



普通高等教育“十一五”国家级规划教材
全国高等院校经济管理类专业基础课系列规划教材

统计学

(第二版)

张德存◎主编



科学出版社
www.sciencep.com

内 容 简 介

本书主要介绍了统计数据搜集、整理和分析的方法,语言通俗易懂,并通过大量的例题和习题通俗讲解,使学生对统计学知识的学习更加容易、便捷,能收到理想的学习效果。本书共分11章,分别介绍统计学的一些基本问题、统计数据的搜集、统计数据描述、概率和概率分布、参数估计、假设检验、非参数估计、方差分析、相关与回归、时间数列分析、统计指数。

本书适合作为经济管理专业本科生教材,也可作为管理类和财经类专业的选修教材,还可作为广大统计工作者和爱好者的入门教材和参考书。

图书在版编目(CIP)数据

统计学/张德存主编. —2版. —北京:科学出版社,2009

(普通高等教育“十一五”国家级规划教材·全国高等院校经济管理类专业基础课系列规划教材)

ISBN 978-7-03-024912-8

I. 统… II. 张… III. 统计学-高等学校-教材 IV. C8

中国版本图书馆CIP数据核字(2009)第110025号

责任编辑:田悦红 / 责任校对:王万红

责任印制:吕春珉 / 封面设计:东方人华平面设计部

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>

张主印刷厂印刷

科学出版社发行 各地新华书店经销

*

2004年9月第一版 开本:B5(720×1000)

2009年7月第二版 印张:25 1/4

2009年7月第二次印刷 字数:496 000

印数:5 001—8 000

定价:35.00元

(如有印装质量问题,我社负责调换〈环伟〉)

销售部电话 010-62134988 编辑部电话 010-62135763-8007 (HF02)

版权所有,侵权必究

举报电话:010-64030229;010-64034315;13501151303

第二版前言

在国内外统计教材出版繁荣的今天互联网纵观 20 世纪统计学的发展,统计数据收集与使用的客观环境发生了很大的变化,特别是计算机与互联网的使用,数据的收集存储、信息交换的客观条件有了质的变化。现代社会所表现出的数据在其容量规模、次元、对时空的依存性、不完全性、不均一性、复杂性及相关性等,均与以往完全不同。社会经济的多元化、金融交易的多样化、国际市场间资本移动的迅猛及电子商务的出现,均对统计学教学活动产生影响。在这种变化中,21 世纪的统计学课程内容应怎样更新,统计学的应用应该如何发展,统计学的基本框架已经形成的时代背景与当今计算机大量普及和统计软件盛行的现实落差如何调和,这些应该是 21 世纪统计学研究与应用的重大课题。在“大统计”的潮流中,在学生追求务实高效的时代,统计学在内容、形式上都面临着有一个极好的发展和改革契机。如何根据新情况和新发展来构建新型的统计学内容体系是修订本书的出发点。

本书自 2004 年 8 月由科学出版社出版以来,很好地满足了使用院校的教学需求。由于质量较高,在教学中反映较好,所以积极参与了 2007 年教育部组织的普通高等教育“十一五”国家级教材规划补充申报工作。经科学出版社申报、专家评审、网上公示,本书于 2008 年 2 月获评为普通高等教育“十一五”国家级规划教材。现在对第一版进行修订,出第二版。

第二版的修订传承了第一版的特点。21 世纪是知识经济的时代,信息技术、计算机技术对统计学理论与方法的发展产生了巨大的推动作用。知识创新是时代的基本特征。统计学理论与方法的创新及应用必将使众多领域和学科发展的价值得到充分体现。统计学与其他学科的紧密结合将产生新的边缘学科,许多学科的发展将依赖于统计理论与技术的应用,更为复杂数据的处理方法将成为统计理论界研究的热点,实用快捷的统计方法与技术将更加普及。本书在传授系统的统计知识的同时能有效地训练学生的数据采集和数据分析能力,以学生为中心开展统计方法的训练,以采集数据和分析数据活动为主要教学手段,以任务为重要的训练步骤,从方法上构建了内容体系。本书是按照“提出研究目的—根据研究目的决定收集什么数据—对收集到的数据进行汇总整理—对数据进行描述性分析—对数据进行深层次的推断性分析”这样的新框架来安排各章节内容的,各章之间的关联性紧密,具有很强的逻辑性。

