



21世纪高等院校教材

统计学基础

耿修林 编著



科学出版社

www.sciencep.com

21 世纪高等院校教材

统 计 学 基 础

耿修林 编著

科 学 出 版 社

北 京

内 容 简 介

本书较为系统地介绍了统计学的基本原理和常用的统计数据分析方法,内容包括:数据字计算及资料采集及质量诊断、数据资料描述中的图表制作、样本特征数应用、统计量的抽样分布、样本参数推断、非参数检验、方差分析与正交试验设计、相关与回归分析、时序分析与动态预测等。另外,对如何利用电子表格处理统计数据也进行了相应的说明。

本书可作为工商管理、财经类本科专业学生的教学教辅材料,也可作为工商管理硕士及从事社会经济实际统计工作人员的参考材料。

图书在版编目(CIP)数据

统计学基础/耿修林编著. —北京:科学出版社, 2010.8

(21世纪高等院校教材)

ISBN 978-7-03-028845-5

I. ①统… II. ①耿… III. ①统计学 IV. ①C8

中国版本图书馆CIP数据核字(2010)第170652号

责任编辑:王伟娟 / 责任校对:郑金红

责任印制:张克忠 / 封面设计:耕者设计工作室

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>

丽源印刷厂印刷

科学出版社发行 各地新华书店经销

*

2010年8月第一版 开本: B5(720×1000)

2010年8月第一次印刷 印张: 15 1/4

印数: 1—3 000 字数: 304 000

定价: 28.00元

(如有印装质量问题, 我社负责调换)

前 言

统计学是一门实用性很强的学科，无论在社会科学领域还是在自然科学领域都有重要的应用，它能为相关学科的定量研究提供数据资料采集方面的指导，也能为后续的数据处理和实证提供必要的分析工具。正因为如此，在1992年国家技术监督局颁布的《学科分类与代码》中，明确地把统计学和数学、经济学等并列为一级学科。今天，对于财经、工商管理专业的学生，掌握现代统计学的基础知识和方法，不仅能丰富自己的知识结构，同时也是专业素养的重要表现。

本书共由十章组成，除第一章是介绍统计学含义和现代统计学性质这些基本问题外，其余各章编写思路大致按照统计活动的逻辑程序展开。其中，第二章介绍了统计数据的来源，并对统计数据的种类和质量诊断问题进行了说明；第三章、第四章属于统计数据初步描述分析的内容，分定性数据和定量数据介绍了常用的图表显示和特征数字的计算；第五章是样本推断的理论基础，着重讲解了正态分布的性质和应用，介绍了在正态分布基础上导出的三大推断分布，以及常用的样本统计量的抽样分布；第六章至第十章介绍了统计推断和分析方面的内容，包括参数估计和假设检验、常用的非参数检验及其应用、方差分析与正交试验设计、相关和回归分析、古典时间序列分析等。另外，还对如何使用 Excel 处理统计数据给出了相应的介绍。

这本书是统计学的基础性读物，它以财经、工商管理专业的学生为对象，综合介绍了统计学的知识在这些学科中的应用。根据长期教学实践的体会，在本书的编写过程中，我们注意了以下几个方面的要求：

第一，尽量使用深入浅出、通俗易懂的语言阐述统计学中的基本概念，以便使初学者和经济管理类学生能够建立直观的理解。

第二，注意引导统计方法的应用。对经济管理专业的学生来说，选修统计学的目的是为专业课程的学习和研究准备必要的工具，而不在于培养自己的数学演算能力，所以在本书的编写过程中，我们有意识地

选编了一些统计学应用的典型例子，力求讲清楚统计方法的应用过程和基本技巧。

第三，在本书编写过程中，注意吸收当今统计学发展的一些新的、具有实用价值的成果。

第四，统计学的广泛应用同计算机软件的普及密不可分。对实际中的研究课题，如果完全依赖手工开展数据分析，不仅工作量浩大，有时也是行不通的。这表明，学习统计学若不同计算机软件的使用结合起来是不会有前途的，基于这样的考虑，在本书的各个章节都针对相应的统计方法，介绍了 Excel 中的“数据分析”功能和使用方法。

本书的编写过程中，我们参考了国内外许多优秀的教科书，在此一并表示诚挚的谢意。尽管在编写过程中，我们付出了很大的努力，但不足乃至错误可能难以避免，在此真诚希望大家给予批评指正。

