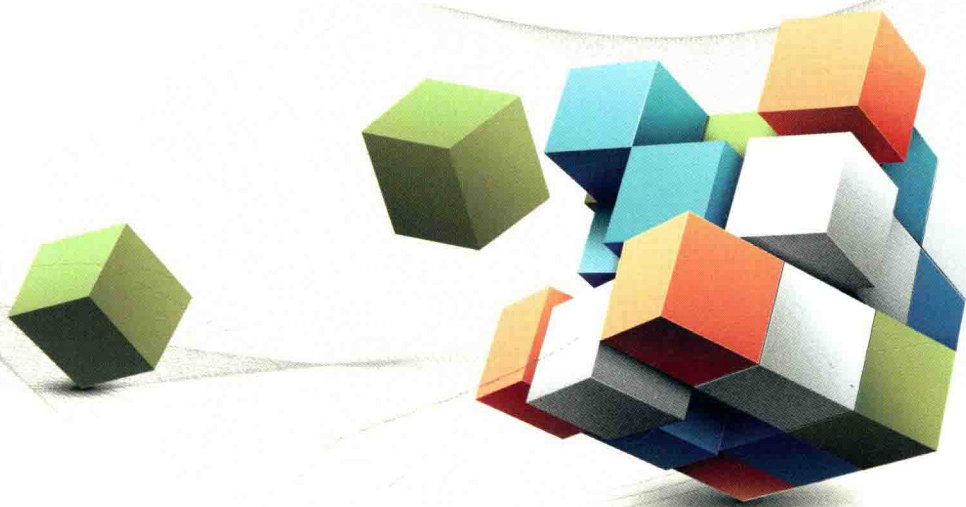




普通高等教育农业部“十二五”规划教材
全国高等农林院校“十二五”规划教材
都市型现代农业特色规划系列教材



多元统计分析 与 SAS

杜晓林 王玉民◎主编

L-39

 中国农业出版社

部“十二五”规划教材
全国高等农林院校“十二五”规划教材
都市型现代农业特色规划系列教材

多元统计分析与 SAS

Multivariate Statistical Analysis and SAS

杜晓林 王玉民 主编

中国农业出版社

图书在版编目 (CIP) 数据

多元统计分析与 SAS / 杜晓林, 王玉民主编. —北京: 中国农业出版社, 2014. 8

普通高等教育农业部“十二五”规划教材 全国高等农林院校“十二五”规划教材 都市型现代农业特色规划系列教材

ISBN 978-7-109-19072-6

I. ①多… II. ①杜… ②王… III. ①多元分析-统计分析-应用软件-高等学校-教材 IV. ①O212.4-39

中国版本图书馆 CIP 数据核字 (2014) 第 070678 号

中国农业出版社出版

(北京市朝阳区麦子店街 18 号楼)

(邮政编码 100125)

策划编辑 朱 雷 魏明龙

文字编辑 魏明龙

中国农业出版社印刷厂印刷 新华书店北京发行所发行
2014 年 8 月第 1 版 2014 年 8 月北京第 1 次印刷

开本: 787mm×1092mm 1/16 印张: 15.75

字数: 366 千字

定价: 32.00 元

(凡本版图书出现印刷、装订错误, 请向出版社发行部调换)

内 容 简 介

多元统计是多变量的综合分析方法，具有坚实的数学理论基础。本书是针对高等农林院校研究生编写的多元统计分析教材，书中除了对数学原理进行了必要的介绍和推理之外，引入了大量生物科学应用实例，极大地方便了农林专业研究中相关数据的分析和处理。同时还以国际上著名的统计分析软件 SAS 作为典型工具，通过实例介绍如何处理数据分析中的各种实际问题。

本书共十二章。第一章为绪论；第二章介绍多元正态分布的定义和性质；第三章介绍多元正态分布参数的估计和假设检验问题；第四章介绍多元数据图表示法。从第五章至第十一章介绍常用的多元统计方法，如方差分析、聚类分析、判别分析、主成分分析、因子分析、典型相关分析和多元回归分析。第十二章介绍定性资料的统计分析方法。

本书适合作为农、林、医等生物应用专业的研究生教材，也可供高等学校高年级本科生和教师作为教材或教学参考书。对于生物类和其他领域中从事应用统计的科技工作者也是一本极好的学习参考书。



