统计数据分析与应用丛书

基于MINITAB的现代实用统计

马逢时 吴诚鸥 蔡 霞 编著

中国人民大学出版社

统计数据分析与应用丛书

基于MINITAB的现代实用统计

马逢时　吴诚鸥　蔡　霞　编著

中国人民大学出版社
·北京·

图书在版编目（CIP）数据

基于 MINITAB 的现代实用统计/马逢时等编著.
北京：中国人民大学出版社，2009
（统计数据分析与应用丛书）
ISBN 978-7-300-10394-5

Ⅰ. 基…
Ⅱ. 马…
Ⅲ. 统计分析-应用软件，MINITAB
Ⅳ. O213

中国版本图书馆 CIP 数据核字（2009）第 028934 号

统计数据分析与应用丛书
基于 MINITAB 的现代实用统计
马逢时　吴诚鸥　蔡　霞　编著

出版发行	中国人民大学出版社		
社　　址	北京中关村大街 31 号	**邮政编码**	100080
电　　话	010－62511242（总编室）	010－62511398（质管部）	
	010－82501766（邮购部）	010－62514148（门市部）	
	010－62515195（发行公司）	010－62515275（盗版举报）	
网　　址	http://www.crup.com.cn		
	http://www.ttrnet.com(人大教研网)		
经　　销	新华书店		
印　　刷	北京鑫丰华彩印有限公司		
规　　格	185 mm×260 mm　16 开本	**版　　次**	2009 年 4 月第 1 版
印　　张	29.25 插页 1	**印　　次**	2009 年 4 月第 1 次印刷
字　　数	688 000	**定　　价**	58.00 元

序言

Preface

随着生产和科学技术的发展，各式各样受随机干扰的数据呈现在人们面前，需要进行处理，从中提取有用信息，并寻找隐藏在随机性背后的统计规律，以达到解决实际问题的目的。在这一过程中必然要使用各种统计方法，因此学习、使用统计学已在中国形成一种时尚，各类统计书籍的编写、出版和销售也都随之兴旺起来。马逢时教授等三人编著的《基于MINITAB的现代实用统计》一书完成了，我很愿意在此书出版之际谈谈自己的看法。

这本书包含三个方面的内容：多元统计分析、可靠性与生存分析及时间序列分析。这三方面知识在实际工作中有着非常广泛的应用，其统计思想丰富，统计方法涉及面广，理论及计算都相当复杂。不论初学者还是专业人士使用这三类方法都很不容易，对缺乏高深统计基础的广大工程技术人员来说就更困难了，加上解决这类问题时一定会涉及大量的数据，因此没有统计软件是很难完成计算与分析的。广大读者迫切需要读到这样一本书，一眼就能看穿这些学科中所包含的统计问题的实际含义，当自己遇见实际问题时能知道此问题属于哪种类型，怎样使用计算机软件去计算，看到计算结果后能准确理解计算结果的含义。总之，读者需要这样的参考书，它是你身边的最好帮手，会帮助你少走弯路，较快达到你的要求。据我的了解，在国内将这些相关内容用本书这样的方式编写出版甚为罕见，尚属首次。它的顺利出版的确是因时代需要应运而生的。

这本书是特别为我国广大工程技术人员和社会科学工作者而编写的。它告诉读者如何从数据出发，寻找适当的统计方法，使用相应的统计软件，解释计算机输出的结果，指明进一步研究的新途径。本书不是从理论到理论，而是要把统计方法变为生产力，为创造财富服务，是将统计的学习应用到实践的好书。

这本书起点较低，用很少的篇幅介绍一些必要的概念与符号，随后就转入对一连串实际问题的研究和对数据处理方法的探讨，在计算机软件的对话框中将问题逐步引向深入。因此，这本书很适合广大工程技术人员阅读，也是入门的好向导。以第2篇为例，该篇先用很少

的篇幅介绍可靠性的基本概念，然后就转入对一连串问题的研究。第一个问题就是寿命分布的识别，这是可靠性研究中首当其冲的问题，利用 MINITAB 软件可以很快地识别一个样本来自 11 个常用寿命分布中的哪一个。假如书中所列的 11 个分布都不合适，那就要用非参数方法处理；若有一个寿命分布适合，就可选用各种参数方法。因为这些常用寿命的参数方法较为成熟，所以将其他寿命分布归于非参数分布族内并不过狭。这一部分的编写在国内外可靠性书籍中都很少涉及，可是本书却对这个问题作了大量陈述并提供了软件使用指导，这是很好的，很难得的。

