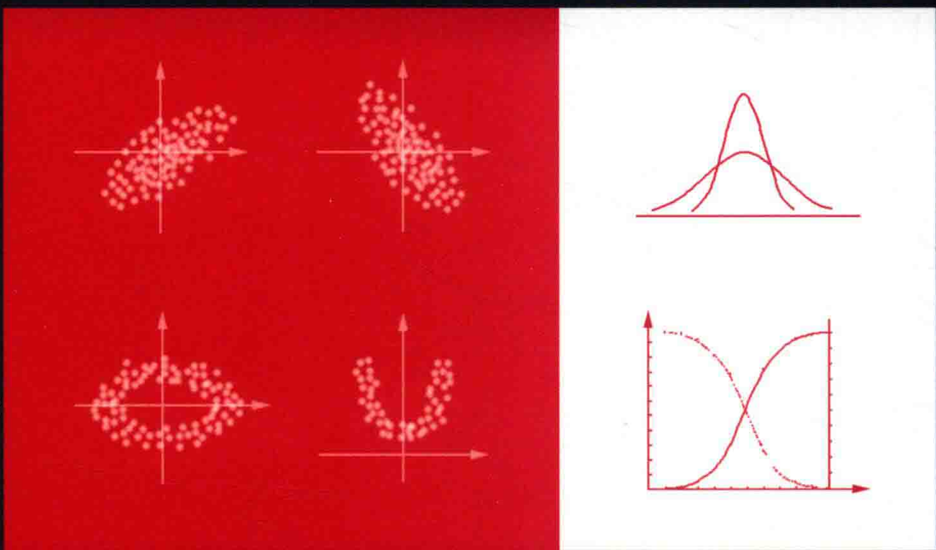
教育统计学

张 磊 姜孟瑞 编著



科学出版社

山东师范大学校级教材资助项目

教育统计学

张　磊　姜孟瑞　编著

科学出版社
北　京

内 容 简 介

统计学素养、方法论层面的思考与训练，对教师提升教育素养、教育评价能力以及研究能力至关重要. 教育统计学致力于唤起人们的方法意识、致力于奠定教师统计学知识基础，将统计学的一般原理和方法应用于教育研究，是一门相对独立的分支应用学科，是教育科学研究方法的重要组成部分，属于方法论范畴. 全书共分十三章. 前十章系统讲述了描述统计和推断统计的基本概念与原理. 包括数据的分析整理、数据特征描述、总体参数估计、假设检验以及方差分析、相关与回归分析等主要内容. 第十一、十二章分别介绍了非参数检验和抽样设计的思想方法，第十三章讲述了用 Excel 电子表格、SPSS 软件处理教育统计数据的重要内容. 本书概念阐述力求准确，论述力求严谨，注重揭示概念和方法本身的思想基础及实际意义，有助于在操作层面掌握概念，有助于更加深刻地理解统计学方法所蕴含的丰富思想.

本书可供课程与教学论专业的研究生、教育硕士以及师范院校本科生使用，可供教育管理者、工作者和广大一线教师参考.

图书在版编目(CIP)数据

教育统计学/张磊，姜孟瑞编著. —北京：科学出版社，2018.2
ISBN 978-7-03-056533-4

I. ①教… II. ①张… ②姜… III. ①教育统计 IV. G40–051

中国版本图书馆CIP数据核字(2018)第025862号

责任编辑：胡 凯 许 蕾 / 责任校对：桂伟利
责任印制：张 伟 / 封面设计：许 瑞

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

北京中石油彩色印刷有限责任公司 印刷
科学出版社发行 各地新华书店经销
*
2018 年 2 月第 一 版 开本: B5 (720 × 1000)
2018 年 2 月第一次印刷 印张: 24 1/2
字数: 494 000

定价: 79.00 元

(如有印装质量问题，我社负责调换)

前　言

追溯教育统计学发展的历史，以数理统计为工具研究教育问题，几乎是随着数理统计学本身的发展而发展的. 20 世纪初期，对统计学理论与实践的研究还多集中在欧洲各国，之后在这方面有较大贡献的有美国心理学家卡特尔(Gattell)、桑代克(Thorndike)和瑟斯顿(L. L. Thurstone)等. 1904 年桑代克出版了《心理与社会测量导论》一书，20 世纪 30 年代，麦考尔(Macall)出版《教育实验法》、瑟斯顿出版《教育统计学纲要》等著作. 但直到 20 世纪 40 年代，小样本的推断统计方法被引入教育研究中来，才真正为教育科学研究开辟了更加广阔的前景. 在我国早在 20 世纪 20 年代之后，就有不少以统计学为工具研究教育学和心理学问题的译著和专著出版，一大批学者从事这方面的研究和教学工作. 尤其是 50 年代以来，如叶佩华等的《教育统计学》、张厚粲的《心理与教育统计学》、王孝玲编著的《教育统计学》等，都为统计方法在教育和心理研究方面的应用基奠了牢固基础，为促进教育科学健康迅速地发展做出了极大贡献.

