	-1.414	0	0	0	0	0	0
21	1	0	0	1.414	0	0	0	0	0
22	1	0	0	-1.414	0	0	0	0	0
23	1	0	0	0	1.414	0	0	0	0
24	1	0	0	0	-1.414	0	0	0	0
25	1	0	0	0	0	0	0	0	0

续表 10.11

x_3x_4	x_3x_4	$x'_1=x_1^2-0.8$	$x'_2=x_2^2-0.8$	$x'_3=x_3^2-0.8$	$x'_4=x_4^2-0.8$
1	1	0.2	0.2	0.2	0.2
-1	-1	0.2	0.2	0.2	0.2
1	-1	0.2	0.2	0.2	0.2
-1	1	0.2	0.2	0.2	0.2
-1	1	0.2	0.2	0.2	0.2
1	-1	0.2	0.2	0.2	0.2
-1	-1	0.2	0.2	0.2	0.2
1	1	0.2	0.2	0.2	0.2
1	1	0.2	0.2	0.2	0.2
-1	-1	0.2	0.2	0.2	0.2
1	-1	0.2	0.2	0.2	0.2
-1	1	0.2	0.2	0.2	0.2
-1	1	0.2	0.2	0.2	0.2
1	-1	0.2	0.2	0.2	0.2
-1	-1	0.2	0.2	0.2	0.2
1	1	0.2	0.2	0.2	0.2
0	0	1.2	-0.8	-0.8	-0.8
0	0	1.2	-0.8	-0.8	-0.8
0	0	-0.8	1.2	-0.8	-0.8
0	0	-0.8	1.2	-0.8	-0.8
0	0	-0.8	-0.8	1.2	-0.8
0	0	-0.8	-0.8	1.2	-0.8
0	0	-0.8	-0.8	-0.8	1.2
0	0	-0.8	-0.8	-0.8	1.2
0	0	-0.8	-0.8	-0.8	-0.8

二、次回归正交设计的统计分析

应用组合设计进行二次回归设计的步骤和统计分析的过程归纳如下：

2. 选择组合设计并查出星号臂 r

例如 对 $p=3, m_c=2^3, m_0=1$ 的情形, 由表 10.10 的 x_1, x_2, x_3 所占列组成试验计划, 查表 10.9, 得 $r^2=1.476$

对于 $p=4, m_c=2^4, m_0=1$ 的情形, 由表 10.8 中 x_1, x_2, x_3, x_4 所占列组成试验计划, 查表 10.9 得 $r^2=2.00$

算出试验次数。

$$N=m_c+2p+m_0$$

2. 确定因子变化范围

假设有 p 个变量 $z_1, z_2, \dots, z_p$, 其中第 j 个变量上、下界值分别记为 $Z_{2j}, Z_{1j}, (j=1, \dots, p)$

根据二次回归正交设计的要求安排试验,

$$\text{令 } Z_{0j} = \frac{Z_{1j} + Z_{2j}}{2} \quad \Delta_j = \frac{Z_{2j} - Z_{0j}}{r}$$

3. 编制因子水平编码表

这与一次回归正交设计类似, 对因子作取值的线性变换。

$$x_j = \frac{Z_i - Z_{0j}}{\Delta}$$

则有

x_i	因 子			
	Z_1	Z_2	$\cdots$	Z_p
r	Z_{r1}	Z_{r2}	$\cdots$	Z_{rp}
1	$Z_{01} + \Delta_1$	$Z_{02} + \Delta_2$	$\cdots$	$Z_{0p} + \Delta_p$
0	Z_{01}	Z_{02}	$\cdots$	Z_{0p}
-1	$Z_{01} - \Delta_1$	$Z_{02} - \Delta_2$	$\cdots$	$-\Delta_p$
$-r$	Z_{r1}	Z_{r2}	$\cdots$	Z_{rp}

三、回归系数的计算与检验

根据上述二次回归正交设计的原理, 试验结果数据, 利用结构矩阵 X 的正交性, 容易得到信息矩阵 A , 常数项矩阵 B 和相关矩阵 C 。

$$A = X'X = \begin{bmatrix} N & & & & \\ & S_1 & & & \\ & & \ddots & & \\ & & & S_p & \\ & & & & S_{12} & \\ & & & & & \ddots \\ & & & & & & S_{p-1,p} \\ & & & & & & & \ddots \\ & & & & & & & & S_{11} \\ & & & & & & & & & \ddots \\ & & & & & & & & & & S_{pp} \end{bmatrix} \quad (10.30)$$

$$\text{其中} \quad S_j = \sum_k x_{kj}^2 \quad (k=1, 2, \cdots, N)$$

$$S_{ij} = \sum_k x_{ki} x_{kj} \quad i \neq j$$

$$S_{jj} = \sum_k (x'_{kj})^2$$

$$B = X'Y = \begin{bmatrix} B_0 \\ B_1 \\ \vdots \\ B_p \\ B_{12} \\ \vdots \\ B_{p-1,p} \\ B_{11} \\ \vdots \\ B_{pp} \end{bmatrix} \quad (10.31)$$

$$B_0 = \sum_k y_k \quad (k=1, 2, \dots, N)$$

$$B_j = \sum_k x_{kj} y_k$$

$$B_{ij} = \sum_k x_{ki} \cdot x_{kj} y_k \quad i \neq j$$

$$B_{jj} = \sum_k x_{kj} y_k$$

$$C = A^{-1} = \begin{bmatrix} \frac{1}{N} & & & & & \\ & S_1^{-1} & & & & \\ & & \ddots & & & \\ & & & S_p^{-1} & & \\ & & & & S_{12}^{-1} & \\ & & & & & \ddots \\ & & & & & & S_{p-1,p}^{-1} \\ & & & & & & & \ddots \\ & & & & & & & & S_{11}^{-1} \\ & & & & & & & & & \ddots \\ & & & & & & & & & & S_{pp}^{-1} \end{bmatrix} \quad (10.32)$$

由于回归系数 $b = A^{-1}B$

$$b'_0 = \frac{1}{N} \sum_k y_k \quad (k=1, 2, \dots, N)$$

$$b_j = \frac{B_j}{S_j} = \frac{\sum_k x_{kj} \cdot y_k}{\sum_k (x_{kj})^2} \quad (10.33)$$

$$b_{ij} = \frac{B_{ij}}{S_{ij}} = \frac{\sum_k x_{ki} \cdot x_{kj} \cdot y_k}{\sum_k x_{ki} x_{kj}} \quad i \neq j, \quad (10.34)$$

$$b_{jj} = \frac{B_{jj}}{S_{jj}} = \frac{\sum_k x_{kj} y_k'}{\sum_k (x'_{kj})^2} \quad (10.35)$$

这样可以得到回归方程

$$\hat{y} = b'_0 + \sum_{j=1}^p b_j x_j + \sum_{i < j} b_{ij} x_i x_j + \sum_{j=1}^p b_{jj} x_j^2, \quad (10.36)$$

由 10.27, 可得到回归方程的另一形式

$$\hat{y} = b_0 + \sum_{j=1}^p b_j x_j + \sum_{i < j} b_{ij} x_i x_j + \sum_{j=1}^p b_{jj} x_j^2 \quad (10.37)$$

$$b_0 = \bar{y} - \frac{1}{N} \sum_{j=1}^p b_j \cdot \sum_k x_{kj}^2 \quad (10.38)$$