本书案例丰富。每类实际问题都至少配有一个含有实际数据的案例，详细说明这些数据的类型和操作程序，直到结果输出，并对输出也作出详尽的说明和解释。如果读者对于前面的相关内容尚有一些不明之处，通过案例也都能弄明白。如果需要了解更严谨的理论分析及公式细节，还可以参见网上资源。这些富于启发性的案例对学习和使用统计方法起到了指导的作用，本书能做到这一点是难能可贵的。

一本书不可能解决所有问题，本书末尾所列的参考文献正是读者进一步深入研究需要阅读的指南。

本书特别适合广大工程技术人员和大学生、研究生阅读，在学科内容方面，适用于非统计专业的各理工科工程技术、医药卫生、生命科学、环境工程、管理、经济、体育、考古、教育和各社会科学领域。本书既可供上述领域科技工作者学习使用，也可作为上述各领域本科生和研究生的教学参考书。

我相信，本书的出版一定会受到广大读者的热烈欢迎。它的出版将会帮助广大工程技术人员掌握统计方法，并把统计方法变为生产力，在推动我国生产发展和科学技术进步方面起到积极、显著的作用。我预祝他们获得更大成功！

茆诗松*

于上海华东师范大学

---

* 上海华东师范大学金融统计学院终身教授，博士生导师，上海现场统计研究会理事长，上海质量科学研究院终身研究员，中国质量协会六西格玛管理推进工作委员会专家委员会顾问。曾任中国概率统计学会副理事长。

前言
Preface

目前我国正处于经济蓬勃发展的时期，现代化科学技术的发展也进入了知识爆炸年代。在各项科学理论及技术开发的研究工作中，新知识、新概念、新方法不断产生，这些都给各项事业的发展提供了基础。而新成果的涌现很多都是面对大量数据，利用统计方法予以分析而获得有意义的信息的结果，因此可以说，应用统计学是各项实际工作中必不可少的支撑工具。特别是近几十年来，与计算机密切结合的统计学发展更加快速，统计学的应用领域越来越广泛和深入，使用的人也越来越多。但是对于非统计专业出身的广大科技、医药卫生、管理、经济和社会科学工作者来说，理解统计思想并掌握现代统计计算工具并不是一件易事。目前在我国已经出版了大量有关统计学的书籍，其中不少是为这些非统计专业人士而编写的，受到普遍欢迎，这说明广大读者对于学习应用统计的迫切需要。

一般介绍应用统计的书籍（例如马逢时、周暐、刘传冰编著的《六西格玛管理统计指南——MINITAB使用指导》等），通常只讨论在实际工作中遇到的最简单的情况，即只考虑单个变量，所有观测数据是相互独立的，大多数情况下它们来自正态分布，等等。但仔细观察可以发现，现实问题是复杂的，有时需要同时考虑多个变量，有时数据间不一定是相互独立的，分布可以多种多样。很多人会认为，在多个变量问题中，如果对每个变量都分析清楚，那么多个变量问题自然就可以分析清楚了。可事实并不是这样。我们不能将多个变量简单地分别加以分析讨论，例如，获得的连续300天的温度数据以及200天的股市行情不能当作相互独立的数据用普通的统计分析方法来处理，因为各个数据间明显是有相关性的。此外，在实际工作中为了进行可靠性测试，常常会遇到各种“删失”数据（例如，只知某零件的失效是在500～600小时发生的，并不能得到准确的数据记录），寿命为Weibull分布等非正态的情况，处理这些数据时，不能忽视其特点而只使用一般的统计方法。本书就是分三篇分别讨论这三大类更复杂的问题。

第1篇多元统计分析介绍的是多变量数据分析方法。除了与单个

变量统计分析相平行的单总体、多总体的参数估计、假设检验问题之外，还介绍了多变量所特有的判别分析、聚类分析、主成分分析、因子分析和对应分析等方法，用这些方法可以得到很多很有价值的结论。第 2 篇可靠性与生存分析介绍的是工程技术领域中的可靠性问题及生命科学中的生存分析问题所用的数据统计分析方法，适用于多种常见的寿命分布及多种类型的删失数据，包括参数方法和非参数方法。第 3 篇时间序列分析介绍的是对不能看成独立数据的时间序列的趋势分析、平滑方法以及目前应用最广泛的 ARIMA 模型，它不但能提供对实际生活中常见的各种类型时间序列的规律性认识，而且可以提供预测结果，这些在实际工作中都是非常有用的。

