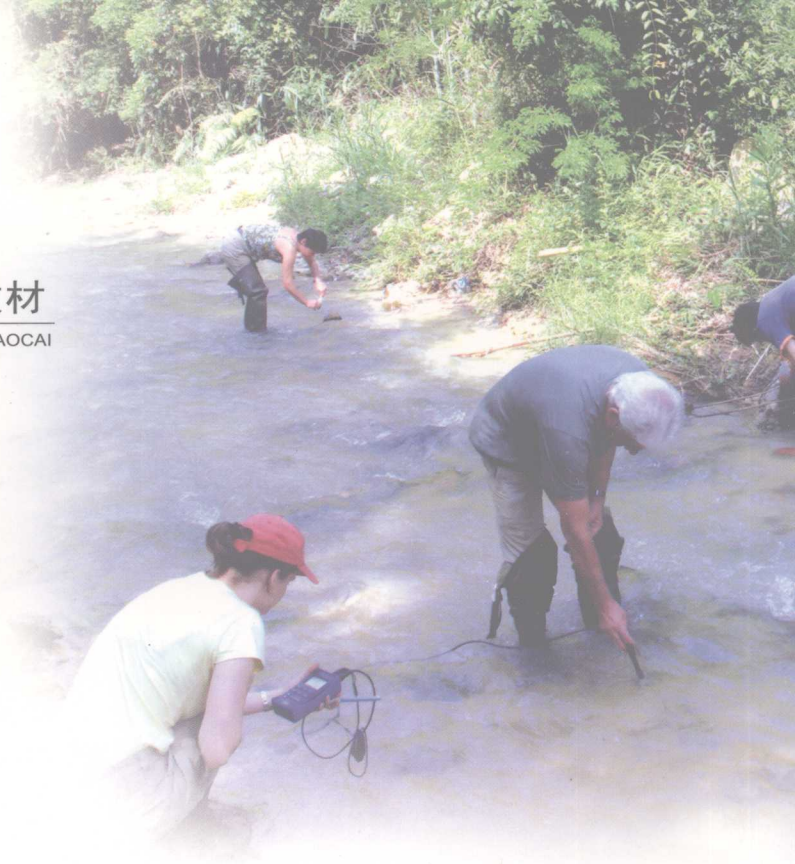

高等学校规划教材
GAODENG XUEXIAO GUIHUA JIAOCAI

赵秀兰 主编

环境研究法

HUANJING YANJIUFA

西南师范大学出版社
XINAN SHIFAN DAXUE CHUBANSHE

高等学校规划教材
GAODENG XUEXIAO GUIHUA JIAOCAI

HUANJINGYANJIUFA
环境研究法

主　编 赵秀兰
副主编 郝庆菊

西南师范大学出版社
XINAN SHIFAN DAXUE CHUBANSHE

内容简介

本书从数据资料的收集、数据资料的整理及论文的写作等方面系统介绍环境科学领域的基本研究方法，共分10章，包括：统计学的基础知识、试验研究、调查研究、数据资料的初步整理、统计假设检验、非参数检验、方差分析、环境科学中变量间的相关分析和科技论文的写作。适宜作环境科学与工程专业本科生的教材，也可供从事环境保护的科技人员参考。

图书在版编目(CIP)数据

环境研究法/赵秀兰主编. —重庆：西南师范大学出版社，2007.12

ISBN 978-7-5621-4026-9

Ⅰ.环… Ⅱ.赵… Ⅲ.环境科学—研究方法—高等学校—教材 Ⅳ.X-03

中国版本图书馆CIP数据核字(2007)第198877号

环境研究法

赵秀兰 主 编

郝庆菊 副主编

责任编辑：张浩宇

整体设计：CASTALY 尚品视觉 周娟 钟琛

出版、发行：西南师范大学出版社

重庆·北碚 邮编：400715

网址：www.xscbs.com

印 刷：四川外语学院印刷厂

开 本：787mm×1092mm 1/16

印 张：13

字 数：340千字

版 次：2008年6月第1版

印 次：2008年6月第1次印刷

书 号：ISBN 978-7-5621-4026-9

定 价：24.00元

前言

环境科学是多学科、多门类的综合性科学。高校的环境教育事业蓬勃发展，21世纪的环境教学更是要培养具备数学、物理、化学、生物、地学等方面的基础知识，掌握环境科学及相关专业的基本理论和实践技能，熟悉国家环境法律法规，了解环境科学研究及应用前沿及动态，具有较强创新型和能解决问题的应用型人才。因而，环境专业的学生除要掌握必备的基础知识和实践技能外，更应该学会如何进行科学研究。为此，我校自2003年开始招收环境科学专业以来，就开设了《环境研究法》这门课程。为配合课程的教学及提高学生应用知识的能力，编者在两年教学过程中形成了该教材的初稿，并对初稿进行了较大修改，随后在西南师范大学出版社的资助下，该教材得以出版。