《教育统计分析方法》一书也正是在多次讲授教育统计学课程的基础上，经过五年多不断修改、打磨，撰写而成. 最初写作该书的动议，源于研究生期间学习这门课时，就深感这方面知识对教师从事教育和教学工作的重要意义，那时这方面资料还很稀缺. 2001 年秋季接手这门课时，需要的教育统计学教材也不能很容易买到，于是一边着手准备上课，一边收集资料为进一步研究做准备. 然而，这项工作却绝非想象的那样简单. 教育统计学致力于唤起人们的方法意识、致力于奠定教师统计学知识基础，将数理统计学的一般原理和方法应用于教育研究，是一门相对独立的分支应用学科，是教育科学研究方法的重要组成部分，属于方法论范畴. 在思想方法层面把这门课讲清楚并不轻松，若用文字表达出来，让别人看懂、有所领悟，并形成系统、完整的理论体系和自己独特的风格，又远比在课堂上讲明白、让学生听懂困难得多. 所以尽管自认为具有较厚重的数学基础，却还是远远不够，力不从心是常态，倘若不是从小就喜欢数学却没能学习数学专业，有意借这本书补偿一下自己早先的缺憾(事实证明，补偿缺憾之外对思想方法的深入思考才是更有意义的)，倘若不是对提高教师评价能力以及教育研究能力的一种责任与使命感，绝不可能坚持几年做这项工作而没有放弃. 正是内心深处对数学的热爱和深厚的情感，正是作为一名普通教师对这门课价值的真切的认同感，才能欣然接受各方面的挑战，才能心甘情愿付出时间与努力，全身心、全力以赴投入到这本书的写作之中，尽管辛苦却也乐在其中，尽管单调乏味却也享受那份

目　录

绪　论

一、应用数理统计方法研究教育问题是教育科学研究的必然要求

教育领域存在大量的随机与模糊现象. 如同在任何其他领域一样, 在教育领域里除确定性现象以外, 也存在着大量的随机现象和模糊现象. 众所周知, 确定性现象是指在一定条件下进行观察或试验, 其结果是确定的, 是人们可以预知的, 或者说是必然发生或必然不发生的一类现象. 例如, 水在任何温度下都会蒸发, 同种电荷相互排斥, 异种电荷相互吸引, 物体之间总有万有引力存在等, 都属于确定性现象. 微积分、微分方程、线性代数、物理学等就是研究这一类确定性现象的学科, 是人们较为熟知的一类知识与理论系统.

随机现象则是指在一定条件下可能发生、也可能不发生的现象. 因此, 在对这类现象进行观察或试验之前, 不能预知其结果. 例如, 在 5 个实验考试题中只抽一题进行实验, 将抽到第几题? 某系明年的考研率能达到多少? 如何才能使考研率达到更高? 像这样受诸多偶然因素影响的问题, 其结果是不确定的, 也是不可预知的. 但是, 事实证明在许多不确定的问题中隐含着确定性的规律, 这就是对随机现象进行大量重复地观察时, 其结果出现的统计规律性. 例如, 重复多次抛一枚质地均匀的硬币, 出现正面朝上和反面朝上的次数大致会各占一半. 因此, 随机现象的特点, 就表现在每次试验中结果的不确定性和在大量重复试验中呈现的某种统计规律性这两个方面. 概率论与数理统计正是研究和揭示随机现象统计规律性的一门数学学科. 而模糊现象所描述的事物本身的含义是不确定的, 如年轻人的集合(35 岁还属于年轻人这一群体吗?). 事实上, 现实世界中许多事物是这样模糊的、不精确定义的类型, 它们没有精确定义的判别准则, 表现出“亦此亦彼”的模糊性. 以模糊现象为研究对象的模糊数学是定量处理模糊现象的有效工具.