作者

2010年6月于南京大学

目 录

前言

第一章 绪论	1
第一节 统计学的渊源及其体系	1
第二节 统计学的概念	5
第三节 现代统计学的性质与特点	5
第二章 数据来源与数据质量	7
第一节 统计数据来源	7
第二节 统计数据的测量尺度	12
第三节 统计数据质量问题	15
第三章 定性数据的统计描述	18
第一节 定性数据的频数分布	18
第二节 定性数据的图形表示	22
第三节 定性数据的特征数字	24
第四章 定量数据的统计描述	31
第一节 定量数据的频数分布	31
第二节 定量数据的图形绘制	34
第三节 定量数据的特征数字	39
第五章 统计量的抽样分布	59
第一节 参数、统计量与抽样分布	59
第二节 正态分布及其应用	63
第三节 χ^2 分布、t分布与F分布	68
第四节 常用的抽样分布	76
第六章 样本参数推断	86
第一节 概述	86

第二节	参数估计	87
第三节	参数假设检验	98
第七章	非参数检验	107
第一节	卡方检验	107
第二节	符号检验	112
第三节	秩和检验	118
第四节	游程检验	121
第八章	方差分析与正交试验设计	124
第一节	方差分析原理	124
第二节	单因素方差分析	127
第三节	双因素方差分析	134
第四节	试验设计基本问题	142
第五节	正交试验分析	146
第九章	相关与回归分析	152
第一节	相关分析	152
第二节	简单回归分析	159
第三节	多元回归分析	172
第十章	时序分析与动态预测	179
第一节	时间序列的含义及作用	179
第二节	时间序列的分解与假定模型	182
第三节	长期趋势分析	186
第四节	季节变动分析	199
第五节	循环变动分析	208
参考文献		211
附录	常用统计表	212
A.	随机数表	212
B.	标准正态概率表	213
C.	t分布表	215

D. χ^2 分布表	218
E. F分布表	220
F. 常用正交表	226
G. 符号检验表	231
H. 秩和检验表	231
I. 游程数检验表	233
J. 游程长度检验表	233

第一章 绪 论

统计学是怎样产生和发展起来的？现代统计学的知识体系构成如何？统计思想与统计方法有哪些基本特征？本章将对这些问题进行介绍和说明。

第一节 统计学的渊源及其体系

统计活动源远流长，可以说自从人类社会有了数的概念、有了计数活动，统计就开始了。但作为一门科学的统计学，它的出现却晚得多。统计学的发展经历了以下几个阶段。

一、起源时期

根据统计史学的研究成果，统计学一般有三个来源。

（一）“政治算术”

17世纪中叶，在英国首次出现了有意识地用数字语言说明问题，注重从数量角度探索客观事物变化规律的研究活动。早期“政治算术”的杰出代表是配第(W.Petty)、格朗特(J. Graunt)等人。配第在《政治算术》(1676)这本书以及其他有关著作中，采用了不同于前人的研究方法，明确地用大量的数据资料分析问题，试图把结论建立在可靠的事实根据上。格朗特(J. Graunt)在《关于死亡表的自然观察与政治观察》(1662)中，也使用大量的人口数据，分析揭示了人口变化的规律。

（二）“国势学”

与“政治算术”产生的时期相近，在欧洲大陆的德国也出现了一门统计学——“国势学”。“国势学”主要研究“国家的有关显著事项”，即有关国家兴衰强弱的重大问题与治理国家必备的知识。康令(H.Conring)是“国势学”的奠基人，他对“国势学”的研究目的、研

究对象、研究方法等基本问题，作了具体深入的阐述。经过 100 多年的发展，到 18 世纪中期，“国势学”达到了顶峰。阿肯瓦尔(G.Achenwall)在继承康令开创的研究体系和研究方法的基础上，全面发展了“国势学”，尤其是给“国势学”起了一个统计学(statistik)的新名称。这一德文词转译成英文“statistic”后，逐渐被人们的接受并沿用至今。不同于“政治算术”的研究和表述风格，“国势学”最大的特点是它较少使用数字工具，而主要诉诸文字记述和逻辑比较。

(三) 概率论

统计学另一个重要起源是“概率论”，13、14 世纪在工商业比较繁荣的意大利和地中海沿岸其他地区，由于赌博游戏的盛行和保险活动的萌起，人们已经对“机会”问题发生了兴趣。不过真正意义上的“概率论”，是从 17 世纪开始的。帕斯卡(B. Pascal)和费马(P. Fermat)关于“得点问题”的讨论，奠定了概率论的基础。从现在的角度看，“政治算术”研究的是简单的、确定的数量关系，而概率统计则研究复杂的、随机性现象，这极大地充实和深化了数量问题研究的内容。以概率论为基础，统计学进入了一个崭新的发展阶段。

二、“凯特勒时代”