都市型现代农业特色规划系列教材编审委员会

- 主任委员** 王慧敏（北京农学院院长）
邢克智（天津农学院院长）
崔英德（仲恺农业工程学院院长）
- 副主任委员** 范双喜（北京农学院副院长）
孙守钧（天津农学院副院长）
向梅梅（仲恺农业工程学院副院长）
- 编委会委员** 张喜春 李奕松 乌丽雅斯 马文芝 王立春
卢绍娟 朱立学 石玉强 洪维嘉

都市型现代农业特色规划系列教材学术委员会委员

- 主任** 范双喜
- 副主任** 孙守钧 向梅梅
- 委员**（按姓名笔画排序）
- | | | | | |
|-----|-----|-----|-----|-----|
| 马文芝 | 马吉飞 | 马晓燕 | 王厚俊 | 朱立学 |
| 乔秀亭 | 刘开启 | 刘金福 | 李 华 | 李奕松 |
| 杨逢建 | 吴宝华 | 吴锡冬 | 宋光泉 | 张喜春 |
| 陈 俐 | 周厚高 | 郭 勇 | 阎国荣 | 梁 红 |
| 潘金豹 | | | | |

编写人员名单

主 编 杜晓林 王玉民

副主编 颜亭玉 孔素然 陈晓昕
张艳芳

参 编 刘建慧 于景华 梁宏英
张俊芳 侯首萍

审 稿 徐兴忠

总 序

都市型现代农业作为一种新型的农业发展模式，自 20 世纪 90 年代进入迅速发展阶段，目前已显示出明显的经济、社会和生态效益。尽管国家间、地区间发展很不平衡，但随着人们生活水平的提高、城市人口的扩张以及资源与能源供求的集聚，都市型现代农业必将成为大城市及城郊经济社会发展的重要组成部分，其重要意义和独特优势已不同程度地显现出来。都市型现代农业要在满足不断增长的城市需求的过程中获得高效益，又要做到资源节约和环境友好，其发展必须依靠产业的融合和多学科的交叉以及现代高新技术的应用。实现都市型现代农业的高水平发展，科技是动力，人才是保证，这为都市型农业院校提出了一个既具体又有一定创新性的任务，即责无旁贷地要为都市型现代农业发展提供科技和人才支撑。长期以来，由于常规农业的发展需要和相应人才培养方案的惯性延续，使人才培养和都市型现代农业发展需求之间存在一定差异。参照国内外都市农业发展对人才种类需求的调查结果，都市型现代农业对以下三大类型人才有共同的需求。

第一种：经济功能类人才。这类人才是推动都市农业发展的关键因素，是实现各类新兴农业和涉农产业经济效益的核心。这类人才包括：懂科技、能经营、会管理的涉农企业家与经营管理人才；厚基础、复合型、多学科的科技创新人才；懂技术、高技能的技能型人才；懂科技、有经验的科技成果转化和推广人才。

第二种：生态功能类人才。建设都市农业对内强化生态功能，因此对生态环境功能有更高要求，对这类人才将有更大需求。这类人才包括：环境公益类人才、生态类人才、环境改造及创意类人才、区域规划和布局类人才、安全食品产业链监控人才等。

第三种：服务功能类人才。适应都市农业服务功能的需要，以服务带动农业产业发展。这类人才包括：旅游管理人才、物流人才（包括涉农外贸）、会展人才、农业信息技术人才等。这就要求都市型高等农业教育要更加注重都市型现代农业发展需求，适时调整教育目标和教学内容。其中，深化高校教学改革是都市型高等农业院校发展的主旨与核心，而做好高质量教材建设与创新是教学改革的重点。如何构建适应都市型现代农业发展与高校人才培养的特色教材体系是众多

都市型高等农业院校面临的现实任务，也是长期任务。

基于北京农学院、天津农学院、仲恺农业工程学院等地方高等农业院校的区位特点和办学特色，为了强化对地区经济的服务功能，逐步完善支撑都市型现代农业发展的课程体系及课程内容，2008 年天津农学院主持召开了“都市型现代农业规划系列教材”编写会议，确定了编写教材的指导思想、特色要求等内容，成立了以三院院长、分管教学的副院长、教务处长及有关专家组成的编写委员会。2009 年 9 月以北京农学院院发〔2009〕46 号、天津农学院农院政〔2009〕34 号、仲恺农业工程学院仲字〔2009〕7 号联合发布了“关于都市型现代农业特色教材建设指导性意见”，进一步明确了都市型现代农业特色规划系列教材的定位、遴选原则、组织领导、出版使用等方面的要求。在系列教材编写过程中，三院多次组织、邀请各参编高校开展特色教材编写研讨会，并聘请各高校同行专家对教材初稿进行全面审阅，共同商榷，认真修改，集思广益，确保教材的高质量出炉。同时也陆续得到了更多兄弟院校的支持，并纷纷加盟。