其实随着科学研究的不断深入, 人们需要研究的关系不仅越来越复杂, 而且对系统的判别和推理的精确性要求也越来越高. 正是为了精确描述愈加复杂的现实对象, 各类新的数学分支不断产生和发展起来, 应用较为广泛的就有: 确定性数学模型、随机性数学模型和模糊性数学模型三大类别. 比较三类数学模型, 容易发现前两种模型的共同特点是所描述的事物本身的含义确定, 具有“非此即彼”的清晰性, 而模糊性数学模型不具有这种特点. 另外, 随机性数学模型虽然具有不确定性, 但它与模糊性数学模型是有区别的. 随机性是就对事件的某种结果的机会而言, 由于条件不充分, 导致各种可能的结果, 这是因果律的破缺而造成的

不确定性. 模糊性则指存在于现实对象中的不分明现象，如“年轻”与“不年轻”之间没有明确的边界；思想品德的“优秀”与“良好”之间也找不到明确的界限. 这种不分明现象表明，从差异的一方到另一方，中间经历的是一个从量变到质变的连续过渡过程. 这是排中律(形式逻辑基本规律之一. 是说在同一时间和同一条件下，互相矛盾的两个判断中，必有一真，不能都假. 如有一个是假的，另一个一定是真的，不能有中间情况)的破缺造成的不确定性. 如果说，概率论与数理统计将数学的应用范围从必然现象扩大到随机现象的领域，那么，模糊数学则将数学的应用范围从清晰现象扩大到模糊现象的领域.

数学的发展为教育领域的定量研究提供了基础与保障. 例如，从简单的学生年龄和身高的登记统计，到较复杂的学科平均成绩的计算，甚至推断影响学生成绩的主要因素等，都离不开数学工具的应用和科学理论的支持或支撑. 这本书的目的重在研究怎样应用概率论与数理统计的基本原理和方法解决教育研究领域里的许多现实问题. 当然它不是概率论与数理统计在教育领域里的简单应用，而是根据教育现象本身的特点和规律，使这部分内容自成体系. 所以，即使有概率论与数理统计学基础，在从事教育教学研究时，仍需要有这方面知识的支持. 另外，它毕竟属于应用学科，因此，书中不会过分强调数学原理严密的逻辑推证，而是在不失准确性和科学性的基础上，侧重于数学思想和方法的具体应用以及对结果实际意义的解释，从而为教育决策或相关部门制定政策提供基础.

教育现象及其测量的随机性决定了将统计学方法引入教育研究领域的可能性与必要性. 人们早已熟悉要测量长度必须首先选择一个测量范围及精度都合适的量尺，然后进行测量. 具体测量过程中，比如要测量某人的身高，虽然被测者的身高是一个确定值，但由于诸多偶然因素的影响，不同测量者对其身高的测量结果一般不同，甚至同一测量者对同一被测者身高进行多次测量时，每次结果一般也各不相同. 所以，测量结果具有随机性. 这时，为了使测量结果更加准确，人们一般采用统计上求算术平均值的方法获得身高的测量值. 但实际上，通常无法将一切可能的测量结果全部列出. 因此，这里用求算术平均值的方法获得身高的测量值，只是求其一部分测量结果的算术平均值，而绝不是其全部测量结果的平均值. 于是测量结果的随机性还充分表现在：就不同的测量者而言，为什么偏偏用这一个测量者的测量结果而不是另外一个测量者的测量结果？或就某一测量者而言，为什么用其今天某一时间的测量结果而不是另外某天某一时间的测量结果？事实上，教育领域里其他许多类型的测量也都具有与此类似的特点，意味着教育领域中所获得的大量数据都具有随机性，如学生的学科成绩、对教师的年终考核得分等. 教育现象及其测量的随机性处处可见，例如，推断某省某年份高考某学科的平均成绩时，尽管从表面上看，把该年份参加该学科高考的所有学生的成绩求平均值最为简单方便，但就一般而言，尤其是面对像某人的身高测量值这

样一个无限总体的情况下，根本无法做到直接求解总体的平均值，则通常运用统计学方法，依据从总体中随机抽取的样本平均数，推断总体的平均数. 即使像高考这样人数确定的情况，一般也不会直接计算全部考生的平均分，而是从该学科高考成绩总体中随机抽取一部分学生的考分计算其平均值，并由此推断该省该年份该学科高考的平均成绩. 当然抽取的学生不同，推断结果也往往不同，产生的误差也不一样，所以抽样结果以及推断结果的随机性总是存在. 将统计学方法引入教育研究领域，就是要研究在怎样的情况下，能使教育测量以及推断结果的误差最小或更接近于实际情况.