本书以生物统计、数理统计等学科为基础，结合环境科学与工程领域的研究及应用成果，从科研课题的选题、数据资料的收集（调查研究、试验研究）、数据资料的整理及试验结果的整理等方面系统介绍环境科学领域的基本方法，适宜作为环境科学与工程专业本科生的教材，也可供从事环境保护的科技人员参考。

本书共十章，第一、二、三、四、六、七章由赵秀兰编写，第五、八、九、十章由郝庆菊编写，全书由赵秀兰统稿。

本书得到西南师范大学出版社的资助，得到西南大学教务处、西南大学资源环境学院领导的关心，得到西南大学环境科学与工程系的所有老师的支持，魏世强研究员为本书提出了宝贵的修改意见，部分研究生为本书收集了大量资料。本书的编写还参阅并引用了大量的有关资料。在此，编者向所有支持、关心和付出辛勤劳动的领导、老师、朋友、学生及资料的原作者表示深深的感谢。

由于编者学识有限，不妥之处在所难免，敬请相关专家和读者批评指正。

编　者

2007年8月

1 第一章 绪论
5 第二章 统计学基础知识
第一节 基本概念 5
第二节 概率基础 9
第三节 几种常见的概率分布 10
第四节 抽样分布 16

21 第三章 试验研究
第一节 概述 21
第二节 试验设计 26
第三节 正交设计 33

38 第四章 调查研究
第一节 概述 38
第二节 专门调查 41
第三节 抽样调查 43

52 第五章 数据资料的初步整理与显示
第一节 数据资料的初步整理 52
第二节 特征数的确定 58
第三节 正交试验结果的直观分析 64
第四节 可疑值的取舍 66
第五节 数据整理结果的显示 70

76 第六章 统计推断
第一节 概述 76
第二节 总体平均数比较的假设检验 80
第三节 百分数资料的假设测验 85
第四节 正态总体方差的假设检验 88
第五节 参数估计 90

CONTENTS 目录

96 第七章 非参数检验
第一节 卡平方(χ^2)检验 96
第二节 符号检验 102
第三节 秩和检验 104

111 第八章 方差分析
第一节 方差分析的基本原理 111
第二节 单因素试验资料的方差分析 118
第三节 双因素试验的方差分析 122
第四节 三因素试验的方差分析 128
第五节 正交试验的方差分析 131
第六节 方差分析的基本假定和数据转换 137

140 第九章 环境科学中变量间的关系分析
第一节 回归和相关的概念 140
第二节 直线回归分析及检验 142
第三节 直线相关分析 148
第四节 曲线回归分析 152
第五节 多元线性回归分析 157

162 第十章 科技论文与科技报告的写作
第一节 科技论文的写作 162
第二节 科技报告的写作方法 169

172 参考文献

174 附表

第一章　绪论

一、科学研究

科学研究是一种复杂的具有探索性和创造性的思维活动，是扩充、修改和整理知识的活动过程。它既包括探索、发现迄今人类尚未掌握的知识和规律，并对现今思想、活动所依据的学说和理论不断进行检验、修正的思维活动；也包括在实验室或一定范围内，在科学理论指导下发明、创造出在生产实践中可以直接利用的新产品、新工艺、新技术、新方法、新设计、新规范等的思维活动、实践活动和技术活动。科学研究的实质是通过各种科学研究方法对客观存在的事实和确凿的材料进行加工整理，从感性认识上升到理性认识，找出客观事物和过程的发展变化规律，创造出新的科学知识。

二、环境科学研究的内容及任务

环境科学是20世纪随着环境问题的不断加重而产生的一门新兴学科，其研究内容和主要任务是：(1)探索全球范围内环境演化的规律，包括环境的基本特性、环境结构的形式和演化机理。其目的是在人类改造自然的过程中，使环境向有利于人类的方向发展，避免向不利于人类的方向发展；(2)揭示人类活动同自然生态之间的关系，使环境中的物质和能量在输入和输出之间维持相对平衡，避免排入环境的废弃物超过环境自净能力以致造成环境污染，避免从环境中获取可更新资源超过其再生能力造成资源枯竭；(3)探索环境变化对生物生存的影响，主要研究污染物在大气、土壤、水和生物环境中的物理、化学变化过程，在生态系统中的迁移转化的机理及归属，以及进入人体后发生的各种作用，包括致畸作用、致突变作用和致癌作用及对其它生物体的毒理作用和遗传变异影响的机理和规律，对生态系统结构和功能的影响。这些是保护人类生存环境、制定各项环境标准、控制污染物排放的依据；(4)研究区域环境污染综合防治的技术措施和管理措施等，如大气污染防治技术、水污染防治技术、固体废物的处理和利用、噪声控制技术等。并研究环境污染综合防治，以及运用系统分析和系统工程的方法，从区域环境的整体上寻求解决环境问题的最佳方案。