都市型现代农业特色规划系列教材的编写注重都市农业特点，注重人才培养目标领域的拓宽，注重把“教材”向“学材”转变，注重教材内容实用性的优化，重点强调以下几方面的特色：注重学科发展的大背景，拓宽理论基础和专业知识，着眼于理论联系实际与可应用性，突出创新意识；体现都市型现代农业发展的特征；借鉴国内外最新的资料，融合当前学科的最新理论和实践经验，用最新知识充实教材内容；在结构和内容的编排上更注重能力培养，强化自我学习能力、思维能力、解决问题能力；强化可读性，教材中尽量增加图表内容，将深奥的理论通俗化，图文并茂。

感谢参加本系列教材编写和审稿的各位教师所付出的大量卓有成效的辛勤劳动。由于编写时间紧、相互协调难度大等原因，本系列教材还存在一些不足。我们相信，该批特色规划系列教材的编写作为都市型高等农业院校教学改革的重要环节，将会为培养 21 世纪现代农业高等人才提供重要保障，对都市型现代农业多功能的充分发挥和更好地服务于大都市和农村将具有重要的推动作用。在各位老师和同行专家的努力下，本系列教材一定会不断地完善，在我国都市型高等农业院校专业教学改革和课程体系建设中定能发挥出应有的作用。

都市型现代农业特色规划系列教材编审委员会

2013 年 7 月

前 言

多元统计分析是运用数理统计方法研究多变量问题的理论和方法，是一元统计学的推广。自 20 世纪 70 年代起 40 余年来，我国不仅在多元统计分析相关理论的研究上取得了长足进展，也在将其理论和方法引入生物、农业、经济和管理等多个学科领域的应用方面取得了丰硕的成果，有些研究成果已达到了国际先进水平。

农业科学研究是在周密设计的前提下，通过实验和调查获得实验数据，运用数学方法整理分析，把实验数据转化为研究结果，从而揭示事物的内在规律。而多元统计分析是一种综合分析方法，它能够在多个对象和多个指标相互关联的情况下分析它们的统计规律，非常适合农业科学研究。

根据现代农业科学的需要，高等农林院校普遍开设研究生多元统计课程。本书编者多年从事研究生多元统计分析课程的教学工作，深刻体会到一本适用于农林院校学生的多元统计分析教材对于提升他们科研水平和能力的重要性，故谋划多年组织编撰，今日终偿此愿。本书可作为农科各专业研究生多元统计分析课程的教材，也可作为其他对多元统计分析方法感兴趣的非数学专业本科生、研究生和科研人员的参考用书，为他们提供一个易于理解、掌握和操作应用的工具。

本书力求突出以下特点：

1. 明显不同于纯理工类教材，强调从一元统计到多元统计的过渡，增加了一元统计分析的基础知识，深入浅出地介绍多元统计分析的基本思想和主要内容。

2. 增加了多元统计分析所必需(包括线性代数和概率论方面)的基本知识。

3. 重点介绍农业科学研究中常用的多元方差分析和多元回归分析的内容，减少或删除了不常用的内容。

4. 强化实例应用，弱化理论推导，着重介绍各种方法的实际背景和统计思想，增加涉农的实际案例。

5. 本书在各章最后一节详细介绍了各种方法的 SAS 统计软件操作方法，可以帮助读者很快掌握多元统计方法的实际操作使用。

全书共分 12 章。第一章绪论介绍多元统计分析的主要内容和学习这门课程的数学基础知识；第二章介绍多元正态分布的定义和性质；第三章介绍多元正态分布参数的估计和假设检验问题；第四章介绍多元数据图表示法。从第五章至第十一章介绍常用的多元统计方法，如方差分析、聚类分析、判别分析、主成分分析、因子分析、典型相关分析和多元回归分析。第十二章介绍定性资料的统计分析方法。