教育测量的间接性与教育现象的复杂性. 教育现象还具有更加复杂的一面，就测量学生身高而言，总能找到并选择合适的量尺，直接进行测量. 可是在教育领域里绝大多数待测量物都无法直接测量，这是由于教育领域里绝大多数问题涉及人的心理现象，而心理现象是无法进行直接观察与测量的，如学生的记忆力、推理能力、实验操作能力、写作能力等，到目前为止都还无法进行直接测量. 而对这些量的了解又是进行教育教学研究的重要依据和客观基础，这正是教育领域中问题复杂性的一个重要方面. 比如，为了解学生的数学推理能力，人们不得不选择若干大家普遍认同的项目或试题组成试卷，针对学生对题目的反应情况，间接推断学生推理能力的高与低. 正因为教育测量的间接性，使教育教学研究中许多变量的值无法直接给出量的规定. 考虑到学生的情绪、态度及周围环境也会影响到学生对题目的反应等，就更增加了教育现象的复杂性.

尽管教育现象复杂多变，但总是可以测量的. 1918 年，美国心理学家桑代克提出：“凡物之存在必有其数量.”1922 年美国测量学家麦考尔又进一步指出：“凡有数量的事物都可以测量.” 这是说，任何事物、现象都有程度上的不同或数量上的差异，因此，也就为进行数量化的测定提供了可能，如学生的学科成绩有高低之分、思想品德有好坏之别等. 尽管这种好与坏之间没有明确的界限，其间的区别具有模糊性，但这种差别的存在毕竟保证了教育现象的可测量性.

教育测量以统计学理论为基础. 由于对教育现象的测量，一般是在环境、学生身心状况等诸多因素影响下对学生心理或精神特质的测定，决定了对教育现象进行测量的间接性，也大大增加了这一领域里测量的难度. 但无论如何，在教育领域里通过测量及教育调查、教育实验等多种手段获得的大量数据，为在一定条件下，对教育现象进行定量分析、处理提供了物质前提. 对数据采集的研究属于教育测量学的内容，但对数据的处理需要统计学理论的支撑. 本书正是运用概率与数理统计的理论与方法，研究如何以有效的方式整理、分析测量中所获得的数据资料，并以此为据，对所考察的问题作出尽可能精确、可靠的判断和预测，从而为教育决策提供科学依据. 这也就决定了本书最感兴趣的部分是测量数据的随机性问题.

二、概率论与数理统计及其在教育领域应用的发展

概率论起源于博弈问题. 15~16 世纪意大利数学家帕乔利(L. Pacioli)等的著作中曾讨论过“如果两人赌博提前结束，该如何分配赌金”等概率问题. 到 1654 年左右，费马与帕斯卡讨论类似的合理分配赌金的问题，并用组合方法给出了正确解答. 这引起了荷兰数学家、物理学家惠更斯的兴趣，他在 1657 年发表的《论赌博中的计算》是最早的概率论著作，但当时主要以代数方法计算概率. 一般认为，概率论作为一门独立数学分支，其真正的奠基人是雅各布·伯努利，他首次提出后来以“伯努利定理”著称的极限定理，揭示了概率的客观存在. 之后，拉普拉斯、高斯和泊松等对概率论作出了进一步的奠基性贡献. 尤其是拉普拉斯在 1812 年撰写的《概率的分析理论》一书开辟了概率论发展的新时期. 正是在这部著作中，拉普拉斯给出了概率的古典定义.

19 世纪末，概率论在统计物理等领域得到广泛应用，也同时提出了对概率论的基本概念与原理进行更加明晰解释的要求. 前苏联数学家科尔莫戈罗夫(1903~1987)在这方面做出了极其卓越的贡献. 1933 年，他以德文撰写的经典性著作《概率论基础》为概率论在更广阔范围的应用奠定了坚实的理论基础.

统计学作为一门独立学科尽管发展较晚，但简单的统计工作自古就有. 例如，古希腊和罗马时期，许多国家对人口进行统计调查、财产登记，中国古代把田产分成不同的等级等. 18、19 世纪逐渐出现统计推断的思想，如 1763 年，英国的贝叶斯提出“贝叶斯定理”给出最早的一种统计推断公式，拉普拉斯和高斯等利用这一公式估计参数，同时也建立起其他的分析方法，逐步使统计学摆脱了对观察数据的单纯描述，而向强调统计推断的方向发展. 到 20 世纪，以概率论为理论基础的统计理论已告成熟.