本书此次修订分工如下:由张德存设计修订版本的框架,第 1 章、第 4~9

章及附录由张德存修订，第 2~3 章、第 10~11 章由李雪梅修订，全书由张德存总纂定稿。本书在修订过程中参考了大量的统计学文献，在此对其作者表示深深的谢意。

多年来，本书的出版一直得到科学出版社的大力支持，在使用过程中不少兄弟院校提出了宝贵的意见，在此向相关人员一并表示衷心的感谢。

由于编者水平所限，本书不足之处在所难免，恳请广大读者不吝赐教。

第一版前言

面向 21 世纪的信息社会，仅有理学性质的理论研究是很难解决实际问题的。应用于经济管理中的统计学，与以往被按照研究对象或研究方法分门别类的经济学、管理学、计算机科学互相渗透，互相结合，特别是随着信息科学的进步，统计应用的范围越来越广。统计数据的计算变得简单了，但根据研究对象的不同，数据处理及数据采集挖掘的方法呈现出多样化，统计分析方法也相对复杂化、专业化。统计学的应用不仅要不断提高理论统计学的基本素质，还要注重掌握经济学的理论、金融交易制度及金融理论、管理科学的理论与计算机的技术方法。统计理论与应用的紧密结合显得比以往任何一个时期都更为迫切，更加重要。

原美国统计学会会长伯克斯 (G. E. P. Box) 在他的学术报告 (JASA, 1979, Vol. 74) 中曾经指出：以往统计学者的分布表现为一边是以研究统计理论为主的统计学者，另一边是以研究统计应用为主的统计学者，即呈现出中间结合部分偏低较薄弱的双峰分布。但理想的分布应是中央部分较高较宽厚的分布，即以统计理论与统计应用相结合为主所构成的单峰分布。此话较精辟地概括了 20 世纪统计学的现状及问题，值得统计学者的深思。

纵观 20 世纪统计学的发展，统计数据收集与使用的客观环境发生了很大的变化，特别是通过计算机与互联网的使用，数据的收集存储、信息交换的客观条件有了质的变化，现代社会所表现出的数据在它的容量规模、次元、对时空的依存性、不完全性、不均一性、复杂性及相关性等，均与以往完全不同。社会经济的多元化、金融交易的多样化、国际市场间资本移动的迅猛以及电子商务的出现，均对我们的日常生活产生影响。在这种变化中，21 世纪的统计学理论应怎样更新？统计学的应用应该如何发展？统计理论的基本框架已经形成的时代背景与当今计算机大量普及的现实落差如何调和？这些应该是 21 世纪统计理论研究与应用的一个重大的课题。针对我国统计理论和实践的发展，我们提出我国统计实践应该关注的一些问题：

1. 实际统计工作中迫切需要解决的抽样调查理论和方法问题

- 抽样调查如何满足多目标、多层次的需要
- 小样本理论及其应用
- 小型工商业与小业主抽样调查问题
- 地理信息与抽样调查
- 人口抽样调查问题

- 住户抽样调查问题
- 2. 中国宏观经济模型理论、方法及实证研究
 - 适合我国国情的宏观经济模型
 - 宏观经济政策效应的统计分析
- 3. 国家经济安全问题的统计研究
 - 金融风险的统计分析
 - 社会保障问题的统计研究
- 4. 数据分析与数据质量问题研究
 - 数据分析
 - 数据质量
- 5. 国民经济核算问题研究
 - 国民经济核算不同方法的统一性问题
 - 国家核算与地方核算的统一性问题
 - 绿色 GDP 核算
 - 投入产出核算
- 6. 政府统计改革问题研究
 - 统计指标体系研究
 - 政府统计与部门统计的衔接
 - 统计标准化问题
 - 调查方法体系问题
 - 管理体制问题
- 7. 国家、地区竞争力研究
 - 国家竞争力研究
 - 地区竞争力研究
- 8. 统计教育改革问题研究
 - 统计教育体系
 - 社会统计教育
- 9. 信息技术在统计中的应用
 - 统计软件
 - 统计信息自动化
 - 统计与网络
- 10. 统计新理论、新方法及其应用
 - 空间统计学的发展和应用
 - 知识经济与统计
 - 其他统计新理论

实践问题的解决需要理论的支持,本书就是在这样的指导思想下编写的。同时统计学作为应用性最广的学科,大部分科系的大学生都要学习统计课程。因此,在不同科系学生的教育中,统计方法的研究已具有重要的角色。本书的宗旨是让统计学变得更有趣,让读者能够更广泛、更容易应用它。本书有三大特色支持此宗旨:其例子和练习的内容均来自真实生活,本书介绍清晰和简洁,本书的教学方法完善。

