

文科应用数学

主 编 段耀勇 王松敏 王丙参
副主编 刘丽芳 安晓伟 李育安

文科应用数学

主 编 ① 段耀勇 王松敏 王丙参

副主编 ① 刘丽芳 安晓伟 李育安

西南交通大学出版社

· 成 都 ·

图书在版编目 (C I P) 数据

文科应用数学 / 段耀勇, 王松敏, 王丙参主编. —
成都: 西南交通大学出版社, 2021.8
ISBN 978-7-5643-8220-9

I. ①文… II. ①段… ②王… ③王… III. ①应用数
学 - 高等学校 - 教材 IV. ①O29

中国版本图书馆 CIP 数据核字 (2021) 第 171580 号

Wenke Yingyong Shuxue

文科应用数学

主编 段耀勇 王松敏 王丙参

责任编辑 孟秀芝

封面设计 原谋书装

出版发行 西南交通大学出版社
(四川省成都市金牛区二环路北一段 111 号
西南交通大学创新大厦 21 楼)

邮政编码 610031

发行部电话 028-87600564 028-87600533

网址 <http://www.xnjdcbs.com>

印刷 成都蜀通印务有限责任公司

成品尺寸 185 mm × 260 mm

印张 27.25

字数 613 千

版次 2021 年 8 月第 1 版

印次 2021 年 8 月第 1 次

定价 59.00 元

书号 ISBN 978-7-5643-8220-9

课件咨询电话: 028-81435775

图书如有印装质量问题 本社负责退换

版权所有 盗版必究 举报电话: 028-87600562

随着 21 世纪知识经济时代和信息时代的到来，人文领域中许多研究对象的数量化趋势越发明显，加上计算机的普及和应用，给人们一个现实的启示：每一个想成为有较高文化素质的现代人都应当具备较高的数学素质，数学教育对文科大学生来说必不可少。马克思在几百年前就曾指出：“一种科学只有在成功地运用数学时，才算达到了真正完善的地步。”可见，数学的基础作用，无不体现在其他学科的深入研究中。

对研究生教育而言，数学是重要的基础课。通过相应的数学课程的学习，学生可为自己后续的科学研究、学术论文及毕业论文中的理论推导和科学计算打下坚实的基础，这点在理工科专业尤为突出。国外高校在这方面的意识更早，体制也更健全，很多非数学专业的学生都要选修大量的数学课。在我国，目前随着大数据时代的到来，大家越来越意识到数学的重要性，国内高校对文科研究生的数学教育也越来越重视，很多高校都针对性地开设了一些数学课程。

“文科应用数学”是中国人民警察大学于 2013 年针对边管、警卫、维和、法学、政工、情报与战术等文科硕士研究生开设的公共基础课程。2013 年被称为大数据元年，大数据的数学基础是统计学。统计学的一个典型特点是利用一定的资料对所关心的问题做出尽可能精确可靠的预测，依据所做预测，并考虑到行动的后果而制订一种行动方案，这使得它在人文社会科学中的应用越来越广泛和深刻。另外，博弈论既是现代数学的一个新分支，也是运筹学的一个重要学科，它在国际关系、政治学、军事战略、法学等学科都有广泛的应用。基于社会大背景，结合学校

研究生的专业方向，我们选取统计学和博弈论两部分内容来讲述这门课，并借助数学软件完成统计学中的相关计算。经过这几年的实践，从早期借助 Excel 完成相关计算，到现在利用 SPSS 完成相关统计分析，从没有统一的讲义和课件，逐步过渡到较成型的教学理念、课件和详细的讲义，在此基础上编写了本书。

本书具体编写分工如下：第 1 章“统计学基础”由王丙参编写；第 2 章“数据的整理与显示”和第 3 章“数据分布的描述与分析”由安晓伟编写；第 4 章“相关与回归分析”和第 9 章“完全信息静态博弈”由刘丽芳编写；第 5 章“时间序列分析”和第 7 章“博弈论导引”由王松敏编写；第 6 章“多元统计分析”由李育安编写；第 8 章“博弈论的表述方式”由李秀林编写；第 10 章“完全信息动态博弈”由姜春艳编写。第 11 章“附录”由段耀勇和张聪编写。