北京理工大学徐兴忠教授审阅了本书全稿，并提出了许多中肯的修改意见，在此表示衷心感谢。

由于编者水平所限，疏漏和不妥在所难免，欢迎读者批评指正。

编 者

2014 年 1 月

目 录

总序
前言

第一章 绪论	1
1.1 从一元统计分析到多元统计分析	1
1.1.1 一元统计分析的研究对象	1
1.1.2 多元统计分析的研究对象	1
1.2 多元分析的数学基础	2
1.2.1 线性代数中的有关概念、定理	2
1.2.2 数理统计中的有关概念、定理	4
1.3 多元分析的主要内容	7
1.3.1 多元统计分析方法的主要内容	7
1.3.2 多元统计分析方法在农业统计中的具体应用	9
1.3.3 多元统计分析方法在农业统计中应用的一般步骤	10
习题一	11
第二章 多元正态分布	12
2.1 一元正态到多元正态的扩展	12
2.1.1 一维随机变量和多维随机向量的基本概念	12
2.1.2 多维随机向量的数字特征	14
2.1.3 多元正态分布	16
2.2 多元正态分布的基本性质	17
2.3 SAS 简介与基本操作	17
2.3.1 SAS 统计软件简介	17
2.3.2 SAS 统计软件的特点	18
2.3.3 SAS 统计软件的基本操作	18
习题二	29
第三章 多元正态总体的统计推断	31
3.1 多元正态分布的参数估计	31
3.1.1 多元样本的基本概念和常见统计量	31
3.1.2 多元正态总体 μ 和 Σ 的最大似然估计	32
3.2 均值向量的检验	33

3.2.1 多元正态总体的常见抽样分布	33
3.2.2 一个总体均值向量 $\mu=\mu_0$ 的假设检验	37
3.2.3 两个总体均值向量 $\mu_1=\mu_2$ 的假设检验	39
3.3 协差阵的检验	41
3.3.1 似然比准则的一般原理	41
3.3.2 一个总体协差阵 $\Sigma=\Sigma_0$ 的假设检验	42
3.3.3 多个总体协差阵相等的假设检验	43
3.4 统计推断的 SAS 计算	45
3.4.1 两样本均值的区间估计和 t 检验	45
3.4.2 正态性检验	46
习题三	47
第四章 多元数据图表示法	50
4.1 轮廓图	50
4.1.1 轮廓图的概念	50
4.1.2 轮廓图的做法与例子	50
4.2 雷达图	51
4.3 调和曲线图	52
4.3.1 调和曲线图的概念	52
4.3.2 调和曲线图的作法与例子	52
4.4 脸谱图	53
4.5 SAS 作图法	54
4.5.1 一维数据的图形	54
4.5.2 多维数据的图形	64
习题四	66
第五章 多元方差分析	68
5.1 一元方差分析到多元方差分析	68
5.1.1 一元方差分析的基本概念和基本假设	68
5.1.2 单因素一元方差分析	69
5.1.3 无交互作用的双因素一元方差分析	73
5.1.4 有交互作用的双因素一元方差分析	77
5.1.5 多元方差分析	80
5.2 多元方差分析的数学基础	81
5.2.1 单因素多元方差分析	81
5.2.2 无重复的双因素多元方差分析	86
5.2.3 等重复的双因素多元方差分析	90
5.3 多元方差分析的 SAS 计算	95
5.3.1 单因素方差分析	95

5.3.2 双因素方差分析	100
5.3.3 多元方差分析	103
习题五	105
第六章 聚类分析	111
6.1 聚类分析原理	111
6.2 距离和相似系数	111
6.2.1 指标(变量)测量尺度的类型	111
6.2.2 样品或指标的亲疏程度的度量	112
6.3 基本系统聚类方法	114
6.4 基本性质	120
6.4.1 系统聚类法的基本性质	120
6.4.2 聚类分析的主要步骤	120
6.5 聚类分析的 SAS 计算	121
习题六	123
第七章 判别分析	127
7.1 判别分析基本原理	127
7.2 距离判别法	127
7.2.1 两个总体的距离判别法	128
7.2.2 多个总体的距离判别法	130
7.3 费舍尔(Fisher)判别法	135
7.3.1 不等协差阵的两总体费舍尔判别法	135
7.3.2 多个总体的费舍尔判别法	136
7.4 贝叶斯(Bayes)判别法	138
7.4.1 基本思想	139
7.4.2 多元正态总体的贝叶斯判别法	139
7.5 逐步判别分析	141
7.5.1 基本思想	141
7.5.2 引入和剔除变量所用的检验统计量	141
7.5.3 具体计算步骤	142
7.6 判别分析的 SAS 计算	146
习题七	149
第八章 主成分分析	152
8.1 主成分分析基本原理	152
8.1.1 数学模型	152
8.1.2 主成分的几何意义	152
8.2 基本性质	153