本书以财经类专业本科学生为主要对象,专科学生使用本书时,部分章节的扩展应用和难度较大的内容可以略而不讲。此外,本书也可供统计工作者和企业管理人员自学参考。本书设计课时为每周4课时,一学期18周共72课时为课堂讲授时间。

本书编写分工如下:全书共分14章由张德存设计编写框架,第1章、第4~9章及附录由张德存撰写,第2~3章、第10~11章由李雪梅撰写,全书最后由张德存总纂定稿。本书在编写过程中参考了大量的统计学文献资料,在此对相关作者表示深深的谢意。

由于编者水平的限制,加之时间仓促,本书难免存在不足之处,恳请专家、读者赐教。

目 录

第二版前言

第一版前言

第 1 章 导论	1
1.1 统计实践历史与统计学的产生和发展	1
1.2 统计与统计学	6
1.3 统计学的几个基本范畴	17
1.4 统计学的方法	22
习题	24
第 2 章 统计数据的搜集	27
2.1 统计数据的类型	27
2.2 统计数据搜集的组织形式	31
2.3 统计数据的搜集方法	40
2.4 统计数据的搜集方案与问卷设计	46
2.5 数据质量	56
习题	58
第 3 章 统计数据描述	60
3.1 统计数据的预处理	61
3.2 品质数据的描述	62
3.3 定量数据的描述	66
3.4 定量数据描述的数值方法	79
习题	98
第 4 章 概率和概率分布	104
4.1 随机变量及其分布	104
4.2 抽样分布	112
4.3 正态分布	117
4.4 几个重要的概率分布	120
习题	125
第 5 章 参数估计	127
5.1 参数估计的基本概念	127
5.2 参数估计的基本方法	131

5.3 总体均值和总体比例的区间估计	137
5.4 两个总体均值和两个总体比例之差的置信区间估计	143
习题	147
第6章 假设检验	151
6.1 假设检验的一般问题	151
6.2 一个正态总体参数的假设检验	155
6.3 两个正态总体参数的假设检验	160
6.4 假设检验中的其他问题	164
习题	166
第7章 非参数估计	168
7.1 非参数估计的特点	168
7.2 单样本非参数检验	169
7.3 两样本非参数检验	182
7.4 秩相关分析	190
习题	199
第8章 方差分析	203
8.1 单因素方差分析	204
8.2 无交互作用的双因素方差分析	210
8.3 有交互作用的双因素方差分析	215
习题	218
第9章 相关与回归	221
9.1 相关与回归概述	221
9.2 相关关系的测度	225
9.3 一元线性回归分析	231
9.4 多元线性回归模型	243
9.5 曲线回归	253
习题	260
第10章 时间数列分析	265
10.1 时间数列概述	265
10.2 时间数列的水平分析指标	267
10.3 时间数列的速度分析指标	272
10.4 时间数列的分解	279
习题	294
第11章 统计指数	299
11.1 指数的意义与分类	299

11.2 综合指数	302
11.3 平均指数	305
11.4 指数数列	318
11.5 几种常用的经济指数	320
习题	324
附录 有关统计的应用数据	331
1 二项分布表	331
2 泊松分布表	343
3 标准正态分布表	345
4 正态分布分位数表	348
5 t 分布表	351
6 χ^2 分布表	353
7 F 分布表	355
8 Kolmogorov-Smirnov 检验临界值表	367
9 符号检验界域表	369
10 两样本 K-S 检验统计量	370
11 游程检验临界值表	373
12 Mann-Whitney-Wilcoxon 秩和分布	375
13 Spearman 秩相关系数检验临界值表	379
14 相关系数 $\rho=0$ 的临界值表	382
15 正交表	383
16 Kendall τ 检验临界值表	385
17 组合数	386
18 $m=n$ 时 Smirnov 检验临界值	387
19 $m \neq n$ 时 Smirnov 检验临界值	388
20 随机数字表	390
参考文献	391