本书系统地介绍了上述三个领域中涉及的有关统计学背景知识及其统计思想，把各方面统计内容的介绍与计算机 MINITAB 软件的操作使用结合起来，使广大读者在学完本书后，对于上述三个领域常用的统计学内容有更深入的理解与认识，能更广泛、更方便地使用。这三个领域的统计内容几乎是相互无关的，学习时并没有固定的先后顺序要求。

本书的编著者将此书的特点定位于下述几个方面：

(1) 统计方法的应用指南。与一般的统计教科书不同，本书大大增加了实际应用方法的介绍。从统计学科内容来说，这三个领域中所使用的统计和数学工具比单变量统计复杂得多，理解起来也困难得多。介绍多元统计分析、可靠性与生存分析及时间序列分析的统计书籍有很多，但广大工程技术人员和大学生很少能读懂，更不知如何使用。本书并不强调公式与理论的推导（最多给出公式），但要求学会理解统计思想和基本方法并使用统计工具，结合计算机软件 MINITAB 直接解决具体问题。本书的阅读对象是一般工程技术人员和非统计专业的大学生，起点低，易学习，只要求读者有大专水平，具备简单的应用统计基础。

(2) 统计软件 MINITAB 操作的实际应用的全程指导。与一般的软件使用说明书或软件附带的帮助文件不同，本书不只讲操作，而是特别强调通过统计方法的背景和统计思想、统计概念的介绍，说明有关计算的统计含义。如果实际问题同时有多种统计方法可以处理，本书还会介绍各种方法的共同点及其差异，使读者明白如何进行统计方法的选择。因此，本书一方面强调，在理解统计概念的基础上运用计算机软件最终获得实际问题的具体分析结果，避免纸上谈兵；另一方面强调，对于分析结果要有比较深入的统计解释，要在理论上达到足够的高度和深度，避免只讲操作，不讲道理。

(3) 特别强调在学习统计知识方面的可读性。在介绍统计知识时，本书尽量避免使用专业的数学语言，努力做到叙述通俗化，例题生动、具体、多样，稍微深入的或不常见的内容都给予解释，从而使一般的工程技术人员和非统计专业大学生都能看得懂、学得会。本书力求打破一般读者对于统计学特别是高深统计学的神秘感、惧怕心，把多元统计分析、可靠性与生存分析以及时间序列分析这些复杂的统计内容变得易于理解掌握，让读者在学习后能顺利地将这些知识应用于各自的领域。

(4) 强调统计方法应用的广泛性和实用性。多元统计分析、可靠性与生存分析以及时间序列分析这三类统计方法的应用范围非常广，不但在工程技术领域有深入的应用，在医药卫生、生命科学、环境工程、管理、经济、体育、艺术、考古、股市、保险、教育、社会工作等领域也都有出色应用的范例。本书在例题的选择上力求有更广泛的代表性，使读

者容易举一反三。不论你所遇到的问题属于哪个领域，都可以从本书所介绍的例题中找到可借鉴的思路，通过自己的学习探索实践，最终能解决相应问题。

(5) 使用的是MINITAB最新版全中文软件。任何与统计有关的工作离开计算机都是不可能的。目前全世界通用的统计软件不下百种，但MINITAB是在各高等院校及工程技术人员中使用最多的最普及的统计软件。统计学应用的现代化当然离不开统计软件的现代化。2007年1月MINITAB软件公司推出的最新的R15版，不但增加了功能，而且增加了简体中文版界面。中文版界面的出现将使广大的中国读者更易于接受、使用，有了问题也能更容易地查到相关的详细的中文帮助信息，这将大大推动应用统计及计算机软件在中国的普及应用。本书力图使统计的应用达到现代化的水平，例如，本书使用的最新的R15版软件中，包含有多元质量控制图、用Logistic回归进行判别分析、主成分回归等一些新成果，还增补了该软件中尚未解决的一些重要问题的宏指令。我们在全部使用中文版操作说明的同时，在说明之后用括号注明英文原文的内容，这对于使用英文原版或早期其他版本的读者也很方便。当然，任何软件都会不断更新，但这些基本模式会适用相当长的时期。限于篇幅，本书在输出结果的界面内只给出了中文输出结果，使用英文原版界面的读者对照两种输出不难理解其含义。为了便于两种文本的对照，在本书的网上资源中还包括带计算机检索功能的英汉、汉英统计词汇对照表。