对于二次回归正交设计的回归系数的检验与一次回归正交设计的检验完全类似。这里不加描述了。

二次回归正交设计主要用于寻找最佳配方和建立生产过程的数学模型,例如林木的施肥试验中复合肥料的配比和建立林木产量的收获模型,以及林木害虫的防治措施确定等。

下面举例说明二次回归设计的具体分析过程。

例 10.2 研究某树种苗木的水培试验,试验目标是研究某种培养液中的 5 种主要成份与苗木主根生长量的数学模型,采用二次回归正交设计法,这 5 种成份为 A, B, C, D, E。

正交组合设计中 $m_i = 2^{5-1}$, 只需做 27 次试验,变量的设计水平见下表

变量	零水平变化间距		变量设计水平 ($r=1.547$)				
			-1.547	-1.0	0	1.0	1.547
X_1-A	1.2	0.4	0.58	0.8	1.2	1.6	1.82
X_2-B	0.8	0.4	0.18	0.4	0.8	1.2	1.42
X_3-C	1.0	0.5	0.23	0.5	1.0	1.5	1.77
X_4-D	1.5	0.5	0.73	1.0	1.5	2.0	2.27
X_5-E	0.8	0.4	0.18	0.4	0.8	1.2	1.42

根据正交组合设计安排的 27 次试验计划,进行试验并观察苗木的发根量,其数量记在表 10.12 的最后一列中,结构矩阵亦见表 10.12,表中各平方项按下式计算。

$$x'_{kj} = x_{kj}^2 - \frac{1}{N} \sum_k x_{kj}^2 = x_{kj}^2 - x_{rj}^2 = 0.77$$

根据 10.23 式可以计算出 b_0

$$b_0 = \bar{y} - \frac{1}{N} \sum_{j=1}^p b_{jj} \sum_{k < j} x_{kj}^2 = \frac{1}{27} \sum y_i - \frac{b_{11}}{27} \sum x_1^2 - \\ b_{22} 27 \sum x_2^2 - \frac{b_{33}}{27} \sum x_3^2 - \frac{b_{44}}{27} \sum x_4^2 - \frac{b_{55}}{27} \sum x_5^2$$

由 10.33~10.35 式可以计算出各回归系数,为了说明问题方便,将其计算结果列在表 10.12 最后一行上,从而建立该培养液中 5 种成份对主根生长量的数学模型。这只要将诸 b_{ij} 列出即得(由于下面会看到的原因,我们不把该模型详细列出)

最后进行回归显著性检验,我们有

$$L_{\text{总}} = \sum_k (y_k - \bar{y})^2 = 0.124615$$

$$L_{\text{回归}} = \sum_k (\hat{y}_k - \bar{y})^2 = 0.112465$$

$$L_{\text{剩余}} = \sum_k (y_k - \hat{y}_k)^2 = 0.012150$$

表 10.12

试验号	x_0	x_1	x_2	x_3	x_4	x_5	x_1x_2	x_1x_3	x_1x_4	x_1x_5
1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	-1	-1	1	1	-1	-1
3	1	1	1	-1	1	-1	1	-1	1	-1
4	1	1	1	-1	-1	1	1	-1	-1	1
5	1	1	-1	1	1	-1	-1	1	1	-1
6	1	1	-1	1	-1	1	-1	1	-1	1
7	1	1	-1	-1	1	1	-1	-1	1	1
8	1	1	-1	-1	-1	-1	-1	-1	-1	-1
9	1	-1	1	1	1	-1	-1	-1	-1	1
10	1	-1	1	1	-1	1	-1	-1	1	-1
11	1	-1	1	-1	1	1	-1	1	-1	-1
12	1	-1	1	-1	-1	-1	-1	1	1	1
13	1	-1	-1	1	1	1	1	-1	-1	-1
14	1	-1	-1	1	-1	-1	1	-1	1	1
15	1	-1	-1	-1	1	-1	1	1	-1	1
16	1	-1	-1	-1	-1	1	1	1	1	-1
17	1	1.547	0	0	0	0	0	0	0	0
18	1	-1.547	0	0	0	0	0	0	0	0
19	1	0	1.547	0	0	0	0	0	0	0
20	1	0	-1.547	0	0	0	0	0	0	0
21	1	0	0	1.547	0	0	0	0	0	0
22	1	0	0	-1.547	0	0	0	0	0	0
23	1	0	0	0	1.547	0	0	0	0	0
24	1	0	0	0	-1.547	0	0	0	0	0
25	1	0	0	0	0	1.547	0	0	0	0
26	1	0	0	0	0	-1.547	0	0	0	0
27	1	0	0	0	0	0	0	0	0	0
Σx^2	27	20.7864	20.7864	20.7864	16	16	16	16	16	16
$B = \Sigma xy$	13.48	0.33	-0.9556	-0.3164	-0.03	0.01	-0.25			
$b =$	0.5230	0.2105	0.3259	0.0184	0.0006					
$B/\Sigma x^2$	0.0159	-0.0460	-0.0152	-0.0019	-0.0156					

续表

x_2x_3	x_2x_4	x_2x_5	x_3x_4	x_3x_5	x_4x_5	x'_1	x'_2	x'_3	x'_4	x'_5	y_j
1	1	1	1	1	1	0.23	0.23	0.23	0.23	0.23	0.54
1	-1	-1	-1	-1	1	0.23	0.23	0.23	0.23	0.23	0.50
-1	1	-1	-1	1	-1	0.23	0.23	0.23	0.23	0.23	0.62
-1	-1	1	1	-1	-1	0.23	0.23	0.23	0.23	0.23	0.52
-1	-1	1	1	-1	-1	0.23	0.23	0.23	0.23	0.23	0.48
-1	1	-1	-1	1	-1	0.23	0.23	0.23	0.23	0.23	0.34
1	-1	-1	-1	-1	1	0.23	0.23	0.23	0.23	0.23	0.53
1	1	1	1	1	1	0.23	0.23	0.23	0.23	0.23	0.59
1	1	-1	1	-1	-1	0.23	0.23	0.23	0.23	0.23	0.46
1	-1	1	-1	1	-1	0.23	0.23	0.23	0.23	0.23	0.45
-1	1	1	-1	-1	1	0.23	0.23	0.23	0.23	0.23	0.50
-1	-1	-1	1	1	1	0.23	0.23	0.23	0.23	0.23	0.46
-1	-1	-1	1	1	1	0.23	0.23	0.23	0.23	0.23	0.43
-1	1	1	-1	-1	1	0.23	0.23	0.23	0.23	0.23	0.37
1	-1	1	-1	1	-1	0.23	0.23	0.23	0.23	0.23	0.61
1	1	-1	1	-1	-1	0.23	0.23	0.23	0.23	0.23	0.51
0	0	0	0	0	0	1.62	-0.77	-0.77	-0.77	-0.77	0.52
0	0	0	0	0	0	1.62	-0.77	-0.77	-0.77	-0.77	0.52
0	0	0	0	0	0	-0.77	1.62	-0.77	-0.77	-0.77	0.53
0	0	0	0	0	0	-0.77	1.62	-0.77	-0.77	-0.77	0.37
0	0	0	0	0	0	-0.77	-0.77	1.62	-0.77	-0.77	0.45
0	0	0	0	0	0	-0.77	-0.77	1.62	-0.77	-0.77	0.57
0	0	0	0	0	0	-0.77	-0.77	-0.77	1.62	-0.77	0.57
0	0	0	0	0	0	-0.77	-0.77	-0.77	1.62	-0.77	0.41
0	0	0	0	0	0	-0.77	-0.77	-0.77	-0.77	1.62	0.52
0	0	0	0	0	0	-0.77	-0.77	-0.77	-0.77	1.62	0.55
0	0	0	0	0	0	-0.77	-0.77	-0.77	-0.77	-0.77	0.56
16		16		16		11.4313		11.4313			
16		16		16		11.4313		11.4313		11.4313	
-0.05		0.07		-0.07		-0.3148		-0.1234			
0.47		0.21		0.17		0.0202		-0.0276		0.0920	
-0.0031		0.0044		-0.0044		-0.0275		-0.0108			
0.0294		0.0131		0.0106		0.0018		-0.0024		0.0081	