三、环境科学研究的类型

现代科学体系由基础科学、技术科学和工程科学组成。相应地，环境科学研究的课题可分为基础研究、应用研究和开发研究。环境科学中的基础研究属基础科学领域，是以揭示环境科学领域中某种现象的本质及其发展规律为主要目的，如研究重金属诱导的植物细胞程序化死亡、亚热带农田生态系统中典型污染物的归趋等。其表现的具体形式是理论知识形态的科学考察与实验报告以及学术论文等，它的内容终将纳入自然科学理论的某个基础部分。基础研究的成果可以说明一般的和普遍的真理，可作为普遍通用的法则、理论和定律，对广泛的科学领域长时间产生影响。

应用研究属技术科学领域，以人类维持和改善生态环境现实中存在的许多急需弄清和解决的技术问题或对策为目的，如垃圾渗滤液的处理问题、水体富营养化问题、重金属污染土壤的修复、室内空气污染控制技术等。

开发研究则属工程科学领域，以应用基础研究和应用研究的成果，开发环保新材料、新产品、新设备、新技术、新工艺、新流程和新方法，寻求在技术上可行、经济上合理的最佳解决方案为目的。

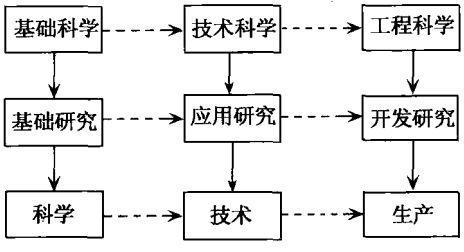

图 1-1　科学研究类型及其相关图

基础研究与应用研究之间没有绝对的界限，只是其主要目的不同。实际上，对于某一课题，既可以从理论的角度进行研究，也可以从应用的角度进行研究，或从二者兼顾的角度来研究。从成果的效用上看，基础研究的成果可用于指导实践，应用研究的成果也可用于理论研究。开发研究则是在基础研究和应用研究的基础上进行的新开发和再创造，其具体形式与应用研究大体相同，但实用性、针对性更强，与生产实际更密切、更接近。三类研究反映了科学、技术、生产三者之间的相互关系，它们之间的关系见图 1-1。

四、环境科学研究的程序及方法

同科学研究的一般过程一样，环境科学研究的程序也可分为准备阶段、实施阶段和总结阶段三个过程。相应地，科学研究的方法可以分为准备方法、数据资料的搜集方法、数据资料的加工与整理方法及材料的总结方法。

1. 准备阶段

准备阶段的主要任务是选择科研课题（选题）和确定研究方案，即解决做什么、为什么做和如何做等问题。其中，选题是科学研究的起点，决定着科学研究成果的理论价值、应用价值和研究者研究成就的大小。为此，研究人员要通过各种途径和线索，提出尚未解决的科学问题，确定研究方向，选择并论证课题，以及把课题具体化。然后，对所研究的问题提出大胆的假设，确定研究的组织形式，掌握研究的工具并确定研究的方法，拟定研究程序和制定研究计划。至于如何选题和确定研究方案将在第三章及第四章进行介绍。

2. **实施阶段**

实施阶段的主要任务是获取文献情报资料和数据，采用的方法包括文献检索法、观察法、试验法和调查研究法。科学的发展是累积性的，文献检索法是获取已有资料的重要方法。研究人员可通过查阅科技文献（包括百科全书、文献索引、专业书籍、文摘、科技期刊和杂志、专利文献等）了解特定领域研究的结论性内容、研究采用的材料及方法，获得某些特定领域的数据资料，及大量的有关产权和技术信息，是一个很重要的信息源。除科技文献外，研究人员还可从报纸、学位论文、政府出版物等获得所需的资料数据。随着现代计算机及网络技术的发展，网络系统则成为文献检索的快速、经济的重要工具。环境科学研究最常用的中文网站包括 CNKI 学术期刊（http://edu. cnki. net）、维普（http://www. cqvip. com），外文网站有 Elsevier SDOL 数据库（http://www. sciencedirect. com）、Kluwer Online Journals（http://kluwer. calis. edu. cn）、Springer 电子期刊（http://www. springerlink. com）、ACS 美国化学会数据库（http://pubs. acs. org）、http://www. blackwell－synergy. com、OVID（http://gateway. ovid. com/autologin. html）和 ProQuest 数据库（http://proquest. umi. com/login）及各期刊的主页等。涉及专利检索的有中国专利信息网（http://www. patent. com. cn）、欧洲专利局网站（http://www. epo. org）等。