总之，本书力图使统计学的应用和普及达到实用程度，实现现代化水平。

为了充分利用网上资源并便于不同水平、不同需求读者的学习使用，本书对于各项内容做了细致安排：基本内容及全部操作都列在书中，一般读者能够顺利阅读及使用；全部数据文件及新增宏指令列为网上资源，可以免费下载供所有读者学习时使用，这些数据文件在给出原始数据的同时还列出了一些关键的计算结果供大家参考；所有例题的全部原始数据也列在网上资源中相同编号的例题中；如果读者对于某些章节内容希望了解更详细的推导或理解更深入的统计知识，可以参考网上资源中关于理论知识和计算方法的介绍；如果初学者希望对于MINITAB的操作有更详细的界面显示，也可以参考网上资源中相应的界面图示；如果需要英汉或汉英统计名词的转换查询或需要对照查看统计用表等，也可以在网上资源中找到相应文件。网上资源中的所有章、节、例、公式、图表等编号皆与正文保持相同（因而不一定连续），其特有部分内容需要编号时才另给编号。读者可以登录中国人民大学出版社经管在线（www.rdjg.com.cn）或前言中提供的电子信箱地址免费下载数据文件及全部网上资源。

本书的编著历时三年，五易其稿。2006年初，马逢时在主持MINITAB软件R15的中文翻译工作时，就组织了天津大学数学系研究生贺广婷、蔡霞及辛凌雯开始编写本书的草稿。2007年1月MINITAB软件R15正式发布后，为适应读者日益提高的需求，编著者决定大幅度增加实例及统计解释，重新撰写本书初稿。其中，吴诚鸥负责第1篇；蔡霞负责第2篇；马逢时负责第3篇。初稿完成后，大家互相讨论，反复修改了多次，最后由马逢时负责完成全书的修改稿。修改稿的审定分别邀请国内各领域权威专家完成，其中第1篇审稿人是云南大学王学仁教授，第2篇审稿人是华东师范大学茆诗松教授，第3篇审稿人是北京大学程乾生教授。根据审稿者的宝贵意见，编著者再次进行了修改。为更充分利用现代网络技术手段以满足不同读者的要求，在出版前再次修订调整，安排了网上资源

部分，这使得本书能以更加精练易读的面貌出现在读者面前。在最后阶段，金海木参加了全书的校核工作。

本书的编写得到了中国质量协会全国六西格玛管理推进工作委员会的关怀与支持，得到了 MINITAB 软件公司及其在中国的上海泰珂玛信息技术有限公司的支持，本书的草稿提供者、校核者以及三位专家为书稿的出版花费了很多心血，编著者在此对他们表示诚挚谢意。

虽然编著者尽了自己的最大努力，但限于各方面的条件和水平，难免会有错误或疏漏，恳请读者批评指正。

编著者联系电子信箱：fengshima@vip.sina.com，chengou_wu@163.com 和 caixiatju@126.com，数据文件及网上资源存放地址：fengshima@163.com，登录密码为：statistics。

提供支持的网站：www.caq.org.cn，www.techmax.com.cn，www.6sq.net。

编著者

CONTENTS 目录

第 1 篇　多元统计分析

第 2 篇 可靠性与生存分析

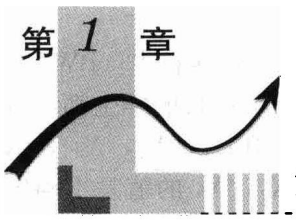
第3篇 时间序列分析

第1篇

多元统计分析

在一般的应用统计教材中，我们可以学习很多统计方法，但这些统计方法仅限于单个随机变量的情况，统称为**一元统计分析**。而我们在实际问题中通常会同时涉及多个随机变量，例如生产一颗螺钉，要同时兼顾螺纹长度、直径、深度等多个指标，一元统计分析却只能取单个指标进行研究。很多人认为，如果分别对每个指标都研究清楚，不就等于把整个情况研究清楚了吗？事实并不是这样。例如，新大学生报到时，一位男学生的身高为 182 厘米，这不值得奇怪，一位男学生的体重仅为 55 千克，这也不值得奇怪，因为在这两项指标上，这些数值虽然与平均值有些偏离，但幅度并不大。但如果同一个人身高为 182 厘米，体重为 55 千克，则这样又高又瘦的人会让人们十分奇怪。这就说明，我们在考虑问题时，如果把多个变量拆分为多个一元变量分别研究分析，常常不能对全局得出准确结论。这种情况之所以发生，是因为多个变量间通常是相关的，必须将所有变量看成一个整体才行。正因为如此，多元问题的统计分析不能被多个单独变量的研究所替代，这就需要通过多元统计分析来同时研究多个随机变量，多元统计分析是一门非常重要又有很多应用的学科。