这里 $p=20, N=27, N-p-1=6$

因为 $F = \frac{L_{\text{回归}}/p}{L_{\text{6剩余}}/N-p-1} = \frac{0.112465/20}{0.012150/6} = 2.78$

查 F 临界值表, 当 $\alpha=0.10$ 时

$$F_{0.10}(20,6)=2.84$$

现在 $F=2.78 < F_{0.10}$

因此认为回归是不显著的。

第十一章 回归旋转设计

这里所说的旋转设计是指某项试验计划,能够使与中心点距离相等的点上预测值 $\hat{y}$ 的方差相等,具有这种性质的设计称为旋转设计,或者说该设计具有旋转性。

旋转设计弥补了回归正交设计的缺点。回归正交设计虽有试验次数比较少、计算简便、消除了回归系数之间的相关性等优点,但是二次回归的预测值 $\hat{y}$ 的方差强烈地依赖于试验点在因子空间中的位置,由于误差的干扰,试验不能根据预测值直接寻找最优区域。

如果能使二次设计具有旋转性,即使与试验中心点距离相等的点上,预测值的方差相等。从这一重要问题出发,前人提出了回归旋转设计。

为了能较全面了解它,本章着重讨论回归旋转设计的基本原理以及如何进行设计和数据处理。

第一节 回归旋转设计的几个重要问题

一、旋转性和正交性条件

实际上,试验也经常希望牺牲部分的正交性而获得旋转性,如何才能使试验计划具有旋转性,具体有什么要求。为了方便叙述,下面以林业科研与生产中常用的三元二次回归方程来讨论这个问题。

在研究苗木生长与 N, P, K 营养要素的关系时,就采用三元二次回归方程来表达。那么二次回归模型为

$$y_{\alpha} = \beta_0 + \beta_1 x_{\alpha 1} + \beta_2 x_{\alpha 2} + \beta_3 x_{\alpha 3} + \beta_{12} x_{\alpha 1} \cdot x_{\alpha 2} + \beta_{13} x_{\alpha 1} \cdot x_{\alpha 3} \\ + \beta_{23} x_{\alpha 2} \cdot x_{\alpha 3} + \beta_{11} x_{\alpha 1}^2 + \beta_{22} x_{\alpha 2}^2 + \beta_{33} x_{\alpha 3}^2 + \varepsilon_{\alpha} \quad (11.1) \\ \alpha = 1, 2, \dots, N$$

它的结构矩阵为

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & x_{13} & x_{11}x_{12} & x_{11}x_{13} & x_{12}x_{13} & x_{11}^2 & x_{12}^2 & x_{13}^2 \\ 1 & x_{21} & x_{22} & x_{23} & x_{21}x_{22} & x_{21}x_{23} & x_{22}x_{23} & x_{21}^2 & x_{22}^2 & x_{23}^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{N1} & x_{N2} & x_{N3} & x_{N1}x_{N2} & x_{N1}x_{N3} & x_{N2}x_{N3} & x_{N1}^2 & x_{N2}^2 & x_{N3}^2 \end{bmatrix} \quad (11.2)$$

对应的信息矩阵 A 为 10 阶对称方阵:(见 389 页)

由 10 阶对称方阵中可以看出,信息矩阵 A 中元素的一般式表示为

$$\sum_{\alpha} x_{\alpha 1}^{\alpha_1} x_{\alpha 2}^{\alpha_2} x_{\alpha 3}^{\alpha_3} \quad (11.3)$$

其中 x 的指数 $\alpha_1, \alpha_2, \alpha_3$ 可分别取 0、1、2、3、4 等非负整数,但它们的和不超过 4,即

$$0 \leq \alpha_1 + \alpha_2 + \alpha_3 \leq 4$$

而且矩阵 A 元素可分为两类:一类元素,它们的所有指数 $\alpha_1, \alpha_2, \alpha_3$ 都是偶数或零;另一类元素,它们的所有指数 $\alpha_1, \alpha_2, \alpha_3$ 中至少有一个奇数。在 p 元 d 次回归设计中,满足下列要求,认为满足其旋转性,这些基本要求称为旋转性条件。

当 $d=1$ 时,在一次旋转设计的信息矩阵 A 中,满足不等式

$$0 \leq \alpha_1 + \alpha_2 + \dots + \alpha_p \leq 2$$

对第一类元素的要求:

$$\sum_{\alpha} x_{\alpha j}^2 = \lambda_2 N \quad j=1,2,\dots,P \quad (11.4)$$

对第二类元素要求:

$$\sum_{\alpha} x_{\alpha 1}^{a_1} \cdot x_{\alpha 2}^{a_2} \cdots x_{\alpha P}^{a_P} = 0 \quad (11.5)$$

当 $d=2$ 时, 第一类元素的要求:

$$\sum_{\alpha} x_{\alpha j}^2 = \lambda_2 N \quad (11.6)$$

$$\sum_{\alpha} x_{\alpha j}^4 = 3 \sum_{\alpha} x_{\alpha j}^2 \cdot x_{\alpha j}^2 = 3\lambda_1 N \quad (11.7)$$

$i, j=1,2,\dots,P; \lambda_1$ 为待定参数。

第二类元素的要求:

$$\sum_{\alpha} x_{\alpha 1}^{a_1} \cdot x_{\alpha 2}^{a_2} \cdots x_{\alpha P}^{a_P} = 0 \quad (11.8)$$

式 11.7 和 11.8 就是 p 元二次回归的旋转性条件, 此外, 为了使旋转设计成为可能, 还必须使矩阵 A 不退化, 为此, 必须有不等式

$$\frac{\lambda_1}{\lambda_2^2} \neq \frac{P}{P+2} \quad (11.9)$$

式 11.9 是 p 元二次旋转设计的非退化条件。待定参数 λ_2, λ_1 的比值不仅与因子个数 P 和试验次数 N 有关, 而且与 N 个试验点所在球面的半径 $\rho_{\alpha} (\alpha=1,2,\dots,N)$ 有关。这样要使设计同时满足旋转性条件和非退化条件, 要求 N 个试验点至少位于两个半径不等球面上, 同时考虑试验点在球面上合理分布即可。