中国人民警察大学（原中国人民武装警察部队学院）的学员在智慧警务学院数据科学教研室（原基础部数学教研室）的教师们的指导下共同参加了全国大学生数学建模、全国研究生数学建模以及全国军事数学建模等竞赛活动，警大学子不畏强手，奋力拼搏，成绩斐然。近 20 年来荣获全国大学生数学建模竞赛一、二等奖；全国研究生数学建模竞赛一、二、三等奖；军事数学建模特等奖和一、二、三等奖以及其他赛事奖项。为了表扬和记住这些辛勤汗水浇灌出来的果实，本书最后附上部分学员获奖论文。

该书的优点是方便阅读，书中注重数学思想的讲解与方法的应用，理论的推导没有过多展开，复杂烦琐的数学计算由 SPSS 完成（SPSS 的

界面与 Excel 类似,且可以直接打开 Excel 文件,不涉及编程,方便操作,比较适合文科类专业学生使用)。整体来看,该书适合文科类专业的研究生或本科生使用。由于作者水平所限,不足之处在所难免,希望广大读者多提宝贵意见,以便今后再版时修订。

天水师范学院统计系统计学博士王丙参作为主编参与该书的编写,使得本书无论是从学术水平还是应用价值上都大有增色。该书的出版得到了中国人民警察大学研究生院、国家自然科学基金(11701446)和科技部重点研发项目(2017YFF0207400)的资助,防灾科技学院基础课教学部王福昌教授审读全稿并提出了宝贵的修改意见,由王松敏统稿完成,在此一并致谢!

编者

2020年12月20日于中国人民警察大学

第 1 章	统计学基础	
1.1	统计学简介	002
1.2	数据的分类、来源与质量	014
1.3	统计调查方法	022
1.4	统计调查的组织方式	028
1.5	统计计算工具	036
第 2 章	数据的整理与显示	
2.1	数据的预处理	044
2.2	品质数据的整理与显示	048
2.3	数值型数据的整理与显示	055
2.4	统计表	065
第 3 章	数据分布的描述与分析	
3.1	集中趋势的测度	070
3.2	离散程度的测度	079
3.3	偏度和峰度的测度	083
3.4	利用 SPSS 计算统计量	085
第 4 章	相关与回归分析	
4.1	相关与回归分析概述	092
4.2	相关分析	095
4.3	一元线性回归分析	101
4.4	多元线性回归分析	107
4.5	非线性回归模型	110
4.6	Logistic 回归	115
第 5 章	时间序列分析	
5.1	时间序列分析概述	128
5.2	趋势线拟合法	130
5.3	移动平均法	142
5.4	指数平滑法	152
5.5	ARIMA 模型	168

第 6 章 多元统计方法	
6.1 主成分分析	188
6.2 因子分析	195
6.3 聚类分析	200
6.4 判别分析	207
第 7 章 博弈论导引	
7.1 博弈论的基本概念	226
7.2 经典案例	231
7.3 博弈中的随机行动	237
7.4 有趣的智力游戏	240
7.5 合作中的博弈	244
7.6 合理制度促进社会进步	245
第 8 章 博弈论的表述方式	
8.1 标准式表述	254
8.2 扩展式表述	261
8.3 两者之间的转换	264
第 9 章 完全信息静态博弈	
9.1 占优均衡	270
9.2 重复剔除的占优均衡	274
9.3 纳什均衡	286
9.4 混合策略的纳什均衡	291
9.5 多重纳什均衡及其甄别	300
第 10 章 完全信息动态博弈	
10.1 完全且完美信息动态博弈	306
10.2 完全非完美信息动态博弈	314
10.3 重复博弈	319
10.4 博弈中的承诺行动	324
附 录 获奖建模论文选编	333
参考文献	428

第 1 章

统计学基础

在日常生活中，我们经常与“数”打交道，网络、电视、报纸上数据无处不在，例如，北京房价下跌 10%、上证指数上涨 8%、就业率升高 6%等。要使这些数据变为对你有用的信息，帮你决策，就需要对这些数据进行处理与分析。统计学就是一套处理与分析数据的基本方法和技术，统计学知识是正确阅读并理解数据、图表等的基础。在很多领域中，统计学都有应用且成绩斐然。下面是统计研究得到的一些结论：