观察法是科学研究中经常使用的一种搜集数据资料的基本方法。狭义的观察是指用眼睛仔细对事物的察看，广义的观察是指在自然发生条件下，或者在人为控制条件下通过视觉、嗅觉、味觉、听觉、触摸等感觉器官或者借助仪器设备对客观事物、现象进行有计划、有目的地察看或测试的一种科学观察方法。环境科学研究中，可以通过视觉和嗅觉观察大气或水环境污染导致的变化，通过对生物体表观反应的观察描述生物受到环境污染的危害程度等。目前，随着科学技术的发展，遥感技术、显微技术使环境科学领域中的观察深度和广度扩大。

试验法和调查研究法将在本书有关章节介绍。

3. **总结阶段**

总结阶段的主要任务是运用逻辑方法、数学方法，统计方法及其他方法对获得的数据进行整理和加工，运用科学思维方法进行归纳、综合，从中找出规律性的东西，或上升为理论，或形成新的开发产品。并依此来撰写研究报告、科技论文或申请专利，对不需要保密的成果应及时组织专家评价，申报请奖，乃至获奖。

五、本书的内容及课程要求

环境是一个区域性的多层次、多介质、多元的动态系统。环境科学是一门新兴的综合性学科，分支学科多，研究领域广。环境科学研究中自然会用到自然科学如地学、生物学、化学、物理学、数学等学科的理论、技术和方法。由于环境问题来自于社会和经济发展过程中人类的生产、生活和消费等活动，环境问题研究也涉及社会学、经济学和法学等学科的知识。另外，对环境污染的治理技术，离不开工程技术的原理和方法。随着科学技术的发展，环境科学的微观研究已进入到分子、原子水平，宏观研究涉及整个地球表层，甚至包括外层空间，在时间上，有历史环境、现代环境、未来环境。目前的环境科学，研究手段多样、方法日新月异，作为环境科学工作者必须掌握其专业领域的最新动态及新的研究手段，才能在研究领域中立于不败之地。但是作为大学生，最基本的还是掌握获得数据资料的基本方法（调查研究和试验研究）及数据资料的分析方法，学会如何整理、显示数据及编写研究报告。

为此，本教材从科学研究的过程系统介绍环境科学的研究方法。主要内容包括以下几个部分：(1)数据资料的收集方法，包括调查研究和试验研究；(2)资料的整理、统计方法，包括数据资料的初步整理、统计假设检验、方差分析、变量间依存关系分析；(3)研究报告的编写。为使同学们能很好的掌握本课程，增加了统计学的基本知识这部分内容。

通过该课程的学习，要求学生掌握环境科学研究中的基本原理与基本概念，学会科学地进行试验设计，掌握搜集、整理与分析资料的基本知识与技能，掌握统计方法的基本思想及有关公式的应用条件和用法，对所研究的问题能做出科学、合理的解释，并可根据这些结果对未来研究的方向和环境管理进行决策。

第二章　统计学基础知识

第一节　基本概念

一、总体与样本

1. 总体

试验研究离不开研究对象，某项试验研究的具体对象的全体称为总体，它是由许多客观存在的具有某种共同性质的个别单位所构成的集合体。构成总体的个别单位称为个体，亦称总体单位。如调查某地区稻田 CH_4 的排放量，该地区所有稻田 CH_4 的排放量就是研究的总体，每一块稻田 CH_4 的排放量就是总体单位。对个体的某种性状加以考察（如称量、度量、计数或分析测定）所得的数值，称为观测值。总体具有以下 3 个特征：

（1）同质性 同一总体的各个个体，必须在某一方面具有相同的性质，才能把它们集合起来，构成某种性质相同的总体。如研究某地区水库、湖泊水中的氮磷含量时，其共同特征是水中的氮、磷含量。所谓同质，不是绝对的，是随研究目的而变化的。

（2）变异性 在构成总体的单位具有同质性的前提下，不同单位之间一般都存在差异，即变异性。如果没有同质性便不称为总体，若没有变异性则无需统计。对总体的统计研究，实际上是研究组成这些总体的个体之间的变异情况。例如，某地区的土壤镉含量，受土壤母质、施肥及三废排放的影响，不同地块的土壤镉含量变异较大；又如同一污染物对植物的毒害作用因植物的种类、生育期及环境条件的不同而异。同质性和变异性是由事物的客观性决定，在同质性的基础上研究总体的变异程度、集中趋势及规律，是统计分析的重要任务之一。