概率与统计学理论的使命. 概率论与数理统计都是研究随机现象统计规律性的学科，不同之处在于概率论是在已知随机变量分布的状况下，着重讨论随机变量性质及随机事件的概率求法等问题；而数理统计则是利用概率论的基本理论，对所要研究的随机现象进行多次观察或试验，研究如何合理地获得数据，如何对获得的数据进行整理、分析，并对所关心的问题作出估计或判断. 数理统计方法应用范围极广，在不同领域的应用又形成各种不同的分支应用学科. 数理统计及其在各种不同领域的应用一起构成统计学这样一个有机的庞大整体.

数理统计方法在教育领域的应用，几乎是随着数理统计本身的发展而发展的. 19 世纪末，英国生物学家、心理学家和优生学创始人高尔顿(Francis Galton)最早将统计学方法应用到对心理测量所获得的资料数据进行定量分析中来，并设计了一种相关系数的计算方法. 1896 年他的学生皮尔逊(Karl Pearson)改进了相关系数的计算，创立了积差相关系数公式，至今仍在广泛应用. 同时期的英国另一

位心理学家斯皮尔曼(Spearman)，1904 年发表论文“普通智力”，首创智力结构的二因素学说，并提出因素分析的思想和方法等，在统计方法的应用方面做出了巨大贡献. 20 世纪初，对统计学理论与实践的研究还多集中在欧洲各国，之后在这方面有较大贡献的有美国的心理学家卡特尔、桑代克和瑟斯顿等人. 1904 年桑代克出版《心理与社会测量导论》一书，系统地介绍了统计方法和测验编制的基本原理，极力倡导以心理学和统计学为工具来研究教育学. 到 20 世纪 30 年代有这方面的专著出版，如麦考尔的《教育实验法》、瑟斯顿的《教育统计学纲要》等. 但直到 20 世纪 40 年代，随着小样本的推断统计方法被引进教育研究中来，才真正为教育的科学推断提供了一件利器，从而为教育科学研究开辟了更加广阔的前景.

在中国，教育统计学方法是随着欧美各国科技成就一同被介绍进来的. 20 世纪 20 年代以后，就有不少这方面的译著和专著出版，一大批学者从事这方面的研究和讲授工作. 尤其在文革以后，在这方面有较大影响的有叶佩华等的《教育统计学》、张厚粲的《心理与教育统计学》、还有王孝玲编著的《教育统计学》等. 都为统计方法在教育和心理研究方面的应用做出了极大的贡献，也为促进教育科学健康、迅速地发展做出了极大贡献.

三、教育统计学的主要内容

将数理统计学的一般原理和方法应用于教育研究，就形成了教育统计学这样一门相对独立的分支应用学科，属于方法论范畴，也是进行教育科学研究的重要工具. 它以辩证唯物主义思想为指导，运用统计学方法，研究如何以有效的方法收集、整理和分析在教育调查、教育实验中所获得的数据资料，从中认识教育现象的本质，发现教育规律，并以此为依据，对所考察的问题作出尽可能精确、可靠的判断和预测，从而为教育决策提供科学依据.

教育统计学与数理统计学的关系还表现在：对教育科学研究的逐步深入，不仅提出了许多需要教育统计学进一步研究解决的问题，也同时为数理统计学提供了新的研究内容，使二者相互促进，相辅相成；另外，在教育领域里应用统计学方法，并不断开发适合教育现象特点，符合教育现象规律的新方法，也就使统计学方法带有了教育的特殊性和复杂性. 从而使教育统计学成为一门相对独立的分支应用学科，也成为教育科学研究方法的一个重要组成部分. 因此，即使具有数理统计学知识的人进行教育科学研究，仍然需要进行教育统计方法的学习、训练和指导.

教育统计学的研究内容是随着历史的发展而不断丰富和发展的，但从具体应用的角度来看，大体可分为描述统计、推断统计和实验设计三个部分.

1. 描述统计

在教学实践中，常以平均分数来衡量学科教学质量的高低. 之所以用平均分数，是因为平均分数往往代表了学科的平均水平，或者平均分数可以被视为这一群体成绩中的典型或代表. 像平均分数这样能代表一组数据典型水平或平均程度的量，在统计上称为集中量. 所以，平均分数是集中量的一种(另外还有其他一些集中量，如中位数、众数等).