- (1) 吸烟有害健康，吸烟男性减少寿命 2 250 天。
 - (2) 身高高的父母，其子女的身高也高。
 - (3) 第一个出生的子女比第二个出生的子女聪明。
 - (4) “怕老婆”丈夫得心脏病的概率较大。
 - (5) 上课坐在前面的学生平均考试分数比坐在后面的学生高。
- 这些结论正确吗？你相信吗？

理解并掌握一些统计学知识对于普通大众是必要的，因为我们每天都关心生活中的一些事情，这其中就包含了很多统计学知识。比如，在外出旅游时，需要关心一段时间内的详细天气预报；在观看世界杯时，了解各球队的技术统计，等等。理解并掌握一些统计学知识对于制定决策的人更为重要，因为在他们做出决策时，如果不懂统计，就可能闹出笑话，甚至损失巨大。

在终极的分析中，一切知识都是历史；在抽象的意义下，一切科学都是数学；在理性的基础上，所有的判断都是统计学。总有一天，统计思维会像读与写一样成为一个有效率公民的必备能力。

本章主要介绍统计学的基本问题，包括统计学简介、数据的分类与来源、统计调查方法与组织方式等，最后简要介绍统计计算工具。

本节将详细介绍统计与统计学的概念，给出统计学的研究对象、分类及统计的应用领域，最后回顾了统计学的基本概念，以供读者参考。

1.1.1 统计与统计学的概念

人们在日常生活中经常接触“统计”一词，但很多人对其一知半解，甚至存在很多误区。“统计”一词起源于国情调查，最早意为国情学，且历史悠久，可以说，自从有了国家就有了统计实践活动。最初，统计只是为了统治者管理国家的需要而搜集资料，弄清国家的人力、物力和财力，作为国家管理的依据。

(1) 中国：公元前 22 世纪的夏禹时代，中国分为九州，人口约 1 352 万人，由此可见人口统计历史久远；《书经·禹贡篇》记述了九州的基本土地情况，被西方经济学家推崇为“统计学最早的萌芽”；西周建立了较为系统的统计报告制度；秦时《商君书》中提出“强国知十三数”，其中包括粮食储备、各国人数、农业生产资料及自然资源等。

(2) 国外：公元前 27 世纪，埃及为了建造金字塔和大型农业灌溉系统，曾进行过全国人口和财产调查；公元前 15 世纪，犹太人为了战争的需要进行了男丁的调查；公元前约 6 世纪，罗马帝国规定每五年进行一次人口、土地、牲畜和家奴的调查，并以财产总额作为划分贫富等级和征丁课税的依据。

一般认为，统计学的学理研究始于古希腊的亚里士多德时代，迄今已有两千多年的历史。今天，“统计”一词已被人们赋予了多种含义，很难给出一个简单的确切定义。人们给统计学下的定义很多，比如，“统计学是收集、分析、表述和解释数据的科学”；“统计是一组方法，用来设计试验、获得数据（data），然后在这些数据的基础上组织、概括、演示、分析、解释和得出结论”。综合地说，统计学（statistics）就是收集、处理、分析、解释数据并从数据中得出结论的科学。这一定义揭示了统计学是一套处理数据的方法和技术。与物理学的假说类似，统计学的模型仅仅是对现实的近似，没有任何模型是“正确”的，也无法证明任何模型是正确的。只能说，在某些可能有争议的准则下，某些模型比另外一些模型更合适一些。在数学逻辑中存在的确定性在统计学中完全不成立，针对不同学科问题而发展的统计学中的数学完全不成为一个封闭的体系，也没有成为一个数学体系。切记，是否解决问题是评价统计方法的最终标准。

在不同场合，“统计”一词具有不同的含义，它可以是统计数据的搜集活动，即统计工作；也可以是统计活动的结果，即统计数据资料；还可以是分析统计数据的方法和技术，即统计学。