（3）大量性 统计研究的目的在于揭示事物的客观规律，而这种规律性只能在大量事物的普遍联系中表现出来，所以总体是由大量个体构成的。总体所包含的个体数目称为总体容量（N）。若总体容量无穷大，则称其为无限总体，如大气、水体、土体等连续体，是无限不可分的，是无限的总体。若总体容量是有限的，则称其为有限总体，如某一时刻的全国人口是一个容量很大的有限总体。

对无限总体的研究只能采取非全面调查的方法进行。对有限总体的研究可以作全面普查，如全国人口普查、土壤普查等，也可进行非全面调查。但是全面普查耗资巨大，有时还难

此为试读，需要完整PDF请访问：www.ertongbook.com

以实施，对于无限总体或有限的大总体，全面普查是不现实的。当观测手段具有破坏性，即使总体容量不太大，也不允许对所有个体加以一一考察。因此，多数情况下只能采用抽样方法进行。

2. **样本**

从总体中随机抽取一部分个体所组成的集合称为样本。由样本特征来推断总体性质是统计分析的基本手段(见图 2-1)。为此，样本必须对总体具有代表性，这要求抽样应满足随机抽样的要求：(1)等可能性，即每次抽样时各个体被抽取的机会相等；(2)独立性，即每次抽样不影响下次抽样时各个体被抽取的机会。

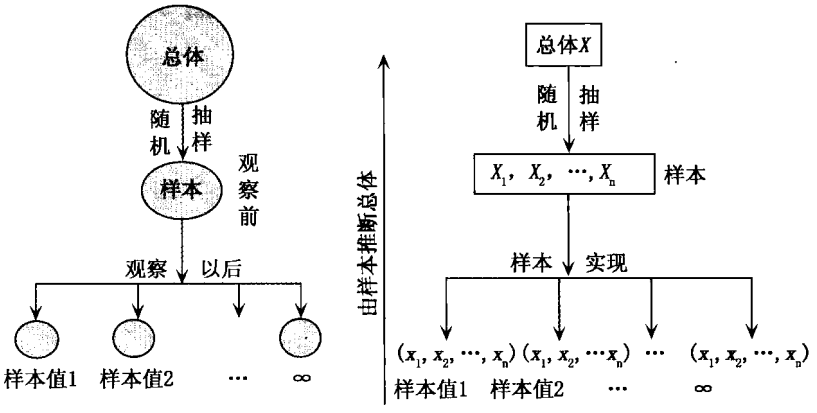

图 2-1　总体与样本

样本所包含的个体数目称为样本容量，常记作 n。$n \geqslant 30$ 的样本称大样本，$n < 30$ 时称小样本，有时大样本与小样本具有不同的统计分布特征。

二、随机变量

受许多因素的影响，总体内部各个体之间普遍存在着变异性，观测值之间也会表现出波动性。如某城市大气中 SO_2 的日均浓度，因该城市的燃煤消耗量、烟气的净化技术及气象条件的不同发生变化，是一个变量；又如，当重复 10 次测定某一土样的镉含量，其结果也会因样品的均匀程度、测定条件及测试者的操作技能等一系列因素的影响而表现出一定的波动性。受许多因素影响而表现出波动性的数据称为随机变量或随机变数，简称变数，它是表示在一个界限内变动着的性状的数值。

将某随机变量的一组具体数值称为变量值，如测定 10 个水样的全氮含量。按变量性质的不同，一般可以分为数量性状资料、质量性状资料和半定量(等级)资料三大类。

1. **数量性状资料**

数量性状是指测试、调查的对象具有可度量或计数的性质，观察测定数量性状而获得的数据就是数量性状资料。

数据变量据变量值是否具有连续性，可进一步分为连续型变量和离散型变量两类。连续型变量又称为计量资料，能用量测手段进行直接测定，如污染土壤的面积、污染物的浓度、地下水中硝酸盐的浓度、水体 DO 的浓度，土壤中的有机碳含量等。这种资料的各个观测值不一定是整数，两个相邻的整数间可以有带小数的任何数出现，小数位数的多少由度量工具的

精度而定。若某变量各变量值之间只能以整数断开而不能表现为小数的，此变量为离散型变量，离散型变量的变量值只能用计数的方法取得，又称为计数资料，如大气或土壤测定时的采样点数、重点污染企业的数量、单位体积水中细菌的数量等。