8.2.1 主成分的推导	154
8.2.2 主成分的性质	156
8.3 计算步骤及实例	156
8.3.1 基于协方差矩阵(Q 型)	156
8.3.2 基于相关系数矩阵(R 型)	157
8.4 主成分分析的 SAS 计算	158
8.4.1 计算主成分分析的 SAS 数据	158
8.4.2 采用 PRINCOMP 对话框计算主成分	160
8.4.3 输出结果与分析	162
习题八	163
第九章 因子分析	164
9.1 因子分析的基本原理及数学模型	164
9.1.1 因子分析的基本原理	164
9.1.2 因子分析模型	165
9.2 因子载荷阵的估计方法	167
9.3 因子旋转	168
9.3.1 因子旋转的目的	168
9.3.2 因子旋转方法	170
9.4 因子得分	172
9.5 因子分析的 SAS 计算	176
习题九	179
第十章 典型相关分析	181
10.1 典型相关分析的基本原理及数学模型	181
10.1.1 典型相关分析的基本原理	181
10.1.2 典型相关分析的数学模型	181
10.2 总体典型相关系数和典型变量	182
10.2.1 典型相关系数	182
10.2.2 典型变量的性质	183
10.3 样本典型相关系数和典型变量	183
10.4 典型相关的显著性检验	185
10.5 典型相关分析的步骤及实例	186
10.6 典型相关分析的 SAS 计算	189
10.6.1 计算典型相关的 SAS 数据	189
10.6.2 采用 CANCORR 过程计算典型相关	190
10.6.3 采用 CANCORR 对话框计算典型相关	190
10.6.4 输出结果与分析	192
习题十	193

第十一章 多重多元回归分析	195
11.1 从一元回归到多元回归的过渡	195
11.1.1 一元线性回归	195
11.1.2 可化为一元线性回归的情形	200
11.1.3 多元线性回归	202
11.2 双重筛选逐步回归分析	204
11.2.1 逐步回归方法	204
11.2.2 双重筛选逐步回归	206
11.3 多元回归分析的 SAS 计算	210
习题十一	220
第十二章 定性资料的统计分析	223
12.1 定性变量数量化方法	223
12.1.1 定性变量的概念	223
12.1.2 定性变量的数量化	223
12.2 列联表及数据分析	224
12.2.1 列联表的概念	224
12.2.2 列联表的数据分析	225
12.3 对数线性模型	226
12.4 Logistic 回归	229
12.4.1 Logit 变换	229
12.4.2 Logistic 回归模型	230
12.4.3 实例	231
12.5 列联表的 SAS 计算	232
习题十二	234
参考文献	235

第一章 绪 论

1.1 从一元统计分析到多元统计分析

1.1.1 一元统计分析的研究对象

随着研究随机现象规律性的科学——概率论的发展,应用概率论的结果更深入地分析研究统计资料,通过对某些现象的观察来发现该现象的内在规律性,并做出一定精确程度的判断和预测;将这些研究的某些结果加以归纳整理,逐步形成一定的数学模型,这些组成了数理统计的内容.

数理统计在自然科学、工程技术、管理科学及人文社会科学中得到越来越广泛和深刻的应用,其研究的内容也随着科学技术和政治、经济与社会不断发展而逐步扩大.但按农学专业的总体设计,我们的数理统计课程只讨论统计推断.数理统计是以概率论为基础,根据试验或观察得到的数据,来研究随机现象统计规律性的学科.目的是让学生了解统计推断检验等方法,并能够应用这些方法对研究对象的客观规律性做出各种合理的估计和判断.掌握总体参数的点估计和区间估计;掌握假设检验的基本方法与技巧;理解方差分析及回归分析的原理,并能运用其方法和技巧进行统计推断.