要解决这问题, 主要用组合设计来实现。组合设计中的 N 个试验点

$$N = m_c + 2P + m_0$$

是分布在三个半径不等球面上的。

m_c 个点分布在半径为 $\rho_c = \sqrt{P}$ 的球面上;

$2P$ 个点分布在半径为 $\rho_r = r$ 的球面上;

m_0 个点分布集中在半径为 $\rho_0 = 0$ 的球面上。

按组合设计选取的试验点是不会使 A 矩阵退化的。至于旋转性条件, 因为它的矩阵 A 元素中

$$\sum_a x_{aj} = \sum_a x_{ai} x_{aj} = \sum_a x_{ai}^2 x_{aj} = 0$$

而它的偶次方元素

$$\sum_a x_{ai}^2 = m_c + 2r^2$$

$$\sum_a x_{ai}^4 = m_c + 2r^4$$

$$\sum_a x_{ai}^2 x_{aj}^2 = m_c$$

均不等于零。为了满足旋转性条件, 根据 11.7 式确定 r 值。在组合设计下, 当 $m_c = 2^P$ (全实施), 则有

$$2^P + 2r^4 = 3 \times 2^P \quad (11.10)$$

解此方程, 即可建立全实施时 r 值。

$r = 2^{\frac{P}{4}}$ 类似可以得到

当 $m_c = 2^{P-1}$ ($\frac{1}{2}$ 实施), $r = 2^{\frac{P-1}{4}}$

$m_c = 2^{P-2}$ ($\frac{1}{4}$ 实施), $r = 2^{\frac{P-2}{4}}$

$m_c = 2^{P-3}$ ($\frac{1}{8}$ 实施), $r = 2^{\frac{P-3}{4}}$

这样当 $P=2, m_c = 2^P$ (全实施) 时

$$r = 2^{\frac{P}{4}} = 2^{\frac{1}{2}} = 1.414$$

其他 $P=3, 4, \dots, 8$ 均可这样计算。

对于二次旋转组合设计中 m_0 的选择, 在一般情况下, 中心点不安排试验, 也不会影响计划的旋转性。但是中心点附近

区域往往是试验者关心的区域,同时适当确定 m_0 还可使二次旋转组合设计具有正交性。

m_0 的确定依下列公式

$$\frac{N}{\lambda_1/\lambda_2^2} = \frac{(m_c + 2r^2)^2(P+2)}{m_c P + 2r^4} \quad (11.11)$$

要求上式中 $\lambda_1/\lambda_2^2=1$, 但实际计算时 λ_1/λ_2^2 不为 1, 而近似于 1, 所以 $\frac{N}{\lambda_1/\lambda_2^2}$ 不是整数时, N 只能选取靠近 $\frac{N}{\lambda_1/\lambda_2^2}$ 的整数值。

现将 P 个因素不同实施情况下设计参数列表 11.1, 供设计时使用。

表 11.1

P	m_c	r	m_r	m_0	N
2(全实施)	4	1.414	4	8	16
3(全实施)	8	1.682	6	9	23
4(全实施)	16	2.000	8	12	36
4($\frac{1}{2}$ 实施)	8	1.682	8	7	23
5(全实施)	32	2.378	10	17	59
5($\frac{1}{2}$ 实施)	16	2.000	10	10	36
6($\frac{1}{2}$ 实施)	32	2.378	12	15	59
6($\frac{1}{4}$ 实施)	16	2.000	12	8	36
7($\frac{1}{2}$ 实施)	64	2.828	14	22	100
7($\frac{1}{4}$ 实施)	32	2.378	14	13	59
8($\frac{1}{2}$ 实施)	128	3.364	16	33	177
8($\frac{1}{4}$ 实施)	64	2.828	16	20	100
8($\frac{1}{8}$ 实施)	32	2.374	16	11	59

用表 11.1 的参数进行设计时, 还得注意消除二次回归旋转组合设计中常数项 b_0 与平方项 b_{jj} 间的相关性, 为此必须对平方项施行中心化变换:

$$x'_{aj} = x_{aj}^2 - \frac{1}{N} \sum_a x_{aj}^2$$

下面列出三个因子的正交旋转组合结构矩阵,见表 11.2

表 11.2

试验号	x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	x'_1	x'_2	x'_3
1	1	1	1	1	1	1	1	0.406	0.406	0.406
2	1	1	1	-1	1	-1	-1	0.406	0.406	0.406
3	1	1	-1	1	-1	1	-1	0.406	0.406	0.406
4	1	1	-1	-1	-1	-1	1	0.406	0.406	0.406
5	1	-1	1	1	-1	-1	1	0.406	0.406	0.406
6	1	-1	1	-1	-1	1	-1	0.406	0.406	0.406
7	1	-1	-1	1	1	-1	-1	0.406	0.406	0.406
8	1	-1	-1	-1	1	1	1	0.406	0.406	0.406
9	1	1.682	0	0	0	0	0	2.234	0.594	0.594
10	1	-1.682	0	0	0	0	0	2.234	0.594	0.594
11	1	0	1.682	0	0	0	0	-0.594	2.234	0.594
12	1	0	-1.682	0	0	0	0	-0.594	2.234	0.594
13	1	0	0	1.682	0	0	0	-0.594	-0.594	2.234
14	1	0	0	-1.682	0	0	0	-0.594	-0.594	2.234
15	1	0	0	0	0	0	0	-0.594	-0.594	-0.594
16	1	0	0	0	0	0	0	-0.594	-0.594	-0.594
17	1	0	0	0	0	0	0	-0.594	-0.594	-0.594
18	1	0	0	0	0	0	0	-0.594	-0.594	-0.594
19	1	0	0	0	0	0	0	-0.594	-0.594	-0.594
20	1	0	0	0	0	0	0	-0.594	-0.594	-0.594
21	1	0	0	0	0	0	0	-0.594	-0.594	-0.594
22	1	0	0	0	0	0	0	-0.594	-0.594	-0.594
23	1	0	0	0	0	0	0	-0.594	-0.594	-0.594

二、二次旋转组合设计的通用性

二次回归旋转组合设计,解决同一球面上各试验点预测值 $\hat{y}$ 的方差相等的问题,但它还存在两个以不同半径球面上试验点预测 $\hat{y}$ 的方差不等的问题。在解决林业试验设计的统计分析工作中,经常会遇到比较不同半径球面上各试验点的预测值 $\hat{y}$ 的问题。

如果某项设计计划不但保持旋转性,同时还具有各试验点与中心点的距离 ρ 在因子空间编码值 $0 \sim 1$ 范围内,其预测值 $\hat{y}$ 方差基本相等的性质,认为设计的计划具有旋转性和通用性。这样的设计称为通用旋转组合设计,从而也解决了上述问题。如何实现计划的通用性,前人已解决了这个问题。

