线性模型分析方法

——适用于动物科学和动物医学

王继华　李旭东 著

科学出版社

线性模型分析方法

——适用于动物科学和动物医学

王继华　李旭东　著

科学出版社

北京

内 容 简 介

本书通过大量实例详细介绍线性模型的建立方法和统计分析的基本原理、方法与常见问题，包括回归分析模型及其应用、方差-协方差分析模型及其应用、多元线性模型及其应用、线性混合模型及其应用、线性混合模型参数估计方法等，统计分析方法包括最小二乘(LS)法、最小范数二次无偏估计(MINQUE)、最大似然(ML)法和约束最大似然(REML)法.

本书介绍的线性模型分析方法和技术，只需要理工科大学本科的数学基础. 书中给出的算法规律性很强，便于理解、使用和记忆. 本书还对线性模型建立和分析中常见的问题及注意事项做了详细介绍.

本书适合动物科学、动物医学和农学等非数学专业的科技人员、高年级本科生和研究生阅读，对于相关专业的中青年大学教师，也有参考价值.

图书在版编目(CIP)数据

线性模型分析方法：适用于动物科学和动物医学/王继华，李旭东著. —北京：科学出版社，2015. 6

ISBN 978-7-03-044957-3

Ⅰ. ①线… Ⅱ. ①王… ②李… Ⅲ. ①线性模型-分析方法 Ⅳ. ①O212

中国版本图书馆 CIP 数据核字(2015)第 129330 号

责任编辑：李 欣 / 责任校对：钟 洋

责任印制：徐晓晨/ 封面设计：陈 敬

科 学 出 版 社 出版

北京东黄城根北街 16 号

邮政编码：100717

http://www.sciencep.com

北京教图印刷有限公司 印刷

科学出版社发行 各地新华书店经销

*

2015 年 9 月第 一 版 开本：720×1000 B5

2015 年 9 月第一次印刷 印张：16 3/4

字数：338 000

定价：98.00 元

(如有印装质量问题，我社负责调换〈印科〉)

作者简介

王继华(1957年11月～),男,河北省大名县人,教授,硕士生导师,河北省省级优秀教师.主要研究动物科学,包括动物育种原理与方法、动物营养与饲料等.已发表学术论文85篇.已出版学术著作9部:《家畜育种学导论》《鸡饲料配方设计技术》《动物科学研究方法》《仔猪饲料配方设计实用新技术》《仔猪饲料配方设计高级技术》《肥育猪饲料配方设计技术》《猪的生长发育及营养调控技术》《饲用矿物元素配合物的研究及应用》《蛋鸡饲料配方设计技术》等.

李旭东(1982.1.26-),河海大学博士,现在四川省农田水利局工作.

序

目前国内外已有一些线性混合模型类出版物可读，但有的并不适合非数学类工程科技人员，尤其是动物、农业和医学等生命科学领域的科技人员，而这是使用线性混合模型方法最多的人群. 此书通过大量实例，深入浅出地介绍线性模型分析方法的基本技术，包括基本概念、原理、方法和应用的完整技术.

此书系统总结国内外对线性模型分析方法的研究成果和王继华教授多年来在动物科学研究和生产上应用线性模型分析方法的研究心得. 此书介绍的这些分析线性模型的统计方法，只涉及理工科大学本科的数学知识，便于非数学专业的科技人员理解和使用.

此书特点是理论联系实际，利用线性模型分析方法处理动物饲养试验的实际问题，例如，动物饲养试验分析方法、动物营养需要量的研究方法、饲料利用率研究方法等，资料丰富、数据翔实，读者可以按照书中实例模仿使用.

承蒙作者信任和重托，本人有幸率先阅读了这部书稿，看到线性模型分析的新方法和新应用，蕴涵了作者的大量心血，是他多年来从事动物科学研究和实践的智慧结晶. 本人乐为此书作序，并热切希望本书的出版能够提升我国动物科学研究和生产实践水平，促进我国养殖业的可持续性健康发展.