但实践表明，仅有平均分数是不足以准确、全面地说明教学质量状况的. 比如有 A、B 两个平行班级，相同测验的平均成绩相同. 在 A 班中，全班成绩绝大多数都集中在平均成绩附近一个不太大的范围内，这时，平均分数对 A 班的成绩确实具有很好的代表性，能说明 A 班的平均水平或典型状况. 而 B 班中，学生成绩两极分化严重，一部分成绩偏低，而另外一部分成绩又极高，显然这时的平均分数就没有多少代表性，对 B 班的平均程度没有多少说服力. 因此，在计算平均成绩的同时，还要考察成绩的离散情况. 个体成绩越是向平均成绩集中，平均成绩越具有代表性. 反之，则平均成绩并不能说明整体的平均水平. 在统计学中，把表示数据离散程度的量叫做差异量，如标准差、全距等都是差异量. 可见，集中量还表征了一组数据向某点集中的趋势，并且当差异量越小时，表明一组数据向某点集中的趋势越强. 因此，在描述一组数据特征时，应把集中量和差异量结合使用.

实践中还常关心各学科成绩之间的关系，即两组或多组数据之间的关系. 例如，数学与物理成绩之间就常表现出一种协同变化的关系，即表现为两门成绩都好或者两门成绩都差的情况. 但这种关系又不同于因果关系，一般不是数学好的物理也一定好，或者数学差的物理也一定差. 在统计中，是通过计算相关系数一类的相关量，来达到了解同一事物的不同特性间变化关系的一致性程度或协同变化的程度的，如数学与物理成绩间就可以通过计算它们之间的相关系数，以达到描述它们间协同变化程度的目的.

像上面这样，通过计算一组数据的集中量、差异量和相关量(统称为特征量)等，来研究数据的分布特征，就属于统计描述所讨论的内容. 简言之，统计描述就是对已获得的数据进行整理、概括，制成图表，或就这些数据计算其各种特征量，以尽可能全面地反映数据分布的特征. 其目的在于从杂乱无章的数据资料中获得尽可能多的有意义的信息，以便对不同的总体进行推断分析和比较，并尽可能得出符合实际的结论来.

2. 推断统计

假如要了解某年某省高考物理的平均成绩，面对全省几十万人的高考人数，由于人力、物力和时间所限，一般不能对整体进行统计计算，而是随机地从中抽

取一部分物理试卷(样本)，计算这部分试卷的平均成绩(统计量)，然后在一定的置信度下，通过归纳概括推断全省(总体)的高考物理平均成绩(参数). 这里，置信度表示把握性(可靠性)大小. 之所以有一定置信度，是由数据概率分布的特性决定的. 最常用的置信度是95%和99%两种情况.

概括来讲，根据样本所提供的数据信息，依据概率理论，在一定置信度下，推测样本所属总体相应性质的统计方法，都属于推断统计的内容.

根据推断的目的不同，统计推断又包括总体参数估计和假设检验两部分.

3. 实验设计

实验设计是指研究者在教育实验开始之前制订的详细的实验计划，包括随机样本的抽取方式、样本大小(样本容量)的确定以及对实验数据进行统计处理和分析方法的选择等.

以上三部分内容相互联系，相互影响. 描述统计是推断统计的基础，同时，只有良好的实验设计才能保证获得有意义且有价值的实验数据；也只有在对这样的数据进行整理、分析的基础上，才能通过统计推断得出与教育实际相符合的结论. 当然，良好的实验设计必须以统计原理为依据，满足统计方法的要求，否则就无法对实验数据进行统计处理.

但由于实验设计涉及的问题比较复杂，而且往往自成体系，所以本书仅以描述统计和推断统计两部分为主要研究内容.

四、学习统计学应注意的问题

统计学作为科学研究中对数据进行分析处理的思想武器与有力工具，其应用不能脱离对实际问题背景的考虑. 一方面，统计学作为一种科学研究的方法理论，其本身不能决定一项科学研究成果的价值，一个毫无理论意义和任何应用价值的研究，即使用再好的统计学方法也不能提升其研究水平，因此不能滥用统计方法. 尤其用来粉饰自己研究或掩盖实验缺陷甚或弄虚作假地编造数据，都是缺乏科学态度的表现. 另一方面，统计学作为教育科学研究方法的重要组成部分，又必须牢固掌握. 因为，任何一项高水平的研究，若没有对数据进行科学有效的统计分析方法作支撑，也不能将其结果的重大意义准确、合理地呈现出来. 因而，在应用统计方法于教育教学研究中时，切记要以辩证唯物主义思想做指导，以教育学和心理学的理论为依据，具有正确的思想和科学方法观念，才能充分发挥统计学所应有的作用.