第 1 章 导 论

学习目标

了解统计的发展历史；掌握统计的含义；掌握统计学的基本范畴；了解统计学的基本方法。

1.1 统计实践历史与统计学的产生和发展

1.1.1 统计实践历史的追溯

统计作为一种人类认识自然、改造自然的实践活动在原始社会就已产生，原始人掌握了在洞穴石壁上划线、绳上打结等多种计数方法，可说是最早的统计实践活动。我国的统计实践，滥觞于伏羲画八卦，而在公元前两千多年大禹治水时期，依据山川土质、人口和物产、贡赋的多寡，把全国分为九州，编纂《禹贡》九州篇，就形成了统计的雏形。公元前三百多年前，在商鞅的调查研究思想中，已有了各种范围的人口调查登记制度和人口按年龄、职业的分组以及各种数量的对比分析。到 17 世纪中叶，随着统计实践的发展，“统计”一词已约定俗成。我国统计实践历史悠远，只是由于缺乏从事专门研究的学者，使其未能成为一门系统的科学。

国外的统计实践，始于公元前 3050 年，埃及建造金字塔时期，为了征集建筑费用，对全国的人口与财产进行了普查。现代统计发端于工业化、城市化迅速发展的 17 世纪欧洲。由于工业化和城市化导致生产活动规模急剧扩张，城市人口规模越来越大，民族国家迅速兴起，这些都使得对于社会、经济、人口等领域进行宏观定量分析和研究成为一门新兴科学。19 世纪中叶以后，统计得到了日益广泛的应用。

目前，我国在新形势下，为了更好地发展社会主义市场经济，提高社会生产力水平，统计作为一种工具，越来越发挥出无可替代的作用。

1.1.2 统计学的产生和发展

统计学的产生和发展是伴随着资本主义的古典哲学、古典政治经济学和空想社会主义的产生和发展完成的。一般公认，统计学的历史从 17 世纪 60 年代算起，

到现在已有 350 多年。史料说明：统计学起源于定量地说明和研究社会经济问题。它是帮助决策、辅助管理的工具。统计学的发展过程大体可划分为四个阶段。

1. 第一阶段称之为“城邦政情”阶段

“城邦政情”（matters of state）阶段始于古希腊的亚里士多德撰写“城邦政情”或“城邦纪要”。他一共撰写了一百五十余种纪要，其内容包括各城邦的历史、行政、科学、艺术、人口、资源和财富等社会和经济情况的比较、分析，具有社会科学特征。“城邦政情”式的统计研究延续了一两千年，直至 17 世纪中叶才逐渐被“政治算术”这个名词所替代，并且很快被演化为“统计学”（statistics）。统计学依然保留了城邦（state）这个词根。

2. 第二阶段称之为“政治算术”阶段

“政治算术”（political arithmetic）阶段与“城邦政情”阶段并没有很明显的分界点，二者也没有很本质的差别。“政治算术”的特点是统计方法与数学计算和推理方法开始结合。分析社会经济问题的方式更加注重运用定量分析方法。可以用 1690 年英国古典政治经济学的奠基人 W. 配弟（W. Petty）出版《政治算术》一书作为这个阶段的起始标志。W. 配弟在书中强调了自己的立论方法，他说：“我进行这种工作所使用的方法，在目前还不是常见的。因为我不采用比较级或最高级的词语进行思辩式的议论，相反的，采用了这样的方法，（作为我很久以来就想建立的政治算术的一个范例）即用数字、重量和尺度来表达自己的问题，只进行诉诸人们的感觉的议论，借以考察在自然中有可见根据的原因。”

用数字、重量和尺度将社会经济现象数量化的方法是近代统计学的重要特征。因此，W. 配弟的《政治算术》被后来的学者评价为近代统计学的来源，W. 配弟本人也被评价为近代统计学之父。刺激近代统计学发展的主要原因是战争和工业革命。当时政治家需要有人专门致力于收集和分析社会、经济和政治方面的统计数据，帮助他们决策和管理国家。W. 配弟写作《政治算术》大约在 1671—1676 年之间，正值爆发第三次英荷战争（1672—1674 年）期间。与此同时，他还撰写了《爱尔兰的政治解剖》一书。在 W. 配弟的心目中，政治算术和政治解剖这两种方法是二位体的。有些学者认为《政治算术》是“比较统计”（comparative statistics），《爱尔兰的政治解剖》是“个别统计”（individual statistics）。

构成“政治算术”方法的核心是“数字”。W. 配弟在书中使用的数字有以下三类。第一类是对社会经济现象进行统计调查和经验观察得到的数字。因为受历史条件的限制，书中通过严格的统计调查得到的数据少，根据经验得出的数字多。第二类是运用某种数学方法推算出来的数字。这类数据 W. 配弟在书中使用