(1) 统计工作是指搜集、整理、分析和研究统计数据资料的工作过程。统计工作在

人类历史上出现的比较早，随着历史的发展，统计工作逐渐发展和完善，统计成为国家、部门、事业和企业、公司和个人及科研单位认识与改造客观世界和主观世界的一种有力工具。我国各级政府机构基本上都有统计部门，比如统计局，它们的主要职能是从事统计数据的收集。大多数企业也有专门从事统计工作的人员，负责企业生产和销售数据的记录、积累以及向上级部门报送数据的任务。统计工作，可以简称为统计。

(2) 统计数据是统计工作进行搜集、整理、分析和研究的主体及最终成果。数字无言，却最有说服力；数字简洁，却最适合描画过去与未来的轨迹。我们经常会看到专门出版统计数据的出版物，如《统计年鉴》，在网络、报纸、杂志上也常常见到大量统计数据。当你看到或听到“据统计……”这样的说法时，这里的统计一词指的是统计数据。由于统计数据 (statistics) 在英文中以复数形式出现，表明统计数据不是指单个数字，而是指同类的较多数据。因为单个数据如果不和其他数据比较，是不能说明问题的。例如，某学生在统计学考试中得 86 分，如果仅凭这一数字，我们很难对这位学生的知识和能力做出判断和评价。因为 86 分在班级中可能是最高分，也可能是中间分，还可能是最低分。

(3) 统计学是对研究对象的数据资料进行搜集、整理、分析和研究，以显示其总体的特征和规律性的学科，亦可简称为统计。例如，我们所学的课程“统计”，实际上指的是“统计学”课程。

正确理解“统计”概念是必要的，一提到“统计”，就想到统计工作的思维习惯是片面的、狭隘的，而要针对具体情况具体分析，同时三者相辅相成，不可分割。具体关系如下：

(1) 统计数据和统计学的基础是统计工作，统计工作的成果是统计数据，统计学既是统计工作经验的理论概括，又是指导统计工作的原理、原则和方法。原始的统计工作，即人们收集数据的原始形态，已经有几千年的历史，而它作为一门科学，则是从 17 世纪开始的。在英语中，统计学家和统计员是同一个单词，但统计学并不是直接产生于统计工作的经验总结。每一门科学都有其建立、发展的客观条件，统计学是统计工作经验、社会经济理论、计量经济方法融合、提炼、发展而来的一种边缘性学科。

(2) 统计数据的收集是取得统计数据的过程，是进行统计分析的基础。离开了统计数据，统计方法就失去了用武之地，因此如何取得所需的统计数据是统计学研究的基本内容之一。整理数据是对统计数据的加工处理过程，目的是使统计数据系统化、条理化，符合统计分析的需要，它是介于数据收集与数据分析之间的一个必要环节。

(3) 统计数据的分析是统计学的核心内容，它是通过统计描述和统计推断的方法探索数据内在规律的过程。如果不用统计方法分析，统计数据仅仅是一堆数据，甚至杂乱无章，不能得出任何有益的结论。

因此，统计研究的过程描述，如图 1.1.1 所示。



图 1.1.1 统计研究的过程

统计作为一门科学，随着其应用的发展和深入，涉及大量的数据和复杂的模型，因此也需要越来越先进的计算机和数学。事实证明，数学和计算机的大量运用加速了统计学的发展，更新了统计的面貌。当前，统计是计算机数值计算的最重要用户之一。

1.1.2 统计学的研究对象与分类

1. 统计学的研究对象

一般来说，统计学的研究对象是自然、社会客观现象总体的数量关系。不论是自然领域还是社会经济领域，客观现象总体的数量方面都是统计学所要分析和研究的。统计学基本上是寄生的，靠研究其他领域内的工作而生存，这并不是对统计学的轻视，因为对很多寄主来说，如果没有寄生虫就会死；对有的动物来说，如果没有寄生虫，它就不能消化食物。因此，人类奋斗的很多领域，如果没有统计学，虽然不会死亡，但一定会变弱，这也从侧面要求统计工作者要知识渊博，最好是某一领域的专家。