在 P 个因素情况下,其预测值 $\hat{y}$ 的方差为

$$D(\hat{y}) = \frac{(P+2)\sigma^2}{[(P+2)\lambda_4 - P](N/\lambda_4)} \times \left[1 + \frac{\lambda_4 - 1}{\lambda_4} \rho^2 + \frac{(P+1)\lambda_4 - (P-1)}{2\lambda_4^2(P+2)} \rho^4 \right] \quad (11.12)$$

对任一个旋转组合设计方案,因子个数 P 与比值 N/λ_4 是确定的,从式 11.12 可以看出 $D(\hat{y})$ 是 λ_4 和 ρ 的函数。如果一个回归设计可使它的预测方差 $D(\hat{y})$, 在区间 $0 < \rho < 1$ 内基本上保持某一个常数的话,要使旋转组合设计具有通用性,关键在于确定 λ_4 。我们在区间 $0 < \rho < 1$ 内插入如下分点

$$0 < \rho_1 < \rho_2 < \cdots < \rho_n < 1$$

然后来确定 λ_4 , 使得 11.12 式在 ρ_i 处的值与 $\rho=0$ 处的值的差平方和最小,即

$$Q(\lambda_4) = f_0^2(\lambda_4) \sum_{i=1}^n [f_1(\lambda_4)\rho_i^2 + f_2(\lambda_4)\rho_i^4]^2 \text{ 为 min, 其中}$$

$$f_0(\lambda_4) = \frac{P+2}{[(P+2)\lambda_4 - P](N/\lambda_4)} \quad (11.13)$$

$$f_1(\lambda_4) = \frac{\lambda_4 - 1}{\lambda_4}$$

$$f_2(\lambda_4) = \frac{(P+1)\lambda_4 - (P-1)}{2\lambda_4^2(P+2)}$$

对不同的 P , 可以求出满足式 11.13 的 λ_4 , 而后由下式确定 N

$$N = \frac{(m_c + 2r^2)^2 (P+2)\lambda_4}{m_c P + 2r^4}$$

如果计算得结果 N 不是整数时, N 可取其最靠近的整数, 最后再确定 m_0

$$m_0 = N - m_c - m_r$$

对不同的 P 值, 二次通用旋转组合设计参数见表 11.3。

从上述结果可以看出, 适当选择 m_0 , 可满足设计的通用性。

表 11.3

P	m_c	m_r	r	λ_1	N	m_0
2	4	4	1.414	0.81	13	9
3	8	6	1.682	0.86	20	6
4	16	8	2.000	0.86	31	7
5($\frac{1}{2}$ 实施)	16	10	2.000	0.89	32	6
6($\frac{1}{2}$ 实施)	32	12	2.378	0.90	53	9
7($\frac{1}{2}$ 实施)	64	14	2.828	0.92	92	14
8($\frac{1}{2}$ 实施)	128	16	3.364	0.93	165	21
8($\frac{1}{4}$ 实施)	64	16	3.828	0.93	93	13

第二节 二次正交旋转组合设计统计方法

在林业试验设计中, 二次旋转设计最常用的是二次旋转组合设计, 本节着重阐述它的一般方法。

假设某项试验有 P 个因素 $Z_1, Z_2, \dots, Z_P$, 设计中首先拟定每个因素的上、下水平, 以 Z_{2j} 表示第 j 因素的上水平, 用

Z_{1j} 表示第 j 因素的下水平。

以每个因素的上、下水平 Z_{2j}, Z_{1j} 计算零水平 Z_{0j} 。

$$Z_{0j} = (Z_{2j} + Z_{1j}) / 2 \quad (11.14)$$

某因素 j 的变化间隔 Δ_j

$$\Delta_j = (Z_{2j} - Z_{1j}) / r \quad (11.15)$$

11.15 式中 r 是待定参数,它是根据二次旋转设计确定的。

对第 j 因素($j=1, 2, \dots, P$)的各水平的取值进行编码变换

$$X_{aj} = (Z_{aj} - Z_{0j}) / \Delta_j$$

然后编制因子水平编码如表 11.4。

表 11.4

X_{aj}	Z_1	Z_2	...	Z_P
$+r$	Z_{21}	Z_{22}	...	Z_{2P}
$+1$	$Z_{01} + \Delta_1$	$Z_{02} + \Delta_2$	...	$Z_{0P} + \Delta_P$
0	Z_{01}	Z_{02}	...	Z_{0P}
-1	$Z_{01} - \Delta_1$	$Z_{02} - \Delta_2$	...	$Z_{0P} - \Delta_P$
$-r$	Z_{11}	Z_{12}	...	Z_{1P}

对因素 $Z_1, Z_2, \dots, Z_P$ 进行水平编码后,以无量纲的规范变量 X_j 表示。

为了设计时方便下面列出二元、三元二次正交旋转组合设计方和它们的结构矩阵。

二次回归正交旋转组合设计试验结果的统计分析与二次回归正交组合设计的结果统计分析方法相似。

表 11.4.1

处理号	x_1	x_2	说 明
m_1 {	1	1	m_1 为二水平全因子试验 $2^2=4$ 次
2	1	-1	
3	-1	1	
4	-1	-1	
m_2 {	1.414	0	m_2 为星号点试验次数 $2P$, 即 $2 \times 2 = 4$
6	-1.414	0	
7	0	1.414	
8	0	-1.414	
m_0 {	0	0	m_0 为中心点的试验
10	0	0	
11	0	0	
12	0	0	
13	0	0	
14	0	0	
15	0	0	
16	0	0	

* 二元二次正交旋转组合设计方案

表 11.4.2

处理号	x_0	x_1	x_2	x_1x_2	x'_1	x'_2
m_c	1	1	1	1	0.5	0.5
	2	1	1	-1	0.5	0.5
	3	1	-1	1	0.5	0.5
	4	1	-1	-1	0.5	0.5
m_c	5	1	1.414	0	1.5	-0.5
	6	1	-1.414	0	1.5	-0.5
	7	1	0	1.414	-0.5	1.5
	8	1	0	-1.414	-0.5	1.5
m_d	9	1	0	0	-0.5	-0.5
	10	1	0	0	-0.5	-0.5
	11	1	0	0	-0.5	-0.5
	12	1	0	0	-0.5	-0.5
	13	1	0	0	-0.5	-0.5
	14	1	0	0	-0.5	-0.5
	15	1	0	0	-0.5	-0.5
	16	1	0	0	-0.5	-0.5

* 二元二次正交旋转组合设计结构矩阵

表 11.4.3

处理号	x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	x'_1	x'_2	x'_3
m_c	1	1	1	1	1	1	1	0.406	0.406	0.406
	2	1	1	-1	1	-1	-1	0.406	0.406	0.406
	3	1	-1	1	-1	1	-1	0.406	0.406	0.406
	4	1	-1	-1	-1	-1	1	0.406	0.406	0.406
	5	1	-1	1	1	-1	1	0.406	0.406	0.406
	6	1	-1	1	-1	1	-1	0.406	0.406	0.406
	7	1	-1	-1	1	1	-1	0.406	0.406	0.406
	8	1	-1	-1	-1	1	1	0.406	0.406	0.406
m_c	9	1	1.682	0	0	0	0	2.234	-0.594	-0.594
	10	1	-1.682	0	0	0	0	2.234	-0.594	-0.594
	11	1	0	1.682	0	0	0	-0.594	2.234	-0.594
	12	1	0	-1.682	0	0	0	-0.594	2.234	-0.594
	13	1	0	0	1.682	0	0	-0.594	-0.594	2.234
	14	1	0	0	-1.682	0	0	-0.594	-0.594	2.234
m_o	15	1	0	0	0	0	0	-0.594	-0.594	-0.594
	16	1	0	0	0	0	0	-0.594	-0.594	-0.594
	17	1	0	0	0	0	0	-0.594	-0.594	-0.594
	18	1	0	0	0	0	0	-0.594	-0.594	-0.594
	19	1	0	0	0	0	0	-0.594	-0.594	-0.594
	20	1	0	0	0	0	0	-0.594	-0.594	-0.594
	21	1	0	0	0	0	0	-0.594	-0.594	-0.594
	22	1	0	0	0	0	0	-0.594	-0.594	-0.594
	23	1	0	0	0	0	0	-0.594	-0.594	-0.594