中国畜牧兽医学会动物遗传育种学分会 理事长
中国农业大学 动物科技学院教授、博士生导师

张 勤

2015 年 3 月 1 日

前　言

多年来，方差-协方差分析和回归分析一直是统计分析模型的基础，这些技术有个基本假定，那就是模型残差或误差项是独立同分布的. 线性混合模型方法放松了这一基本假定，使我们可以处理更加复杂的数据结构.

线性混合模型在理工农医等各个领域得到迅速推广和应用，原因是在实际问题中，参与试验的个体是随机抽取的，并且我们的研究目标常不是这些个体本身的特征，而是它们所在总体的特征，这时把个体效应视为随机效应引入模型，可以巨大提高模型精度. 有时候一些固定效应也属于干扰因素，例如，在不同猪场测定某些添加剂的使用效果，猪场效应就是干扰因素，可以在设计试验时把一些固定效应引入模型. 设计试验时在模型中引入干扰因素，无论引入随机效应还是引入固定效应，都可巨大提高模型精度，巨大提高模型参数的推断质量. 所以统计学家一直非常重视线性混合模型的研究，使得线性混合模型理论迅速发展.

使用混合模型有很多好处，有时候可以提高估计量的精度，得到更广泛的推断；有时候使用更加合适的数学模型使我们能够更加深入地洞察数据结构，启迪我们的研究思路和研究方法. 尤其是，常见通用软件，如 SAS 和 SPSS 都增加了线性混合模型分析程序，使得线性混合模型技术更便于普及. 但是，如何合理使用这些软件包，还需要正确理解线性混合模型分析方法. 本书没有介绍有关软件包的用法，是出于两种考虑：一是，网上可以下载到有关程序使用说明，例如，SAS 的程序 Proc Mixed 或 GLIMMIX，和 SPSS 的程序 MIXED 等；二是，本书使用 OFFICE 的 EXCEL 软件计算全部实例，有助于读者理解线性混合模型的算法和过程.

本书介绍的线性模型分析技术适用于许多应用领域，希望为非数学专业的工程科技人员提供有用的工具，包括动物科学、动物医学和农学专业的研究人员，或想从事量化研究的非数学专业的研究生，以及相关专业的中青年大学教师.

本书介绍的线性模型分析方法和技术，是作者多年研究与教学实践的结晶，这些方法只需要理工科大学本科的数学基础（包括微积分、线性代数和数理统计或生物统计学）. 本书初稿作为动物科学专业的高年级本科生和研究生讲课的讲稿使用，已在本校内部印刷了 2 次. 教学实践表明，动物科学专业高年级本科生完全可以理解和正确使用我们给出的线性模型分析方法. 本书给出的添加约束的方法不仅便于理解和使用，而且，由于模型约束条件本身的特性，使得这些方法具有很强的规律性，与传统的线性混合模型正规方程一样，非常便于写出，便于记忆.

中国畜牧兽医学会动物遗传育种学分会理事长张勤教授百忙中仔细审阅了本

书全稿，提出不少修改意见和建议，使本书质量大大提高. 先生的学风和境界，学为人师，德为世范，我们再次对先生的厚爱和支持表示衷心的敬意和感谢.

本书的初始形态为讲稿，十几年来，每遇见简单易懂便于学生理解的精辟论述，就会毫不犹豫地插入讲稿，所以本书实际上参考了很多文献，在这里向原作者致以衷心的感谢.

限于本人知识和技术水平，书中难免有不足之处，恳望读者不吝赐教，电子信箱是 hdwangjihua@126. com.

王继华
2015 年 5 月 1 日
于河北工程大学动物科学系

目　录

第1章 引 论

在数量遗传学和数量生态学带动下，20世纪70年代开始兴起用数学模型方法研究动物现象. 科学家已由真理发现者转换为模型建立者(model builder)，构造假设、建立模型、发展理论是科学家的首要任务. 科学研究过程由归纳-推理过程转变为假设-求证过程. 科学理论的判断标准也不再是看它能否被证明绝对正确无误，而是看它能否被重复检验.