统计学研究对象的特点有：

1) 数量性

统计学的研究对象是自然、社会经济领域中现象的数量方面，故统计学是定量分析学科，数字是统计的语言，数据资料是统计的原料。事物的数量是我们认识客观现实的重要方面，通过分析研究统计数据资料，研究和掌握统计规律性，就可以达到统计分析的研究目的。

2) 总体性

统计的数量研究是对总体普遍存在着的事实进行大量观察和综合分析，得出反映现象总体的数量特征和资料规律性。一般统计分析的目的并不局限于某个个体或者小团体，而是反映更大范围的群体在某个方面的特征和属性，比如通过抽查部分大学毕业生的就业来推断大学生的就业状况。一般情况下，自然、社会经济现象的数据资料和数量关系等是在一系列复杂影响因素下形成的，有些因素起主要作用，有些因素起次要作用。由于各种原因，每个个体具有一定的随机性质，而对于有足够多个体的总体来说又具有相对稳定的共同趋势，显示出一定的统计规律性。例如，人的身高有高低之分，各不相同，但经分析可发现，高个子父母的子女一般比矮个子父母的子女要高。

3) 具体性

统计研究的是具体现象的数量方面，而不是纯数量的研究，它具有明确的现实含义，这一特点也是统计学与数学的分水岭。数学是研究事物的抽象空间和抽象数量的科学，而统计学研究的数量是客观存在的、具体实在的数量表现，统计学研究的这一特点也决定了它的实用性。

4) 变异性

统计研究对象的变异性是指构成统计研究对象的总体各单位，除了在某一方面必须是同质的以外，在其他方面又要有差异，而且这些差异并不是由某种特定的原因事先给

定的，否则就没有必要进行统计分析研究了。例如，学生作为统计数据资料对象，每个学生性别、年龄、身高等方面会有不同表现，这样统计分析研究才能对其表现出来的差异探索统计规律性。

思考：什么是社会现象？什么是自然现象？

提示：社会现象与自然现象是一个相对的概念。太阳东升西落，四季轮回等都被称为自然现象，而社会现象强调人的参与，是指人类活动的产生、发展、变化密切联系的现象，例如农村儿童留守现象、欺诈现象、腐败现象等。

2. 统计学的分类

统计方法已被应用到自然科学和社会科学的众多领域，统计学也发展成为由若干分支学科组成的学科体系。依照不同的标准，统计学的分类也不同。

1) 按统计方法的构成分类

(1) 描述统计学 (descriptive statistics)，是通过图表或数学方法，对数据资料进行整理、分析，并对数据的分布状态、数字特征和随机变量之间关系进行估计和描述的方法。描述统计是一套处理和分析数据的基本方法和技术，可分为集中趋势分析、离散趋势分析及相关分析三大部分。

(2) 推断统计学 (inferential statistics)，是研究如何根据样本数据去推断总体数量特征的方法，它是在对样本数据进行描述的基础上，对统计总体的未知数量特征做出以概率形式表述的推断。例如，我们想研究教育背景是否会影响人的智力测验成绩。可以找 100 名 24 岁大学毕业生和 100 名 24 岁初中毕业生，采集他们的智力测验成绩。用推断统计方法进行数据处理，最后会得出类似这样的结论：“研究发现，大学毕业生组的成绩显著高于初中毕业生组的成绩，二者在 0.05 水平上具有显著性差异，说明大学毕业生组的智力测验成绩优于中学毕业生组。”值得注意的是，智力测试成绩是智商的主要标志，在相同条件下，成绩越高，智商也越高。如果条件不同，则没有可比性，因此，虽然大学毕业生组的智力测试成绩比初中毕业生组的高，但我们不能说这些大学毕业生的智商比初中毕业生的智商高，因为他们的教育背景不同。