在学习统计学时，应注意把握它的特点. 这大致包括以下几个方面：

(1) 从解决问题方法的视角来看，统计推断是归纳法，即是一种从特殊到一般的辩证思维路线. 这从前面所提到的，推断某省高考物理平均成绩一例中，已经体现出来，是一个统计、分析与归纳的过程.

(2) 以归纳法为基础的统计推断与确定性数学模型中思维方法的“推理”不同，而具有不确定性. 这种不确定性以数据的概率分布特征为客观基础，一般用置信度或可靠度来表征.

(3) 统计推断对所建立的假设往往采用反证法. 但这里的反证法又与以往所熟悉的反证法不同，而具有概率的性质. 这一点也是由数据的概率分布决定的. 统计推断的反证法实际上遵从基于概率分布的证伪的思想路线. 因而与证实的思想方法中，某一命题被证实后接受，不被证实则被抛弃不同，运用证伪方法时，总是试图找反例否定某一命题，只有在无法找到反例或者证据不足以否定它时，才不得不接受该命题，同时保留继续检验或证伪它的可能性. 证伪思想方法是对证实方法的补充更是一种态度，它强调真理的相对性、尊重人认识上的局限性，注重以质疑地态度面对知识，主张不再理所当然地接受事物或真理.

总之，统计学方法与以往确定性数学模型中处理问题的方法有许多不同之处. 因此，应注意用新思路解决新问题，而不能把以往的知识和经验生搬硬套. 统计处理不需要多么高深的数学基础，只要着重掌握统计学的特点和规律，学好它是很容易的事情. 学习统计学绝不能执着于数学的运算，而是要注重理解统计学方法中所蕴含的思想方法及其内涵.

五、学习统计学之前应先熟悉的一些概念

1. 随机试验

前面已经揭示出教育现象的随机性. 随机现象有两个突出特点：一是在一次试验中，结果呈现不确定性；二是在大量重复试验中，结果所具有的统计规律性. 这意味着，随机现象的统计规律性只有在大量重复试验中才呈现出来. 因此，对随机现象统计规律性的认识，离不开对大量重复试验的观察. 实际上，试验这一概念本身就泛指在一定条件下，对自然与社会现象进行的**观察或实验**. 所以，寻找随机现象的统计规律，离不开大量的重复试验.

当试验满足以下条件时，被称为随机试验：

(1) 试验可以在相同条件下重复进行；

(2) 试验结果不止一个，但试验之前能明确指出所有可能结果的范围；

(3) 每次试验之前不能准确预知哪一种结果会出现.

2. 随机事件

随机试验总有各种不同的结果，通常把试验的每一个最基本的结果称为基本事件. 如投掷一枚质地均匀的骰子，“1”在上面(或“2”在上面……)就是一个基本事件. 在一次试验中，只能有一个基本事件出现.

由若干个基本事件组合而成的试验结果叫复合事件. 如投掷一枚质地均匀的骰子，上面点数是“偶数”(2, 4, 6)，就是一个复合事件，它包含 3 个基本事件. 无论是基本事件还是复合事件，在随机试验中出现与否都具有随机性，因此，统称为随机事件，简称为事件，通常用大写字母 A，B，C，…表示. 如用 A，B，C 分别表示以上试验中上面点数是“2”、“偶数”(2，4，6)和“1”的情况. 则当上面出现“2”点时，我们说 A 事件发生，B 事件也发生，但 C 事件不发生. 并且我们会很自然地关心一次试验中 A 事件发生的可能性有多大？B 事件和 C 事件发生的可能性又如何呢？

在随机试验规定的条件下，必然出现的事件叫必然事件；而必然不出现的事件叫不可能事件. 必然事件和不可能事件已不具有随机性，但为了方便，仍把它们视为随机事件，是随机事件的两个极端情况，如同静止与匀速直线运动是匀变速直线运动的两个特例一样.