很多,其推算方法可分为三种:“①以已知数或已知量为基础,循着某种具体关系进行推算的方法;②通过运用数字的理论性推理来进行推算的方法;③以平均数为基础进行推算的方法”。第三类是为了进行理论性推理而采用的例示性的数字。W. 配第把这种运用数字和符号进行的推理称之为“代数的算法”。从W. 配第使用数据的方法看,“政治算术”阶段的统计学已经比较明显地体现了“收集和分析数据的科学和艺术”特点,统计实证方法和理论分析方法浑然一体,这种方法即使是现代统计学也依然继承。

另外英国的J. 格朗特(J. Graunt)在1662年出版了《关于死亡表的自然观察》。他通过大量观察的方法,研究并发现了人口与社会现象中重要的数量规律性。法国的B. 帕斯卡尔(B. Pascal)和P. 费马特(P. Fermat)在数学家们对机会游戏研究的基础上,他们通过通信的方式,将赌博中出现的各种具体问题,归结为一般的概率原理,为概率论和统计学的发展奠定了重要的基础。特别值得提出的是比利时统计学家L. A. J. 凯特勒(L. A. J. Quetelet),他对统计理论方面最大的影响是把概率与统计学相结合,从而提出了关于统计学的新概念,他的主要功绩在于使统计方法获得普遍应用。凯特勒对各种科学均有造诣,像天文学、数学、物理学、人类学、行政、生命及社会统计学、气象学等,并努力使人们把统计学理解为一种可用于任何学科的一般研究方法。在“政治算术”阶段出现的统计与数学的结合趋势逐渐发展形成了“统计分析科学”。

3. 第三阶段称之为“统计分析科学”阶段

19世纪末,欧洲大学里开设的“国情纪要”或“政治算术”等课程名称逐渐消失,代之而起的是“统计分析科学”(science of statistical analysis)课程。当时的“统计分析科学”课程的内容仍然是分析研究社会经济问题。“统计分析科学”课程的出现是现代统计发展阶段的开端。

经过历代统计学家的努力和发展,到19世纪末建成了古典统计学的基本框架。20世纪初,1908年,英国的W. S. 戈赛特(W. S. Gosset)用笔名Student(“学生”氏)发表了关于 t 分布的论文,这是一篇在统计学发展史上划时代的文章。它创立了小样本代替大样本的方法,开创了统计学的新纪元。现代统计学的代表人物首推比利时统计学家凯特勒,他将统计分析科学广泛应用于社会科学、自然科学和工程技术科学领域,因为他深信统计学是可以用于研究任何科学的一般研究方法。

后来,统计学家R. A. 费雪(R. A. Fisher)给出了 F 统计量、极大似然估计、方差分析等,J. 内曼(J. Neyman)和E. S. 皮尔逊(E. S. Pearson)提出了置信区间估计和假设检验,A. 沃尔德(A. Wald)发明了序贯抽样和统计决策函数,到20世纪50年代,构筑了现代统计学的基本框架。20世纪50年代以后,统计学进入了一个全新的发展阶段,新的研究领域层出不穷,像多元统计分

析、非参数统计、现代时间序列分析、线性统计模型等。

4. 第四阶段称之为“统计学帝国主义”（imperialistic stage of statistics）阶段

从世界范围看，自 20 世纪 60 年代以后，统计学的发展有三个明显的趋势：第一，随着数学的发展，统计学依赖和吸收的数学方法越来越多；第二，向其他学科领域渗透，或者说以统计学为基础的边缘学科不断形成；第三，随着统计学应用日益广泛和深入，由于受计算机和新兴科学的影响，使统计学越来越依赖于计算技术，成为数量分析的方法论科学。这一时期统计学的研究和应用范围越来越广，如海洋学中水面的高程、电磁能（特别是无线电波）的流量、工业过程的状态、生物的状态，都能用计算机收集，而没有计算机这些是不实际或不可能的。计算机节省了大量人力，特别是在同样输出的重复计算上，例如，在计算多元分析的相关系数和其他检验统计量。由于计算机软件 SPSS 和 SAS 可用于所有通常的检验（见表 1.1），特别是关于方差分析，则节省更多时间。由于利用软件很方便。2003 年度诺贝尔经济学奖授予两位著名计量经济学家 R. F. 恩格尔（R. F. Engle）和 C. 格兰杰（C. Granger）。20 世纪 80 年代两位获奖者发明了新的统计方法来处理许多经济时间数列中两个关键属性：时变性（time-varying volatility）和非平稳性（nonstationarity），R. F. 恩格尔的研究方向主要是利率、汇率和期权的金融计量分析，提出谱分析回归等创新性统计方法；C. 格兰杰的研究涉及统计和经济计量学，特别是时间序列分析、预测、金融、人口统计学、方法论等领域。这足以证明统计方法应用的领域越来越广泛，可以应用于各行各业的数据分析，使统计学成了一门“万能”的方法论学科。美国 SCIENCEFY 有一篇文章列出近百年来最有用的科学，统计高居前十名。