思考：智商到底是什么？如何客观评估自己与竞争对手的智商？

2) 按统计方法的研究和应用分类

(1) 理论统计学，也称数理统计，指统计学的数学原理，主要研究统计学的一般理论和统计方法的数学理论。数理统计学家对目前广泛应用的大量统计模型有着重要贡献，然而这些似乎脱离某一两个具体应用领域的表面现象及使用的复杂数学工具，使得有些人认为统计（或数理统计）就是数学或数学的一个分支。实际上，从思维方式来说，统计和数学在研究目标和思想方法上是有差异的。数学以公理系统为基础，以演绎为基本思想方法的逻辑体系。它属于少数可以和世界具体事物无关的自成体系的学科。数学可以完全脱离实际而存在，而其他科学均是以实际事物为研究对象的。统计为各个领域服务，它以归纳为基本思想方式，而且归纳和演绎并用。现代统计学几乎用到了所有方面的数学知识，由于概率论是统计推断的数学和理论基础，广义地讲，统计学也应该包括

概率论在内。理论统计学是统计方法的理论基础，没有理论统计学的发展，统计学就不可能发展成为像今天这样一个完善的科学知识体系。因此，作为从事统计理论和方法研究的人员必须要有坚实的数学基础，否则将寸步难行，这也是很多人不喜欢理论统计学的原因之一。

(2) 应用统计学，是研究如何应用统计方法去解决实际问题的方法。在统计研究领域，相对来说，从事理论统计学研究的人是很少一部分，而大部分则是从事应用统计学研究的。

没有概率论，我们无法真正理解统计，没有理论统计学，我们也无法真正掌握应用统计学。理论牢固了，知识系统了，我们才能应用统计学解决复杂的问题。可见，要想掌握统计软件实现数据分析，就要掌握基本的概率统计理论。可能有人会说，专业的事情找专业人士做，但只有你懂点专业，你才能和专业人士交流。

1.1.3 统计规律

统计学是探索数据内在规律的一套方法。那么，什么是统计数据的内在规律呢？为什么统计方法能通过对数据的分析找到其内在的数量规律呢？下面我们通过几个例子进行说明。

1. 人口性别

众所周知，就单个家庭而言，新出生婴儿的性别可能是男，也可能是女。如果不限生育，多个子女的家庭可能全部是男孩，也可能全部是女孩。表面看，新生儿的性别好像没有任何规律，但如果对大量家庭的新生儿进行统计分析，就会发现：新生儿的男孩略多于女孩，男女比例大概为 107 : 100。为什么男婴的出生率会高于女婴呢？拉普拉斯从概率论的观点解释说：这是因为含 X 染色体的精子与含 Y 染色体的精子进入卵子的机会不完全相同。其实，从某种角度而言，女人 XX 染色体比男人 XY 染色体的可靠性高，这是因为 XX 可以看作并联系统，而 XY 可以看作串联系统。另外，由于雄性激素的作用，男人更易具有危险动作，如打架斗殴、酗酒。这样，在自然状态下，男人的死亡率会略高于女人，即使男婴出生率高点，但到了结婚年龄，二者比例很接近。进入中老年后，男性的死亡率仍然高于女性，导致男性的平均寿命低于女性，老年男性人数反而少于女性。

由于生育人口在性别上保持大致平衡，保证了人类社会的进行和发展，所以对人口性别的研究是统计学的起源之一，也是统计方法探得的数量规律之一。中国在社会主义初级阶段的某时期，重男轻女思想与计划生育政策的存在，导致了性别比例失调，希望我们能认真对待并解决这个问题。

2. 掷硬币游戏

在投掷一枚硬币时，既可能出现正面，也可能出现反面，预先做出确定的判断是不可能的，但是假如硬币均匀，直观上看，出现正面与反面的机会应该相等，即在大量的

试验中出现正面的频率应接近 50%。历史上有不少人做过抛硬币试验，其结果见表 1.1.1，从表中的数据可看出，随着试验次数的增加，出现正面的频率逐渐稳定于 0.5。

表 1.1.1 抛掷硬币试验记录

实验者	抛硬币次数	出现正面的次数	频率
德摩根 (De Morgan)	2 048	1 061	0.518 1
蒲丰 (Buffon)	4 040	2 048	0.506 9
费勒 (Feller)	10 000	4 979	0.497 9
皮尔逊 (Person)	12 000	6 019	0.501 6
皮尔逊	24 000	12 012	0.500 5