2. 质量性状资料

质量性状是指能观察到而不能直接测量的性状。在环境与资源研究中，有些研究对象的一些属性能观察而不能度量，如污染水体的颜色、污染物的气味等，具有这类性状的资料为质量性状资料。对这类资料，可采用赋值法使其数量化，即对某种属性的不同类别赋予不同的数值，例如对污染水体的不同颜色赋予不同的数值，如取红色为－1，无色为 0，绿色为 1 等，然后进行统计分析。

3. 半定量或等级资料

半定量或等级资料是指将观察单位按所考察的性状或指标的等级顺序分组，然后清点各组观察单位的次数而得的资料。这类资料既有次数资料的特点，又有程度或量的不同。如利用显色法快速测定土壤浸提液中的磷含量，根据磷含量高低有无色、浅蓝、深蓝等颜色，利用指数法评价空气中某污染物污染的严重程度有无污染、轻污染、中度污染和重污染之分。

三、参数与统计数

在同质性的前提下，总体具有变异性和大量性的特性。用于反映总体内部个体间的变异程度或集中性趋势等特征的指标为总体参数，简称参数。如总体平均数 μ、总体方差 σ^2、总体标准差 σ 等。相应地，利用样本资料计算得到的用于描述样本内部个体间的变异程度或集中性趋势等特征的一些指标，如样本平均数 $\bar{x}$、样本标准差 S 等称为样本统计数，简称统计数，它是总体参数的估计值。在数理统计中常用 $\bar{x}$ 估计 μ，用 S 估计 σ 等。

四、误差与错误

1. 系统误差

在一定条件下某客体所具有的真实数值即为真值。该客体的测定值与真值之差为误差，在一切科学试验过程中自始至终存在误差。误差可分为系统误差和随机误差两类。系统误差是由某种确定的原因所引起的误差。其特点是在相同条件下重复测定时，以相同的大小和正负性重复出现，系统误差是可以测定并校正或消除的。在资源与环境研究领域中，系统误差的主要来源有以下几方面。

(1)方法误差 因分析方法本身不够完善所造成的误差。如重量分析中沉淀不完全或有共沉淀现象、滴定分析反应不完全、比色分析中存在干扰离子等，都会使测定结果偏高或偏低。

(2)仪器误差 因所用仪器不够精确而引起的误差。如滴定管、容量瓶的刻度或仪表的刻度不准确等。

(3)试剂误差 由于试剂不纯造成的误差。不纯的试剂或去离子水带入干扰物，使测定结果偏高或偏低。

(4)操作误差 因分析人员的差异产生的误差，如对滴定终点的颜色变化敏感程度不一。

(5)环境条件的变化 如土壤肥力的方向性变化会造成试验地特定部位的供试作物产量偏高或偏低。

系统误差是重复地以固定形式出现的，不能通过增加重复次数加以消除，对分析测试中的系统误差可通过对照试验、空白试验、校准仪器等办法校正；对于田间试验中土壤肥力的方向性变化可通过随机区组设计加以局部控制。

2. **随机误差**

随机误差又叫偶然误差，由很多不可避免且无法控制的偶然因素引起。如在坩埚称重时，即使严格控制冷却时间，因空气湿度随时都在变化，各次称量可能有微小差异；控制滴定终点时，因视觉判断能力的限制，滴定终点不可能完全一致，读取滴定管体积时虽尽了最大努力，仍有±0.1ml的允许误差。对生物试验结果而言，偶然因素就更为复杂，如供试材料不均一，种子质量、秧苗素质就不可能完全一致；光照、温度、湿度等影响生长的环境因子也可能随时随地发生变化；试验时操作的不一致性以及其他不可预测的自然或人为因素的干扰均会使试验结果产生误差。

随机误差的特点是产生原因不确定，其误差大小无规律性，有时大有时小，有时正有时负，不具“单向性”或“重现性”的随机误差也称不可测误差。

随机误差虽不可避免，也不能校正，但是若在同样条件下对同一试样进行多次测定，就会发现随机误差的出现是服从统计规律的，可以利用数理统计方法对试验数据进行分析处理，通过增加试验重复次数来减小随机误差对试验结果的影响。

在试验过程中由于工作出错造成的观测值与真值的差异，称为疏失误差，即错误。如因违反操作规程而导致测定错误或数据抄错、算错等。错误是在工作中疏忽大意产生的，与误差在性质上是两个完全不同的概念。在试验研究过程中必须养成认真细致的工作习惯，保持严谨的科学态度，杜绝疏失误差的产生。