1.1 动物科学中的线性数学模型

(1) 生命存在的方式——质和量. 任何事物都有质和量两方面，一个研究对象，量的表现常可用一个系统来描述，可把影响系统运动的各种因素表示为不同因子或变量(常量可看做是只取一个数值的变量)，用数学模型描述变量间及变量与系统间的关系，由此研究系统的运动规律.

家畜系统也表现为质和量两方面. 影响一个家畜系统的各种因子或变量间的关系，以及这些因子(或变量)与家畜系统之间的关系，同样可以而且应该用数学模型描述. 只有把家畜系统中各变量间以及各变量与系统间的复杂关系用数学模型表述，这门科学才算发展到高级阶段.

(2) 动物科学发展趋势. 过去的动物科学，没有数学也取得了辉煌成绩. 现在和今后的研究是否可以不用数学？科学发展史表明，任何一门科学，初级阶段总是以定性描述开始，深入到一定阶段，积累大量研究结果，对本门科学有一定程度了解后，便开始总结对研究对象的认识，形成知识. 在地球生物圈中，除人类本身外，家畜是人类接触最早、认识最深的一个生命系统，每种家畜都构成一个子系统. 人类对家畜系统进行了千万年观察研究，积累了大量感性认识，已总结出不同侧面的知识，形成了动物科学的不同分支，例如生理学、生物化学、营养学、遗传学、繁殖学等.

模型化方法就是把研究对象(原型)的一些次要的细节、非本质的联系舍去，从而以简化和理想化的形式去再现原型的各种复杂结构、功能和联系. 作为一种现代科学认识手段和思维方法，模型具有两方面的含义，即抽象化和具体化.

数学模型技术正成为现代动物科学研究的核心技术. 世界权威杂志 J. Anim. Sci. 在1970年以前，没发表“模型化”论文，20世纪70年代发表的有3.2%，80年代发表的有17%，90年代发表的有31.5%，2000年后发表的论文有50%涉及数

学模型，这会是一个增长最快的领域. 再者，“模型化”内涵也变为机理建模，而不停留在经验建模水平. 动物系统的机理模型化已为理解动物科学原理和生产过程贡献很多，数学模型化方法正大步进入动物科学所有学科.

(3) 数学模型的种类. 数学模型可按不同方式分类：

按应用领域(或所属学科)可分为遗传模型、营养模型、生态模型、数学生理学模型、人口模型、环境模型等，范畴更大则形成许多边缘学科，如生物数学、医学数学、数量经济学等.

按建模目的分为描述模型、分析模型、预报模型、优化模型、决策模型、控制模型等.

按照对模型结构的了解程度分：有白箱、灰箱、黑箱模型，把研究对象比喻成一只暗箱，要通过模型化来揭示它的奥妙.

按模型的表现特性分：有确定性模型和随机性模型、静态模型和动态模型(是否考虑时间因素)、离散模型和连续模型(指模型中的变量为离散型还是连续型的)等. 虽然从本质上讲大多数实际问题是随机、动态、非线性的，但由于确定性、静态、线性模型容易处理，并常可作为初步近似来解决问题，所以模型化时，常先考虑确定性、静态、线性模型. 而且实际上，很多非线性问题在小范围内可用线性模型逼近. 连续模型便于利用微积分方法求解，作理论分析，而离散模型便于计算机计算，所以用哪种模型要具体问题具体分析. 在具体模型化过程中将连续模型离散化，或将离散变量视作连续，也是常用方法.

按变量间的关系可分为线性模型和非线性模型. 按变量的形式可分为实变量模型和虚变量模型. 实变量模型是指模型中的变量为连续型变量，可取任意实数值，虚变量模型是指模型中的变量为离散型变量，有些变量只取值 0 或 1. 虚变量模型中，按模型中变量的性质，又可把模型分为固定模型、随机模型和混合模型. 根据研究对象的观测值来源可分为单向分类模型、双向分类模型、多向分类模型和系统分类模型.