本书的第 1 篇就讨论多元统计分析的一些应用课题，共包含 6 章。其中第 1 章介绍多元正态分布；第 2 章介绍判别分析；第 3 章介绍聚类分析；第 4 章介绍主成分分析；第 5 章介绍因子分析；第 6 章介绍对应分析。在一元统计中，正态分布具有最重要的地位。对于多元分布也一样，在各种多元分布中，多元正态分布的地位是最重要的。我们将在第 1 章中介绍多元正态分布以及有关的统计分析。由于整个多元统计分析要用到很多关于矩阵的概念及运算，我们将有关矩阵的知识列入网上资源第 1 篇附录，对于矩阵知识或用 MINITAB 软件实施矩阵运算操作尚不熟悉的读者最好先阅读此附录。



多元正态分布及其统计分析

在多元统计分析中，多元正态分布占有相当重要的地位，它是多元统计分析的基础。实际上，许多应用问题涉及的随机向量服从正态分布或近似服从正态分布，或者虽然本身不服从正态分布，但是它的样本均值近似服从正态分布，因此我们以研究多元正态分布为主要对象。此外，对于多元正态分布，其理论与实践都比较成熟，已有一整套行之有效的统计推断方法；而对于不服从多元正态分布的随机向量，其研究往往需要用到专门的数学理论，限于篇幅本书不作论述。我们在介绍多元统计分析的具体方法之前，首先介绍多元正态分布的概念和性质。本章讨论多元正态分布的概念及其统计分析。其中，1.1 节介绍多元正态分布的概念及其参数估计；1.2 节介绍多元正态总体的参数检验，包括单个多元正态总体及双多元正态总体的均值向量的检验，单个及多个多元正态总体协方差阵的检验；1.3 节介绍多元方差分析（MANOVA）；1.4 节介绍多元质量控制图；1.5 节介绍多元正态随机数的产生方法。从上述提要中大家可以看到，在第 1 章中所讨论的实际上包含了与整个一元统计分析相平行的全部内容。在 1.1 节中，由 1.1.3.2 节至 1.1.4.3 节的内容是为进行理论分析而准备的，有关内容都要运用相当深的统计知识才能完全理解，初学者可以跳过这些段落，直接从 1.2 节的具体应用部分开始学习。

1.1 多元正态分布的概念及其参数估计

多元正态分布是我们在生产实践及科学研究中最常见的一类随机向量的分布。

1.1.1 随机向量

在一元统计分析中，我们首先引入了随机变量的概念。在多元统计分析中，我们要同时考虑多个随机变量，这就是随机向量。

1.1.1.1 随机向量的定义

假设有一个焊接技术培训项目，这个项目提供了三门课程：基础焊接技术（BWT），焊接技术提高（AWT）和焊接车间实践（PWW）。每个参加这个培训项目的学生都学习这三门课程，并在课程结束时得到一个 0～100 之间的分数。分别用 X_1，X_2，X_3 表示 BWT，AWT 和 PWW 这三门课程的分数。未参加考试时，X_1，X_2，X_3 的值是不确定的，它们都是随机变量。把每个学生参加这三门课程将要得到的分数排列成一个列向量的形式：

$$X=\begin{pmatrix}X_1\\X_2\\X_3\end{pmatrix}=(X_1,X_2,X_3)'$$

这个列向量记为 X，X 就是一个随机向量。写成行向量转置的形式 $(X_1, X_2, X_3)'$是为了书写方便，以后常常使用这种写法。

在这个例子中，如果一个学生在基础焊接技术（BWT）课程中得到了很高的分数，那么他很可能也在另外两门课程中得到较高的分数，但一门课的成绩又不能完全决定另一门课的成绩：可能有的学员理论成绩好，而动手能力不强，只有把三门课的成绩合在一起才能全面反映学员的理论和动手能力。由此可见，把三门课的分数合在一起作为随机向量来研究是很必要的。