一组重复观测值中往往有个别数据与其他数据相差较大，这一数据称为可疑值或极端值，也就是离群值。如果可疑值不是由于明显的过失造成，就要用统计学的方法确定其取舍。

五、准确性与精确性

准确性指观测对象的观测值与其真值的偏离程度，偏离程度越小结果越准确。

精确性指同一观测对象的重复观测值之间的彼此相符程度，即试验误差的大小，误差越小试验越精确。

在统计工作中，常用样本统计数来估计总体参数。一般用统计数接近参数的程度衡量统计数的准确性高低，而用统计数的变异程度来衡量统计数的精确性高低。可见，准确性与精确性是不同的概念，准确性表示观测值与真值的相符程度，精确性表示重复观测值间的变异程度。

由于一般试验中真值为未知数，试验的准确性难以确定。精确性一般是指试验误差，是可以估计的。如何正确估计试验误差并减小试验误差以提高试验精度是试验方法设计所要解决的核心问题。

第二节　概率基础

一、事件

自然界中一种事物，常存在几种可能出现的情况，每一种可能出现的情况称为事件。随机试验的每一种可能结果，在一定条件下可能发生，也可能不发生，称为随机事件，通常用 A、B、C 等表示。

二、概率

事件 A 的频率：在相同条件下进行 n 次重复试验，如果随机事件 A 发生的次数为 m，那么 m/n 称为随机事件的频率。

随机事件的概率：当试验重复数 n 逐渐增大时，随机事件 A 的频率越来越稳定地接近某一数值，那么就把该数值 P 称为随机事件 A 的概率。

在一般情况下，随机事件的概率 P 是不可能准确得到的。通常以试验次数 n 充分大时随机事件 A 的频率作为该随机事件概率的近似值。

即　　$P(A) = P \approx m/n$　　（n 充分大）　　(2-1)

概率的性质：

(1) 对于任何事件 A，有 $0 \leqslant P(A) \leqslant 1$；(2) 必然事件的概率为 1，即 $P(\Omega) = 1$；(3) 不可能事件的概率为 0，即 $P(\Phi) = 0$。

小概率事件：随机事件的概率很小。例如小于 0.05，0.01，0.001，这样的事件被称为小概率事件。

小概率原理：把小概率事件在一次试验中看成是实际不可能发生的事件称为小概率事件实际不可能性原理，亦称为小概率原理。小概率事件实际不可能性原理是统计学上进行假设检验（显著性检验）的基本依据。

三、随机变量的概率

1. 离散型随机变量的概率分布

对离散型随机变量 x，其取值可以一一列出，其概率分布规律可以用分布律表示，对 x 的一切可能取值 $x_i (i = 1,2,\cdots)$，其对应的概率 p_i 为：

$$P(x = x_i) = p_i \qquad i = 1,2,\cdots \tag{2-2}$$

则称(2-2) 式为离散型随机变量 x 的概率分布。常用分布列来表示：

$$\begin{bmatrix} x_1 & x_2 & \cdots & x_n & \cdots \\ p_1 & p_2 & \cdots & p_n & \cdots \end{bmatrix}$$

离散型随机变量的概率分布具有 $p_i \geqslant 0$ 和 $\sum p_i = 1$ 两个基本性质。

2. **连续型随机变量的概率分布**

对连续型随机变量，其概率由概率分布密度函数来确定。当 x 取值于区间$[a,b)$ 时，其概率为图中阴影部分的面积，即

$$P(a \leqslant x < b) = \int_a^b f(x)\mathrm{d}x \tag{2-3}$$

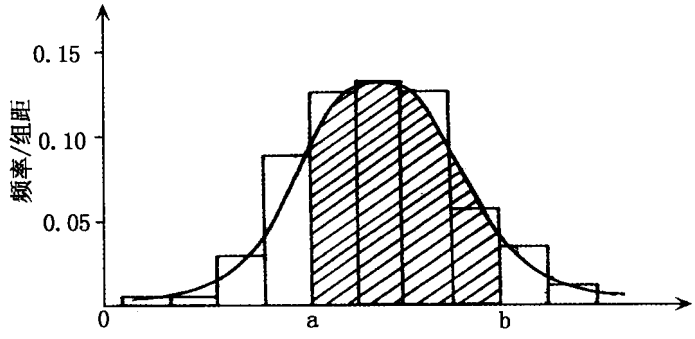

图 2-2　连续型随机变量的分布曲线

连续型随机变量概率分布具有以下性质：

(1) 分布密度函数总是大于或等于 0，即 $f(x) \geqslant 0$；

(2) 当随机变量取某一特定值时，其概率等于 0；即

$$P(x = c) = \int_c^c f(x)\mathrm{d}x = 0 \qquad (c \text{ 为任意实数}) \tag{2-4}$$

因而，对于连续型随机变量，仅研究其在某一个区间内取值的概率，不讨论某一个值的概率。

(3) 在一次试验中随机变量的取值必在 $-\infty < x < +\infty$ 范围内，为一必然事件。所以

$$P(-\infty < x < +\infty) = \int_{-\infty}^{+\infty} f(x)\mathrm{d}x = 1 \tag{2-5}$$