3. 英语字母的频率

在生活实践中，人们逐渐认识到：英语中某些字母出现的频率要高于其他字母。有人对各类典型的英语书刊中字母出现的频率进行统计，发现各个字母的使用频率相当稳定。这项研究对计算机键盘的设计（在操作方便的地方安排使用频率较高的字母键）、信息的编码（用较短的码编排使用频率最高的字母键）等都是十分有用的。

4. 最佳施肥量

在进行农作物试验时，如果其他试验条件相同，我们会发现某种粮食作物的产量会随着某种施肥量的增加而增加。在最初增加施肥量时，粮食产量的增加较快，以后增加等量的施肥量，粮食产量的增加逐渐减少。当施肥量增加到一定量时，粮食产量最高，这时如果继续增加，粮食产量反而会减少。粮食产量与施肥量的这种数量关系（边际效用递减）就是我们要探索的数量规律。如果我们从大量的试验数据中，用统计方法找到粮食产量与施肥量之间的数量关系，就可得到最佳施肥量，进而实现效益最大化。

上述例子说明：就一次观察或试验而言，其结果往往是随机的，但大量试验往往呈现出某种规律性，这种规律性称为统计规律性。利用统计方法可以探索出其内在的数量规律，因为客观事物本身是必然性与偶然性的对立统一，必然性反映了事物的本质特征和规律，偶然性反映了事物表现形式上的差异。如果客观事物仅有必然性的一面，则它的表现形式就会很简单，正是偶然性的存在，才使得事物的表现形式和必然的规律性之间产生偏差，从而形成了表面的千差万别，使得事物的必然性被掩盖在表面的差异中。这正如恩格斯所指出的：“在表面上是偶然性在起作用的地方，这种偶然性始终是受内部隐藏着的规律支配的，而问题只是在于发现这些规律。”概率论的任务是要透过随机现象的随机性揭示其统计规律性；统计学的任务则是通过分析带随机性的统计数据来推断所研究的事物或现象固有的规律性。二者的研究目的都是随机现象的统计规律，但其研究方法存在一定差异，概率论主要利用演绎方法，统计学主要利用归纳方法。

思考：你认为下面命题正确吗？请给出原因。

- (1) 父亲对性格的遗传更多。
- (2) 父亲与母亲对身高的遗传一样大。

1.1.4 统计的应用领域

学者不能离开统计而治学，政治家不能离开统计而施政，事业家不能离开统计而执业。目前，统计方法已被应用到自然科学与社会科学的众多领域，统计学已发展为由若干学科分支组成的学科体系。可以说，统计学几乎用到所有研究领域，比如政府部门、学术研究领域、日常生活中、企业管理中，因而也形成了众多的具有统计学应用性质的学科，如社会统计学、经济统计学、工业统计学、农业统计学、物理统计学、生物统计学、医药统计学、人口统计学、空间统计学，等等。现在，连纯社科领域的法律、历史、语言、新闻等也越来越重视对统计数据进行分析，国外的人文和社会学科普遍开设统计学的课程。可以说，统计方法与数学、哲学一样成为所有学科的基础。

例 1.1.1 (用统计识别作者) 1787—1788 年，三位作者 Alexander Hamilton, John Jay 和 James Madison 为了说服纽约人认可宪法，匿名发表了 85 篇著名的论文。这些论文中的大多数作者已经得到了识别，但是，其中的 12 篇论文的作者身份引起了争议。通过对不同单词的频数进行统计分析，得出的结论是：James Madison 最有可能是这 12 篇论文的作者。现在，对于这些存在争议的论文，认为 James Madison 是原创作者的说法占主导地位。

“红楼梦”后 40 回是否为曹雪芹所写？1985—1986 年，复旦大学李贤平教授带领他的学生做了这项工作，他们创造性的想法是将 120 回看成 120 个样本，然后确定与情节无关的虚词作为变量。为什么要抛开情节？这是因为在一般情况下，同一情节大家描述的都差不多，但由于个人写作特点和习惯的不同，所用的虚词（即变量）是不一样的。然后，统计出每一回里变量出现的次数，作为数据，利用多元分析中的聚类分析进行分类。分析结果然表明，120 回分属两类，即前 80 回为一类，后 40 回为一类，有说服力地证实了全书不是出自同一人的手笔。那么后 40 回是否为高鹗所写？同样，论证结果推翻了“后 40 回是高鹗一个人所写”。这个论证在红学界引发轰动，支持了红学界观点，但《红楼梦》的作者仍存在争议。