* 三元二次正交旋转组合设计结构矩阵

林分生长预测是森林资源管理的重要组成部分。研究林

分生长,一般通过一定数量的临时标准地来完成,这样工作量较大。作者试图应用回归设计方法探讨林分生长预测问题。

要建立林分蓄积量 y 与林分密度 x_1 和林分胸径 x_2 的关系。由于考虑在一个局部范围(如一个林场),本研究在一个林场范围内,根据实际林分状况研究杉木有关方程式,所以因子变幅见表 11.4.4。

表 11.4.4

水平	Z_1 (株/亩)	Z_2 (cm)
上水平 Z_2	200	15.5
下水平 Z_1	100	6.5
零水平 Z_0	150	11

根据二元二次正交旋转组合设计方案,典型调查了 16 个小班,具体结果见表 11.4.5。

根据试验结果,按表 11.4.5 计算各项回归系数,其中 b_0 由下式计算

$$b_0 = \frac{1}{16} \sum y_i - \frac{\sum x_{a_j}^2}{N} \times \sum_{j=1}^2 b_{jj}$$

经计算得到二元二次回归方程

$$\begin{aligned} \hat{y} = & 7.4625 + 0.0806x_1 + 1.5838x_2 + 1.1000x_1x_2 \\ & - 0.5875x_1^2 - 0.3625x_2^2 \end{aligned}$$

依据表 11.4.5 进行回归显著性检验,列出方差分析如表 11.4.6。

表 11.4.5

处理号	x_0	x_1	x_2	x_1x_2	x'_1	x'_2	$y(M^3/\text{亩})$
1	1	1	1	1	0.5	0.5	10.3
2	1	1	-1	-1	0.5	0.5	5.3
3	1	-1	1	-1	0.5	0.5	7.0
4	1	-1	-1	1	0.5	0.5	6.4
5	1	1.414	0	0	1.5	-0.5	5.0
6	1	-1.414	0	0	1.5	-0.5	6.1
7	1	0	1.414	0	-0.5	1.5	8.5
8	1	0	-1.414	0	-0.5	1.5	3.5
9	1	0	0	0	-0.5	-0.5	8.2
10	1	0	0	0	-0.5	-0.5	7.5
11	1	0	0	0	-0.5	-0.5	7.8
12	1	0	0	0	-0.5	-0.5	7.0
13	1	0	0	0	-0.5	-0.5	6.8
14	1	0	0	0	-0.5	-0.5	7.5
15	1	0	0	0	-0.5	-0.5	8.2
16	1	0	0	0	-0.5	-0.5	6.7
$a_j = \sum x_j^2$	16	8	8	4	8	8	$\sum y = 111.8$
$B_j = \sum x_j y$	111.8	0.6446	12.67	4.4	-4.7	-2.9	$SS_B = 37.5975$
$b_j = \frac{B_j}{a_j}$	7.4625	0.0806	1.5838	1.1000	-0.5875	-0.3625	$SS_B = 28.7706$
$u_j = \frac{B_j^2}{a_j}$		0.0519	20.0661	4.8400	2.7613	1.0513	$SS_B = 8.8269$

表 11.4.6

变差来源	离差平方和	自由度	均方	均方比	显著性
x_1	0.0519	1	0.0519	0.059	不显著
x_2	20.0661	1	20.0661	22.733	$\alpha=0.01$ 水平显著
x_1x_2	4.8400	1	4.8400	5.483	$\alpha=0.05$ 水平显著
x_1^2	2.7613	1	2.7613	3.128	$\alpha=0.25$ 水平显著
x_2^2	1.0513	1	1.0513	1.191	不显著
回归	28.7706	5	5.7541	6.519	$\alpha=0.05$ 水平显著
剩余	8.8269	10	0.8827		
总	37.5975	15			

从表 11.4.6 可以看出回归显著,同时可以计算相关系数 $R=0.874$ 。其中 x_2, x_1x_2, x_1^2 项达到不同水平上显著,而 x_1 和 x_2^2 不显著。由于试验具有正交性,消除回归系数之间相关性,故可以把它们从二元二次回归方程中除去,将其平方和及自由度并入剩余项,进行第二次方差分析,其结果趋势是相同的。这样回归方程简化为

$$\hat{y}=7.4625+1.5838x_2+1.1000x_1x_2-0.5875x_1^2$$

第三节 通用旋转组合设计的统计方法

采用通用旋转组合设计时请注意 m_0 的取值,而 m_c 和 m_r 的取值与正交旋转组合设计。 m_0 的取值查表 11.3,其方法步骤阅读本章第二节的内容。

它的回归系数由下式计算

$$b=(X'X)^{-1}(X'Y)$$

其中 $(X'X)^{-1}$ 为设计的相关矩阵, $(X'Y)$ 为常数项矩阵 B , 在通用旋转设计下, 则

$$\begin{bmatrix} b_0 \\ b_{11} \\ b_{22} \\ \vdots \\ b_{pp} \\ b_1 \\ b_2 \\ \vdots \\ b_p \\ b_{12} \\ b_{13} \\ \vdots \\ b_{p-1,p} \end{bmatrix} = \begin{bmatrix} K & E & E & \cdots & E \\ E & F & G & \cdots & G \\ E & G & F & \cdots & G \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ E & G & G & \cdots & F \\ & & e^{-1} & & \\ & & & e^{-1} & \\ & & & & \ddots \\ & & & & e^{-1} \\ & & & & & m_c^{-1} \\ & & & & & m_c^{-1} \\ & & & & & & \ddots \\ & & & & & & m_c^{-1} \end{bmatrix} \begin{bmatrix} \sum y_a \\ \sum x_{a1}^2 y_a \\ \sum x_{a2}^2 y_a \\ \vdots \\ \sum x_{ap}^2 y_a \\ \sum x_{a1} y_a \\ \sum x_{a2} y_a \\ \vdots \\ \sum x_{ap} y_a \\ \sum x_{a1} x_{a2} y_a \\ \sum x_{a1} x_{a3} y_a \\ \vdots \\ \sum x_{ap-1} x_{ap} y_a \end{bmatrix}$$