上式表示分布密度曲线下、横轴上的全部面积为 1。

第三节　几种常见的概率分布

一、正态分布

1. **正态分布的重要性**

正态分布是连续型随机变量的理论分布，在理论和实践上具有十分重要的意义。其理由有四：(1) 客观世界的许多现象的数据都服从正态分布；(2) 即使这些数据本身不符合正态分布，但是可通过适当的变量变换使其趋近正态分布，如平方根变换、对数变换、反正弦变换等；(3) 有些变数的理论分布以正态分布为极限分布，如二项分布、泊松分布、t 分布等；(4) 有些总体并不呈正态分布，但从总体中抽出的样本平均数及其他一些统计数的分布，在样本容量适

当大时仍趋近正态分布，可用它来研究这些统计数的抽样分布。许多统计分析方法都是以正态分布为基础的。

2. 正态分布的定义

若连续型随机变量 x 的概率分布密度函数为：

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{2-6}$$

则称随机变量 x 服从参数为 μ,σ^2 的正态分布，记为 $x\sim N(\mu,\sigma^2)$。$f(x)$ 是一给定变量值 x 的概率密度。正态分布密度函数 $f(x)$ 的图形见图 2-3。

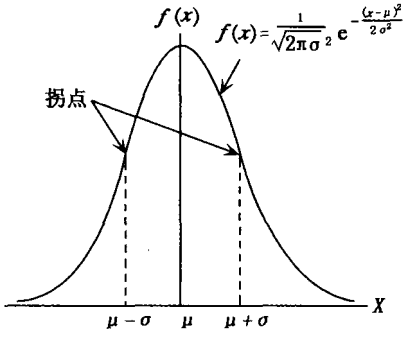

图 2-3　正态分布密度函数 $f(x)$

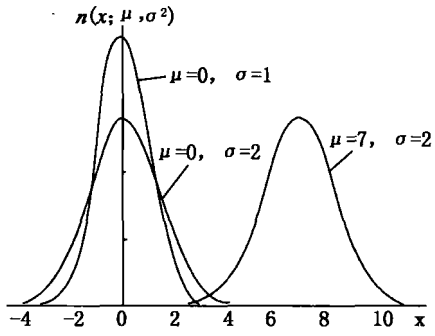

图 2-4　不同 μ,σ 的正态分布密度函数 $f(x)$

3. 正态分布的特征

除具有连续型随机变量的概率分布的分布特征外，正态分布还具有以下特征：

(1) 正态分布密度曲线是单峰对称的钟形曲线，对称轴为 $x=\mu$；

(2) $f(x)$ 在 $x=\mu$ 处达极大，极大值 $f(x)=\dfrac{1}{\sigma\sqrt{2\pi}}$；

(3) $f(x)$ 是非负函数，以 x 轴为渐近线，分布从 $-\infty$ 至 $+\infty$；

(4) 曲线在 $x=\mu\pm\sigma$ 处各有一个拐点，曲线在 $(-\infty,\mu-\sigma)$ 和 $(\mu+\sigma,+\infty)$ 区间上是下凸的，在 $[\mu-\sigma,\mu+\sigma]$ 区间内是上凸的；

(5) 正态分布有两个参数，即平均数 μ 和标准差 σ，其中 μ 是位置参数，确定它在 x 轴上的位置，σ 是变异度参数，确定正态分布的变异度。不同 μ,σ 的正态密度曲线见图 2-4。

4. 标准正态分布

当正态分布的参数 $\mu=0,\sigma^2=1$ 时，正态分布的概率密度函数写成：

$$\varphi_0(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} \tag{2-7}$$

称 x 满足 $x\sim N(0,1)$ 的标准正态分布。

对于任何一个服从正态分布 $N(\mu,\sigma^2)$ 的随机变量，都可以通过标准化变换：

$$u=\frac{x-\mu}{\sigma} \tag{2-8}$$

将其变换为服从标准正态分布的随机变量 u。u 称为标准正态变量或标准正态离差。

对任何大于零的实数 u，分布函数 $\Phi_0(u)$ 的值均可由标准正态分布函数表（见附表 1）直接查得。

此为试读，需要完整PDF请访问：www.ertongbook.com