按模型对实际问题的配合程度看有 3 类模型：真模型(real model)可精确描述数据，没有不可解释的变异. 真模型多数不能精确知道，也不一定是线性模型. 理想模型(ideal model)是真模型的简化或理想化. 可操作模型(apply model)是理想模型的简化版，能用于数据分析. 模型不能无限简化，不能简化到没分析价值. 要掌握理想模型简化为可操作模型的过程所做的假定，才能判断可操作模型的质量.

数学模型化的优点是能够创建科学理论，进行科学推断、解释和预测、得到观测范围外的结果. 缺点是，如果模型过于简化，为在概念上可控，把原型简化成概念的骨架，那么，由此得到的结论在回用于原型时，符合程度就很差. 原型越复杂，过分简化的缺点就越大.

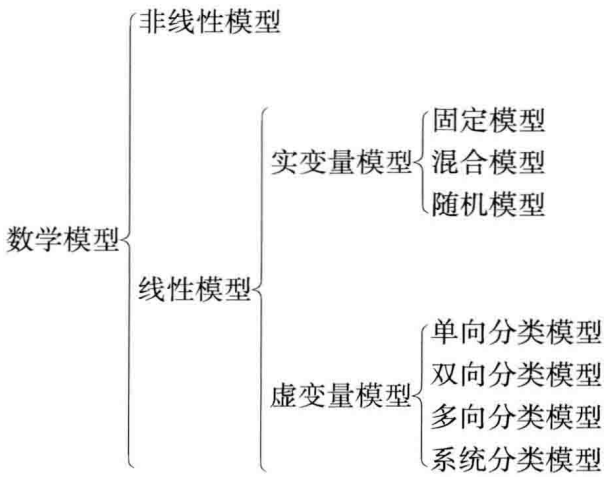
(4)线性数学模型. 在数理统计模型中，数学模型是指描述观测值与影响观测

值变异的各因素间关系的数学方程. 所有统计分析都是基于一定数学模型进行的. 线性统计模型简称线性模型(linear model),Rao(1973)最早给出系统介绍. 这种统计模型统一了许多传统的线性统计问题,例如,平均数估计、线性回归、方差-协方差分析和数量化方法 I,所以是生物学研究和应用领域最重要、最基本的一种数学模型. 多数参数方法都可归为线性模型.

由线性模型发展来的各种方法和技术,可在一定范围内推广到非线性模型. 实际问题多可用线性模型描述,或作为初级近似描述. 模型参数的估计方法有多种,这些方法多可用线性模型理论解决. 大学本科生物统计教材上介绍的经典问题,多可用线性模型表述.

在应用领域,一般从三方面对线性模型分类:一是按因子数目,分为单因子、二因子、三因子和多因子模型等;二是按因子性质分为固定模型、随机模型、混合模型等;三是按模型功能分为回归模型、方差分析模型、协方差分析模型、方差分量模型、混合模型等. 上述分类方法只是考虑到模型某一方面的特征,实践中常要考虑各种特征.

从应用角度,一般把数学模型分为如下种类:



统计模型化有两个基本思想. 模型化一般基于某些假设,所以建立模型前验证这些假设非常重要. 建好模型后还要从两方面评价模型:第一是模型拟合度. 由模型拟合的数据是否接近观测到的样本数据? 拟合数据与样本数据间的差是否呈随机分布? 第二是评价把模型用于预测更广范围的数据时的可靠性. 在一个范围内成立的数学模型,在另一个范围内未必成立,例如,元明粉(无水硫酸钠)对生长猪的生理作用,当日粮中添加量在 0～0.3%时,有健胃、促生长作用;在 0.4%～0.6%时,有健胃、促生长和软化粪便作用;在 0.6%以上时,有倾泻作用.