所以

$$\left. \begin{aligned} b_0 &= K \sum_a y_a + E \sum_{j=1}^p \left(\sum_a x_{aj}^2 y_a \right) \\ b_j &= e^{-1} \sum_a x_{aj} y_a \\ b_{ij} &= m_c^{-1} \sum_a x_{ai} x_{aj} y_a \\ b_{jj} &= (F-G) \sum_a x_{aj}^2 y_a + G \sum_{j=1}^p \sum_a x_{aj}^2 y_a + E \sum_a y_a \end{aligned} \right\} \quad (11.16)$$

其中二次通用旋转组合设计 K, E, F, G 的值见表 11.5。

表 11.5

P	K	$-E$	F	G
2	0.2	0.1	0.14375	0.01875
3	0.1663402	0.056792	0.06939	0.00689003
4	0.1428571	0.0357142	0.0349702	0.00372023
5(1/2)	0.1590909	0.0340909	0.0340909	0.0028409
5	0.0987822	0.019101	0.0170863	0.00146131
6(1/2)	0.1107487	0.018738	0.0168422	0.00121724
7(1/2)	0.0703125	0.00976562	0.000488281	0.00830078

回归方程的显著性检验方法,首先也是进行离差平方和分解,并依下列式子进行

$$L_{\text{总}} = L_{\text{回}} + L_{\text{剩}} \quad (11.17)$$

而各项离差平方和为

$$L_{\text{总}} = \sum_{\alpha} (y_{\alpha} - \bar{y})^2 = \sum_{\alpha} y_{\alpha}^2 - (\sum_{\alpha} y_{\alpha})^2 / N \quad (11.18)$$

$$L_{\text{剩}} = \sum_{\alpha} y_{\alpha}^2 - \sum_{j=0}^p b_j B_j - \sum_{i < j} b_{ij} B_{ij} - \sum_{j=1}^p b_{jj} B_{jj} \quad (11.19)$$

$$L_{\text{回}} = L_{\text{总}} - L_{\text{剩}} \quad (11.20)$$

而各项离差平方和的相应自由度为

$$df_{\text{总}} = N - 1$$

$$df_{\text{剩}} = N - C_{p+2}^2$$

$$df_{\text{回}} = C_{p+2}^2 - 1$$

在通用旋转组合设计中,中心点一般均做重复试验,如果重复次数为 m_0 ,试验结果表示为 $y_{01}, y_{02}, \dots, y_{0m_0}$ 。完全的试验误差是

$$L_{\text{误}} = \sum_{i=1}^{m_0} (y_{0i} - \bar{y}_0)$$

相应的自由度 $df_{\text{误}} = m_0 - 1$ 。剩余项离差平方和与误差项离差平方和之差称为失拟平方和,它是由模型或其他因素引起的误差,记为 L_{IF}

$$L_{\text{IF}} = L_{\text{剩}} - L_{\text{误}} \quad (11.21)$$

$$df_{\text{IF}} = df_{\text{剩}} - df_{\text{误}} \quad (11.22)$$

对于失拟程度必须进行检验,可用 F 统计量

$$F_1 = \frac{L_{\text{IF}}/df_{\text{IF}}}{L_{\text{误}}/df_{\text{误}}} \quad (11.23)$$

如果 $F_1 < F_{0.05}$, 差异不显著,则可直接对回归进行检验;如果 $F_1 > F_{0.05}$, 差异显著,表明存在其他不可忽视的因素对试验结果的影响,需要进一步考察。

回归方程的显著性检验,也用 F 统计量

$$F_2 = \frac{L_{\text{回}}/df_{\text{回}}}{L_{\text{剩}}/df_{\text{剩}}} \quad (11.24)$$

如果 $F_2 < F_{0.05}$, 则认为回归关系不显著,说明此方程不宜应用;如果 $F_2 > F_{0.05}$, 则回归关系显著,表明方程可用。

各变量在回归方程中作用,通过对回归系数的显著性检验来实现。

$$\left. \begin{aligned} t_j &= |b_j| / \sqrt{e^{-1}(L_{\text{误}}/df_{\text{误}})} \\ t_{ij} &= |b_{ij}| / \sqrt{m_c^{-1}(L_{\text{误}}/df_{\text{误}})} \\ t_{jj} &= |b_{jj}| / \sqrt{F(L_{\text{误}}/df_{\text{误}})} \end{aligned} \right\} \quad (11.25)$$

当 F_1 检验结果不显著,可用 $L_{\text{剩}}/df_{\text{剩}}$ 代替 11.25 式中 $L_{\text{误}}/df_{\text{误}}$ 。其中 K 、 m_c 、 F 如前所示,而 $e = m_c + 2r^2$ 。

例 11.1 作者在研究杉木苗木氮磷钾反应模式及最佳施肥组合时,探讨氮、磷、钾三个因素对苗木生物量的影响。

氮、磷、钾三因素分别记为 Z_1, Z_2, Z_3 , 其上下水平列于表 11.6。

表 11.6 kg/亩

水 平	N	P_2O_5	K_2O
Z_{2j}	11.91	17.04	10.29
Z_{1j}	0	0	0

按 11.14 和 11.15 式计算三因素零水平和变化间距, 具体列表 11.7

表 11.7

项 目	Z_1	Z_2	Z_3
Z_{0j}	5.955	8.52	5.15
Δ_j	3.365	4.86	2.99

根据表 11.4 内容计算各因素各水平编码值见表 11.8。

表 11.8

	Z_1	Z_2	Z_3
$+r$	11.91	17.04	10.29
$+1$	9.32	13.38	8.14
0	5.955	8.52	5.15
-1	2.59	3.66	2.16
$-r$	0	0	0

根据三元二次通用旋转组合设计的要求, $P=3, m_c=8, m_r=6, m_0=6, r=1.682$ 。

$$N = m_c + m_r + m_0 = 20$$

依上述的要求可以列出设计方案, 具体见表 11.9 的前三列。试验的观察指标为总生物量 $y(\bar{G})$, 试验结果见表 11.9 最

后一列。

根据试验结果,可得出苗木总生物量的 N 、 P 、 K 反应模式

$$\begin{aligned} \hat{y} = & 8.5068 + 0.1154x_1 + 0.0197x_2 - 0.1873x_3 \\ & - 0.2988x_1x_2 + 0.0663x_1x_3 - 0.3013x_2x_3 \\ & - 0.2142x_1^2 - 0.4582x_2^2 - 0.1965x_3^2 \end{aligned}$$