3. 随机变量

在许多随机试验中，试验的每一种可能结果对应着一个数值，如英语四级的通过率、汽车通过高速公路某给定点的速度等，每一次试验或观察结果都可以用一个数值来表示. 按照通常的做法，也是为了研究和解决问题的方便，人们将这些数值与一个变量的取值联系起来，这个用来表示试验结果的变量就称为随机变量. 统计学处理的变量都是随机变量.

即使有些随机试验的结果不具有数量性，譬如，每天第一节课第一个到教室的是男生还是女生？这时也可以把它与数量联系起来. 比如可以把女生定为“1”，男生定为“0”，从而把试验结果数量化了. 当然，这时的随机变量只有两种取值，并且这时的“0”和“1”也失去了原来的数据性质和大小关系.

随机变量一般用 X，Y，Z 等来表示. 引入随机变量后，随机事件就可以通过随机变量的不同取值来表示. 如上面例子中，“第一节课第一个到教室的是女生”这一事件可表示为$\{X=1\}$；反之，随机变量的不同取值也都对应着一个随机事件. 例如，掷骰子的试验中，可用$\{X=1\}$或$\{X=2\}$等表示上面出现的点子数目.

随机变量的取值状况固然重要，但更重要的是随机变量以什么样的概率取值. 在上例中，或许人们更关心的是，“第一节课第一个到教室的是女生”的概率究竟是多大？

4. 总体、个体和样本

正如在推断全省物理高考的平均成绩时，一般不会对全省所有参加高考的考生的物理成绩直接相加求取平均值，而是从中随机抽取一个合适的样本，算出其平均成绩. 然后依据统计推断的基本原理以及样本的平均值，在一定置信度下，推断全

省物理高考的平均成绩．其中全省所有参加高考的考生的物理成绩就组成一个总体．因而，总体是所要进行研究的具有某种共同特性的个体的集合．而组成总体的每一个基本单位称为个体．个体可以是人，也可以仅指人的某一种属性．该例中，个体就是指每个考生的物理成绩，而所有考生的物理成绩则构成该例中的总体．

总体有有限总体与无限总体之分．有限总体是说总体所包含的个体数目有限；而无限总体所包含的个体数目为无限．譬如，在上面测量某学生身高的试验中，每一个测量结果就是一个个体，而总体是由一切可能的测量结果组成．所以，这个总体就是一个无限总体．总体中所包含的个体数目一般用 N 来表示．对有限总体 N 是一个有限的数值，对无限总体 N 则趋于无穷大．

面对总体(或称研究对象)，似乎最精确的研究办法是对每一个个体进行观测，但事实并非如此．尤其是研究对象为无限总体的时候，这样对所有的个体均进行观测研究，在实际中是行不通的，因而抽取样本总是必要的，即从总体中随机抽取一个样本进行详细研究．样本就是从总体中抽取的作为观察对象的一部分个体．实践中就是通过对观察样本所获得的数据进行分析，再进而对总体的情况做出预测或估计．样本中包含的个体数目称为样本容量，一般用 n 表示．$n>30$ 的样本称为大样本；$n\leqslant30$ 的样本为小样本．在有些特殊情况下，$n\leqslant50$ 的样本也被视为小样本，这在处理具体问题时再详细说明．之所以区分大样本与小样本，是因为在对数据进行统计处理时，对大样本与小样本所采用的统计分析方法一般不同．以什么样的方式从总体中抽取样本以及样本容量 n 的确定等问题，将在本书的第十二章专门讨论．

就总体与样本而言，应注意两点：一是总体的有限与无限的相对性，如有时总体虽然有限，但样本容量 n 与 N 相比小得多，这时可近似把总体视为无限总体；二是总体与样本的相对性，如推断某省物理高考平均成绩时，该省物理高考成绩构成一个总体．但当推断全国的物理高考平均成绩时，该省的物理高考成绩又可被视为一个样本．

5. 统计量和参数

根据样本数据所计算出来的一些量，如平均数、标准差与相关系数等，能够描述这组数据的特征，人们称之为统计量．统计量与样本相对应，是样本上的数字特征．当样本不同时统计量也不同．所以，统计量由样本决定，可以说是样本的函数．由于样本随机抽取，所以统计量也具有随机性，是一个随机变量．

表征总体各种数字特征的量称为参数．如总体的平均数、标准差与相关系数等．它们可能还不为人所知，但是客观存在，是一些常数，参数已不具有随机性，不再是随机变量．

统计推断的目的，就是运用统计学的一般原理与方法，根据样本统计量来推