1.1.2 多元统计分析的研究对象

多元统计分析(简称多元分析)是统计学的一个重要分支.它是应用数理统计学来研究多变量(多指标)问题的理论和方法,它是一元统计学的推广和发展.

多元统计分析是一门具有很强应用性的课程,它在自然科学和社会科学等各个领域中都有广泛的应用,它包括了很多非常有用的数据处理方法.在实际问题中,很多随机现象涉及的变量不止一个,而经常是多个变量,而且这些变量间又存在一定的联系.我们常常需要处理多个变量的观测数据.

由于大量实际问题都涉及多个变量,这些变量又是随机变化的,所以要讨论多维随机向量的统计规律性.多元统计分析就是讨论多维随机向量的理论和统计方法的总称.研究的内容既包括一元统计学中某些方法的直接推广,又包括多个随机变量特有的一些问题.

例 1.1.1 有 n 个不同地区,每个地区记录多种农作物的收获量,用多元统计方法对各个地区的总生产效率进行比较,并对不同的农业区域进行分类.

例 1.1.2 为了节省能源,对某地农用的手扶拖拉机的能源消耗进行抽样调查.调查的内容为拖拉机在田间进行运输、排灌、加工等作业时的燃油耗,在册月数、年平均更变零件数及平均燃油耗.通过对调查资料做统计分析,达到对拖拉机的平均燃油耗做预测并对拖拉机进行分类,划分淘汰类、大修类、小修类和继续使用类的目的.

1.2 多元分析的数学基础

1.2.1 线性代数中的有关概念、定理

1.2.1.1 有关矩阵的概念

定义 1.2.1 将 $m \times n$ 个数 $a_{ij} (i=1, 2, \dots, m; j=1, 2, \dots, n)$ 排成的一个矩形数表

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix},$$

称为一个 m 行 n 列矩阵(Matrix), 简称为 $m \times n$ 矩阵. 其中横向各排称为行, 纵向各排称为列, $m \times n$ 个数叫作矩阵 $\mathbf{A}$ 的元, a_{ij} 叫作矩阵 $\mathbf{A}$ 的 (i, j) 元 ((i, j) entry).

如果矩阵 $\mathbf{A} = (a_{ij})$ 的行数与列数都等于 n , 则称 $\mathbf{A}$ 为 n 阶矩阵(或称 n 阶方阵).

主对角线以外的元全为零的 n 阶方阵称为对角线矩阵(Diagonal Matrix), 简称对角阵,

$$\mathbf{A} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

主对角线上的元都为 1 的 n 阶对角阵称为 n 阶单位矩阵(Identity Matrix),

$$\mathbf{I}_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}.$$

定义 1.2.2 如果两个矩阵 $\mathbf{A}$, $\mathbf{B}$ 有相同的行数与相同的列数, 并且对应位置的元均相等, 则称矩阵 $\mathbf{A}$ 与矩阵 $\mathbf{B}$ 相等, 记为 $\mathbf{A} = \mathbf{B}$.

定义 1.2.3 设有两个 $m \times n$ 矩阵 $\mathbf{A} = (a_{ij})_{m \times n}$, $\mathbf{B} = (b_{ij})_{m \times n}$, 称

$$\mathbf{A} + \mathbf{B} = \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2n} + b_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \cdots & a_{mn} + b_{mn} \end{pmatrix}$$

为 $\mathbf{A}$ 与 $\mathbf{B}$ 的和, 记作 $\mathbf{A} + \mathbf{B}$.

定义 1.2.4 数 λ 与矩阵 $\mathbf{A}$ 的乘积为

$$\lambda \mathbf{A} = \mathbf{A} \lambda = \begin{pmatrix} \lambda a_{11} & \lambda a_{12} & \cdots & \lambda a_{1n} \\ \lambda a_{21} & \lambda a_{22} & \cdots & \lambda a_{2n} \\ \vdots & \vdots & & \vdots \\ \lambda a_{m1} & \lambda a_{m2} & \cdots & \lambda a_{mn} \end{pmatrix},$$

简称数乘.

定义 1.2.5 设 $\mathbf{A} = (a_{ij})$ 是一个 $m \times s$ 矩阵, $\mathbf{B} = (b_{ij})$ 是一个 $s \times n$ 矩阵, 那么规定矩阵 $\mathbf{A}$