统计学新的分支或专门化和应用已经被发展了决策论、时间序列，多元分析、经济计量学、博弈论、临床试验、非参数推断、序贯分析、数学生物分类学、可靠性。例如，W. G. 科克伦（W. G. Cochran）的实验设计理论、Th. W. 安得森（Th. W. Anderson）的复变数分析等。数理统计及其应用正在继续发展和扩大。

D. 萨尔斯伯格（D. Salsburg）在《统计，改变了世界》序言中讲到：17 世纪以来，统计早已跨越这类政府统计的范畴，静悄悄地渗入数理、生物、人文、工程等领域，甚至日常生活里也处处可见统计的踪迹。像民意调查、企业成效的评估等，无一不和统计相关。统计事实上已经改变了我们对自然、心智以及社会的看法，这场静默革命既深刻且广泛，可惜的是一般人虽然身在此山中，却认识不深。

表 1.1 统计软件应用和解决数据处理的领域

反应变量个数	自变量性质	反应变量性质	应采用的方法	SAS 软件	Stata 软件	SPSS 软件
1	无自变量 (1个总体)	连续且正态	单样本 t 检验	SAS	Stata	SPSS
		有序或连续	单样本中位数检验	SAS	Stata	SPSS
		二分类	二项检验	SAS	Stata	SPSS
		分类	卡方优度拟合 t	SAS	Stata	SPSS
	1个自变量 2个水平 (组间独立)	连续且正态	2 独立样本 t 检验	SAS	Stata	SPSS
		有序或连续	Wilcoxon—MannWhitney 检验	SAS	Stata	SPSS
		分类	卡方检验	SAS	Stata	SPSS
			Fisher 精确检验	SAS	Stata	SPSS
	1个自变量, 2个或以上水平 (组间独立)	连续且正态	单因素 ANOVA	SAS	Stata	SPSS
		有序或连续	Kruskal Wallis	SAS	Stata	SPSS
		分类	卡方检验	SAS	Stata	SPSS
	1个自变量 2水平 (组间相关/配对或配伍)	连续且正态	配对 t 检验	SAS	Stata	SPSS
		有序或连续	Wilcoxon 符号秩检验	SAS	Stata	SPSS
		分类	McNemar	SAS	Stata	SPSS
	1个自变量, 2个或以上水平 (组间相关/配对或配伍)	连续且正态	单因素重复测量 ANOVA	SAS	Stata	SPSS
		有序或连续	Friedman 检验	SAS	Stata	SPSS
		分类	重复测量 logistic 回归	SAS	Stata	SPSS
	2个或以上自变量 (组间独立)	连续且正态	析因 ANOVA	SAS	Stata	SPSS
		有序或连续	—	—	—	—
		分类	析因 logistic 回归	SAS	Stata	SPSS
	1个连续性自变量	连续且正态	相关	SAS	Stata	SPSS
			简单线性回归	SAS	Stata	SPSS
		有序或连续	非参数相关	SAS	Stata	SPSS
		分类	简单 logistic 回归	SAS	Stata	SPSS
	1或多个连续性自变量和/或1或多个分类自变量	连续且正态	多重 logistic 回归	SAS	Stata	SPSS
			协方差分析	SAS	Stata	SPSS
		分类	多重 logistic 回归	SAS	Stata	SPSS
			判别分析	SAS	Stata	SPSS
2个或以上	1个自变量, 2个或以上水平 (组间独立)	连续且正态	one—way MANOVA	SAS	Stata	SPSS
2个或以上	2个或以上自变量	连续且正态	多变量多重线性回归	SAS	Stata	SPSS
2组变量	0	连续且正态	典型相关	SAS	Stata	SPSS
2个或以上	0	连续且正态	因子分析	SAS	Stata	SPSS