一般地，为了同时考虑 p 个随机变量 X_1，X_2，…，X_p，把这 p 个随机变量放到一起得到列向量 $X=(X_1, X_2, \cdots, X_p)'$，则称 X 为一个 p 维**随机向量**。

1.1.1.2 随机向量的样本（样本资料阵）

和一元统计一样，多元统计也有总体和样本的概念。例如当你一天生产了许多螺钉时，全体螺钉的三个指标：长度、直径、深度称为总体，我们把总体抽象为随机向量 $(X_1, X_2, X_3)'$。以后本篇所说的总体是指随机向量。当我们随机抽取 1 个螺钉时，它的三个指标观测值称为**样品**。当一次抽取若干个样品时，也把这些观测值称为**样本**。多元统计的样本一般记为**样本资料阵**。下面用一个例子说明样品和样本资料阵的关系。

【例 1—1】 焊接技术培训班有 10 名学生：基础焊接技术（BWT），焊接技术提高（AWT）和焊接车间实践（PWW）的成绩如表 1—1 所示，数据文件为：MV _ 焊接成绩. BTW。

表 1—1　　10 名学生的焊接技术成绩

X_1	X_2	X_3
96	92	91
85	91	80
⋮	⋮	⋮
79	73	80

它们可以看成是随机向量的 10 次观测值：(96, 92, 91)′, (85, 91, 80)′, …, (79, 73, 80)′，每个观测值向量称为一个**样品**，10 个样品合在一起称为一个**样本**。

一般地，如果对 p 维随机向量作第 i 次观测，得 $X_{(i)}=(X_{i1}, X_{i2}, \cdots, X_{ip})'$，称它是

第 i 个**样品**，观测 n 次所得到的 n 个样品就构成一个**样本**。

在多元统计中，通常把 n 个样品的观测值像表 1—1 那样排成一个 $n\times p$ 矩阵，把随机向量的每个分量放在同一列，这样的矩阵称为**样本资料阵**，记为：

$$\boldsymbol{X}=\begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix}=\begin{pmatrix} X_{(1)}' \\ \vdots \\ X_{(n)}' \end{pmatrix}=(X_1,\cdots,X_p)$$

矩阵 $\boldsymbol{X}$ 的第 i 行：$\boldsymbol{X}_{(i)}'=(x_{i1},\cdots,x_{ip})$ $(i=1,\cdots,n)$ 表示第 i 个样品的 p 维观测值。矩阵 $\boldsymbol{X}$ 的第 j 列：

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

表示对第 j 个变量的 n 次观测值。

例 1—1 的样本资料阵为：

$$\begin{pmatrix} 96 & 92 & 91 \\ 85 & 91 & 80 \\ \vdots & \vdots & \vdots \\ 79 & 73 & 80 \end{pmatrix}$$

请注意，样本资料阵在形式上与在 MINITAB 数据文件中的工作表是完全一致的，工作表的第 i 行表示第 i 个样品，工作表的第 j 列表示对第 j 个变量的观测值，变量名称常列在表头。

和一元统计一样，由样本资料阵可以计算出来的，不包含未知参数的量称为**统计量**。

随机向量的样品、样本资料阵和统计量都是数值。和一元随机变量一样，这些量都具有两面性：当给出随机变量的观测值时，它们都是数值；当分析和表述统计量的性质时，没有给出它们的值，它们都是随机的。

1.1.1.3　随机向量的联合分布、边缘分布、条件分布

1. 联合分布

设 $X=(X_1,X_2,\cdots,X_p)'$ 是一随机向量，它的**多元分布函数**是：

$$F(x_1,\cdots,x_p)=P\{X_1\leqslant x_1,\cdots,X_p\leqslant x_p\} \tag{1—1}$$

若存在非负函数 $f(x_1,\cdots,x_p)$，使得随机向量 X 的多元分布函数对一切 $(x_1,\cdots,x_p)\in R^p$ 均可表示为：

$$F(x_1,\cdots,x_p)=\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_p}f(x_1,\cdots,x_p)\mathrm{d}x_1\cdots\mathrm{d}x_p \tag{1—2}$$

则称 X 有**密度函数** $f(x_1,\cdots,x_p)$，并称 X 为连续型随机向量。

2. 边缘分布

p 维随机向量 X 的一些分量（例如 r 个）合在一起，也是随机向量，经常要考虑它