表 11.9

处理号	x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	x_1^2	x_2^2	x_3^2	$y(\bar{G})$
1	1	1	1	1	1	1	1	1	1	1	6.9
2	1	1	1	-1	1	-1	-1	1	1	1	7.96
3	1	1	-1	1	-1	1	1	1	1	1	8.45
4	1	1	-1	-1	-1	-1	-1	1	1	1	7.95
5	1	-1	1	1	-1	-1	1	1	1	1	7.51
6	1	-1	1	-1	-1	1	-1	1	1	1	8.48
7	1	-1	-1	1	1	1	-1	1	1	1	7.51
8	1	-1	-1	-1	1	-1	1	1	1	1	7.63
9	1	1.682	0	0	0	0	0	2.829	0	0	8.11
10	1	-1.682	0	0	0	0	0	2.829	0	0	7.25
11	1	0	1.682	0	0	0	0	0	2.829	0	7.07
12	1	0	-1.682	0	0	0	0	0	2.829	0	6.91
13	1	0	0	1.682	0	0	0	0	0	2.829	7.46
14	1	0	0	-1.682	0	0	0	0	0	2.829	8.00
15	1	0*	0	0	0	0	0	0	0	0	8.675
16	1	0	0	0	0	0	0	0	0	0	8.088
17	1	0	0	0	0	0	0	0	0	0	8.546
18	1	0	0	0	0	0	0	0	0	0	8.664
19	1	0	0	0	0	0	0	0	0	0	8.236
20	1	0	0	0	0	0	0	0	0	0	8.97
$S = \sum x_j^2$	20	13.658	13.658	13.658	8	8	8	24	24	24	
$B = \sum x_j y$	158.369	1.5765	0.2891	-2.5583	-2.39	0.53	-2.41	105.843	101.9394	106.1283	
b	8.5068	0.1154	0.0197	-0.1873	-0.2988	0.0663	-0.3013	-0.2142	-0.4582	-0.1965	
Bb	1347.2134	0.1820	0.0053	0.4792	0.714	0.0351	0.726	-22.6717	-46.7063	-20.853	

* 三因素二次通用旋转组合设计的结构矩阵与总生物量的关系

第十二章 饱和 D —最优设计

在阐述饱和 D —最优设计之前,就饱和 D —最优设计的由来,简要说明其原理。

到现在为止已介绍了多种设计方案,这自然向人们提出一个问题,可能多种方案中存在一个方案所建立的回归方程能使回归预测值与观察拟合最好的方案,即最优方案。人们要识别设计方案,必须有一个评价的标准,否则就无法评价与识别。为此,统计学家先后提出许多标准,例如 G —优良性、 D —优良性、 E —优良性等。本章中只介绍 D —优良性,以及相应 D —最优设计方案。

第一节 D —优良性

一、回归模型扩展

上面所讨论回归模型基本上都是 p 元 d 次完全多项式。由于变量之间的关系不同,例如可能变量之间不存在交互作用,变量的取值不一定按幂函数变化,而是按某一种连续函数。不论变量之间存在何种回归关系,都可以把回归模型加以拓广,表示为

$$y_{\alpha} = \beta_1 f_1(x_{\alpha}) + \beta_2 f_2(x_{\alpha}) + \cdots + \beta_m f_m(x_{\alpha}) + \epsilon_{\alpha} \quad (12.1) \\ \alpha = 1, 2, \cdots, N$$

这里 $\epsilon_{\alpha} (\alpha = 1, 2, \cdots, N)$ 是遵从 $N(0, \sigma)$ 的相互独立的随机变

量。 x_a 是给定的因子区域 $\mathfrak{a}$ 中一点, 假定因子空间为 P 维欧氏空间, 则 x_a 为 P 维向量 $(x_{a1}, x_{a2}, \dots, x_{aP})$; $f_1(x_a), f_2(x_a), \dots, f_m(x_a)$ 都是连续函数; $\beta_1, \dots, \beta_m$ 是 m 个待定参数。

$$F(x) = \begin{bmatrix} f_1(X) \\ f_2(X) \\ \vdots \\ f_m(X) \end{bmatrix} \quad \beta' = (\beta_1, \beta_2, \dots, \beta_m)$$

那么式 12.1 的回归方程用矩阵表示为

$$E(y) = \beta' F(X) \quad (12.2)$$

对于 12.1 模型, 当 $f_1(X), f_2(X), \dots, f_m(X)$ 是一组幂函数时, 模型 12.1 就是多项式模型。

例如, 当 $m=6$ 时

$$\begin{aligned} f_1(x_1, x_2) &= 1 \\ f_2(x_1, x_2) &= x_1 \\ f_3(x_1, x_2) &= x_2 \\ f_4(x_1, x_2) &= x_1 x_2 \\ f_5(x_1, x_2) &= x_1^2 \\ f_6(x_1, x_2) &= x_2^2 \end{aligned}$$

就得到二元二次回归模型。

假定试验计划是由 N 个试验点 $X_1, X_2, \dots, X_N$ 组成, 则模型 12.1 的结构矩阵为

$$X = \begin{bmatrix} f_1(X_1) & f_2(X_1) & \cdots & f_m(X_1) \\ f_1(X_2) & f_2(X_2) & \cdots & f_m(X_2) \\ \vdots & \vdots & \vdots & \vdots \\ f_1(X_N) & f_2(X_N) & \cdots & f_m(X_N) \end{bmatrix}$$

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ \vdots & & \vdots \\ \text{对称部分} & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ \vdots & \vdots & \vdots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ \vdots & & \vdots \\ \text{对称部分} & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ \vdots & & \vdots \\ \text{对称部分} & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ \vdots & \vdots & \vdots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ \vdots & & \vdots \\ \text{对称部分} & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ \vdots & \vdots & \vdots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ \vdots & & \vdots \\ \text{对称部分} & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ \vdots & & \vdots \\ \text{对称部分} & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ \vdots & \vdots & \vdots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ \vdots & & \vdots \\ \text{对称部分} & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ \vdots & & \vdots \\ \text{对称部分} & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ \vdots & \vdots & \vdots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ \vdots & \vdots & \vdots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ \vdots & \vdots & \vdots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ \vdots & \vdots & \vdots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ \vdots & \vdots & \vdots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ \vdots & & \vdots \\ \text{对称部分} & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为

信息矩阵为

$$A = X'X = \sum_{a=1}^N F(X_a)F'(X_a),$$

对于完全多项式回归, 它们的结构矩阵可以完全确定试验计划, 但对拓广后的模型 12.1 不能看出它的试验计划, 需要进一步研究。

假定试验是在给定的因子空间一组点 $X_1, X_2, \dots, X_n$ 上进行, 每个点重复若干次。这组点与其对应的重复次数所组成的集体, 称为一个离散方案, 用 $W(N)$ 表示:

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ n_1, n_2, \dots, n_n \end{bmatrix} \quad (12.3)$$

其中 $X_1, X_2, \dots, X_n$ 称为方案 W 的谱点; $\sum_{i=1}^n n_i = N$ 。

如果把离散方案中每一点的重复数改用其重复频率 $P_i = \frac{n_i}{N}$ 表示, 且 P_i 可以在 $[0, 1]$ 中任意取值, 则称连续方案。

$$W(N) = \begin{bmatrix} X_1, X_2, \dots, X_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (12.4)$$

其中 P_i 称为点 X_i 的测度, 且 $\sum_{i=1}^n P_i = 1$

由式 12.3 和 12.4 可得离散方案的信息矩阵为

$$A(W) = \sum_{i=1}^n n_i F(X_i)F'(X_i) = \begin{bmatrix} \sum n_i f_1^2(X_i) & \sum n_i f_1(X_i)f_2(X_i) & \dots & \sum n_i f_1(X_i)f_m(X_i) \\ & \sum n_i f_2^2(X_i) & \dots & \sum n_i f_2(X_i)f_m(X_i) \\ & & \dots & \vdots \\ \text{对称部分} & & & \sum n_i f_m^2(X_i) \end{bmatrix} \quad (12.5)$$

连续方案的信息矩阵为