1.2 统计与统计学

1.2.1 统计与统计学的含义

统计一词，一般可以有三种不同的含义，即统计工作、统计数据数据和统计学。

1. 统计工作

统计工作，是统计设计、统计调查、统计整理、统计分析和研究统计数据数据的工作过程。统计工作在人类历史上出现得比较早。随着历史的发展，统计工作逐渐发展和完善起来，使统计成为国家、部门、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。

一般而言，任何统计工作都要经过统计设计、统计调查、统计整理、统计分析四个阶段。

（1）统计设计

统计设计是统计工作实际展开前所作的计划性安排，包括统计方案的全面设计，又包括人财物等各种资源的安排。一项统计工作能否顺利进行与统计设计阶段工作是否全面、周密有很大的关系。

（2）统计调查

统计调查充分反映了统计工作的特点，是深入实际进行调查研究的重要环节。统计调查与其他调查活动不同点在于精心设计的统计调查方法的应用更能够保障调查活动起到事半功倍的效果。统计调查方法分为普查、重点调查、典型调查和抽样调查四种，这四种方法各有特点。普查活动搜集的数据最全面、详细、准确，但缺点也很明显，那就是费钱、费时、费人工，因为普查要求对每一个调查对象都要调查到，如人口普查，国家每次人口普查需要消耗大量的社会资源。重点调查和典型调查是我们短期迅速估查的好方法，但结果难以定量化，而且容易产生无法控制的误差，随意性很大。抽样调查是最经常采用的一种调查方法，原因在于它具有时效强、低成本、高质量、科学性的特点，虽然抽样调查的结果存在误差，但这个误差是我们可以接受、可以控制的。关于抽样调查有一整套科学方法，是统计学的重要分支。当然，要切实做好统计调查这项工作也决非易事，除了要掌握方法外，人员的责任心至关重要。

（3）统计整理

统计整理是对通过统计调查搜集到的原始数据统计数据根据研究的需要进行分类、汇总等活动。统计整理是一个对统计调查结果去粗取精、去伪存真的过程。要对统计调查搜集到的原始数据进行“把关”，修匀、剔除一些无效问卷。

统计整理后的结果就是我们通常看到的各种统计数据了。

(4) 统计分析

统计分析就是通过搜集到的统计数据来揭示其内在的一些规律性的特点。这一阶段充分体现了统计学的精华,主要表现为各种统计分析方法的应用,如统计指标、时间序列分析、统计指数、回归分析等,这实际也是我们这门课程要讲解的重点内容。

我国县以上各级政府均设有专门的统计部门和机构。国家统计局系统是自下而上成体系的,包括县以上各级统计局和三支统计调查的专业队伍(城市社会经济调查队、农村社会经济调查队、企业社会经济调查队,这是我国主要社会经济发展的数据搜集、整理、分析的专业化队伍)。每年大量的社会经济发展的宏观数据他们是最权威的发布部门。

2. 统计数据

统计数据 (statistical data),是指可用以推导出某项结论的一些事实或数字。统计数据一般由元素 (element)、变量 (variable) 和观测 (observation) 三部分组成。

(1) 元素

我们所研究的对象由各元素组成。统计数据就是关于各元素的有关特征的信息。表 1.2 中的元素就是各员工。

表 1.2 员工个人数据

姓名	性别	年龄	身高/m	体重/kg	民族	公司服务年限/年	受教育年限/年
甲	男	33	1.85	65	汉	3	18
乙	女	25	1.65	55	回	2	16
丙	男	26	1.72	60	满	1	15
丁	女	35	1.60	53	回	4	16
戊	男	32	1.83	68	汉	2	19

(2) 变量

变量是关于元素的一种属性或特征。

表 1.2 中年龄是各员工的一个特征,这些特征对于不同元素取不同的结果,所以称为变量。其中结果用数字表示的,称为定量变量 (quantitative variable);不可用数字表示的,称为定性变量 (qualitative variable)。

表 1.2 中年龄、身高、体重、公司服务年限、受教育年限是定量变量,性别、民族为定性变量。

(3) 观测

一次观测的结果是指数据中关于某一元素所有各变量的信息。表 1.2 中的每行都表示一次观测的结果。

当一项统计研究工作开始时,首先要明确所研究的问题以及为此需要哪些数