下面将给出统计在工商管理中的部分应用：

(1) 产品质量管理。质量是企业的生命，是企业持续发展的基础。统计质量管理已是质量管理的重要手段，在一些知名的跨国工商， 6σ 准则成为一个重要的管理概念，统计质量控制图已广泛应用于检测生产过程。

(2) 市场研究。企业要在激烈的市场竞争中取得优势，首先必须了解市场，而要了解市场，则需要做广泛的市场调查，取得所需的信息，并对这些信息进行科学的分析，以便作为生产和营销的依据，这些都需要统计的支持。

(3) 经济预测。企业要对未来的市场状况进行预测，经济学家常常要对宏观经济或某一行业进行预测，在进行预测的时候，最常用的方法就是利用各种统计信息和统计方

法。比如，企业要对市场的潜力进行预测，调整生产计划，最大化利润，这就需要利用市场调查取得数据，并进行统计分析。

(4) 人力资源管理。利用统计方法对员工的年龄、性别、受教育程度、工资等进行分析，并作为制订工资计划、奖惩制度的依据。

当然，统计不仅仅在工商管理中 useful，它已经渗透到自然科学和社会科学的各个领域，为多个学科提供通用的数据分析方法。从某种意义上说，统计仅仅是数据分析方法，必须与其他学科结合才能发挥作用，于是，统计工作者的知识面要宽广，最好擅长某一方面，比如生物、医学、农业、教育、经济等。

下面列出统计的一些应用领域（见表 1.1.2），让读者简单浏览并形成概念“统计学非常有用”。

表 1.1.2 统计的应用领域

精算	金融	农业	动物学	考古学
医学	生态学	教育学	计量经济学	管理学
人类学	赌博	地质学	心理学	质量控制
工业	工程	水文学	军事科学	物理学
市场营销学	地理学	语言学	分类学	宗教研究

统计应用上的两个极端是：不用或几乎不用统计、简单问题复杂化。在统计应用中，这两个极端都是不可取的。简单的方法不一定没用，复杂的方法也不一定有用。正如有的学者所说的，最简单的模型往往是最有用的，统计应该恰当地应用到它能起作用的地方。同时，我们不能把统计神秘化，更不能歪曲统计，把统计作为掩盖事实的陷阱。统计是一种从数量上认识客观世界的有力工具，但是若运用不当，即使是科学技术也有可能得出错误的结论，甚至会成为谬论的护身符。实际上，在我们周围，误用统计数字或滥用统计方法的现象不乏其例，这不能不影响到统计科学的严肃性和统计分析的准确性。因为曾经对国家统计局公布的居民收入增长数据表示不解，某网友在某大网站的博客上发明了一个“被增长”的新词。所谓“被增长”，即实际没有增长，但在统计数据中却增长了，使我们都被统计数字“幸福”地笼罩着。这一感受得到了绝大多数人的认同，同时也得到了央行最新城镇储户问卷调查结果的印证。这是因为居民收入的均值往往显著高于中位数。精心设计的模型从根本上说，是经过审慎挑选的有关现实的一组谎言，或者说经过审慎挑选的有关现实的部分真理。更一般地讲，世界上存在着三种谎言：谎言，该死的谎言，统计数字。下面举例介绍统计欺骗中常用手法供读者鉴别。

(1) 有偏样本：有一个装着红、白两色小球的箱子，如果你想要准确知道这个箱子中两种小球的数量，你唯一能做的只有一颗一颗地数小球。用一种更简单的方法也可以估计红球的数量：抓一把小球，假定手中红球所占比例与箱子中红球所占比例相同，只要数一数手中的小球即可。如果你的样本足够大且选择方法正确，在大多数情况下，它能够很好地代表整体。但是，如果以上两个条件不满足，这样的样本就不值得一提。不幸的是，我们所看到的，或者我们自以为了解的许多事物，往往都是根据类似样本所得