凯特勒(A. Quetelet)是统计学发展史上承前启后的重要人物，他一生撰写了大量的有关统计学方面的著作，代表性的就有《社会物理学》(1869)、《概率论书简》(1846)、《犯罪统计学》(1831)、《比利时统计研究》(1829)等。凯特勒在统计学上的突出贡献，是他把概率论全面引进“政治算术”、“国势学”以及其他社会问题的研究，推动了概率论和数学方法在社会科学领域的应用，促进了数量研究由“算术”水平向“数理”阶段的迅速转化。

三、“生物统计学”的兴起

从 19 世纪后半期起，统计学几乎在各个方面都取了创新性成果。其中，尤为引人注目的是兴起于生物遗传学、农业田间试验等领域的所谓统计数学方法的发展。戈尔登(F. Galton)是生物统计学的主创者，受凯特勒的影响，他也利用正态法则研究优生学、遗传学问题，认为正态

法则适应于许多情况，提出了“百分位数”、“中位数”、“四分位数差”、相关与回归等概念和计算方法。K.皮尔逊(K.Pearson)是戈尔登的学生，他系统发展了戈尔登的相关与回归理论，研究了复相关和偏相关，把物理学上“矩”的概念移植到统计学中，给出了极大似然估计方法，导出了重要的 χ^2 分布。以 K.皮尔逊为代表，以大量观察和正态分布为基础的关于总体分布曲线的研究，确立了“大样本”统计理论，奠定了“描述统计学”的基本体系。

四、统计学的“费暄时代”

进入 20 世纪，统计学经历了新一轮飞跃。一方面，新的统计思想和统计方法大量涌现；另一方面，带有归纳性质的统计推断逐渐占据了主流地位。此外，统计方法不仅作为科学研究和管理活动的工具，而且直接融合在科研、管理和生产过程之中，发挥着巨大的社会效益。自苏歇米尔斯(J.Süssmilch)首倡大数法则到 19 世纪末 20 世纪初的大约一个半世纪里，大量观察法一直是统计思想的核心，并被理所当然地接受下来。最先对这一理论发起挑战的是戈塞特(W.Gosset)，基于在酿酒公司多年的工作体会，戈塞特洞察到大样本统计方法并不适合一切情形，有时只能根据少量观察就必须作出结论，按他的话来说，“有些实验不能多次重复进行，这个时候必须依据极少量的实验结果作出判断，像有些化学实验，很多的生物实验和农业实验，便属于这样的情况，但它们也应该成为统计学的研究对象”。经过多年的潜心研究，1908 年，戈塞特终于导出了重要的 t 分布，这是“小样本”统计研究的基石。以此为标志，统计学逐渐实现了由以前的“描述统计”阶段向“推断统计”阶段、由大样本统计向小样本理论的转变。费舍尔(R.A.Fisher)是推断统计学的建立者，他对统计学进行了深入独到的研究，开辟了方差分析、试验设计等统计分支，给出了戈塞特 t 分布的简洁证明，论证了相关系数的抽样分布，提出了 t 检验、F 检验、相关系数检验，并编制了相应的检验概率表。费舍尔的代表著作包括《供研究人员使用的统计方法》(1925)、《试验设计》(1955)、《统计方法与统计推断》(1956)等。费舍尔在统计学发展史上有着辉煌崇高的地位，美国统计学家约翰逊

(P.Johnson)在《现代统计方法：描述和推断》(1959)一书中指出：“从1920年一直到今天的这段时期，称为统计学的费舍尔时代是恰当的，他的名字和他的著作，受到一切不带偏见的人的尊敬和传播。”

五、统计学的当代发展

尼曼(J. Neyman)和 E.皮尔逊(E. Pearson)是继费舍尔之后杰出的统计学家，共同完善了现代统计学的核心内容——区间估计和假设检验理论。20世纪50年代，瓦尔德(A. Wald)提出了“统计决策理论”和质量检验的“序贯分析”，进一步推动了统计学研究和应用的范围。随着计算机在统计中的应用，由威夏特(J. Wishart)、霍特林(H. Hotelling)等人发展起来的多变量统计，又重新活跃起来。20世纪50、60年代以来，稳健统计、时间序列、抽样理论、统计诊断、探索性分析、贝叶斯统计等，都取得了重要的进展。

通过对统计学发展历史的简单回溯可以看出，随着人们认识的不断深化、社会实践需要的推动，统计学始终在不断地丰富和完善着。它经历了从意义和概念不甚明确的阶段，到作为一门独立学科的转化；从数量研究的“算术”水平，到需要较高数学知识的“数理”阶段的转化；从确定型问题的研究，到随机现象研究的转化；从大量观察消除误差干扰以达到对客观现象规律认识的大样本理论，到控制试验次数提高数据质量的小样本推断的转化。目前，统计学已经形成了由若干个分支组成的庞大的学科体系，已经长成了一棵枝繁叶茂的参天大树(图 1.1)。

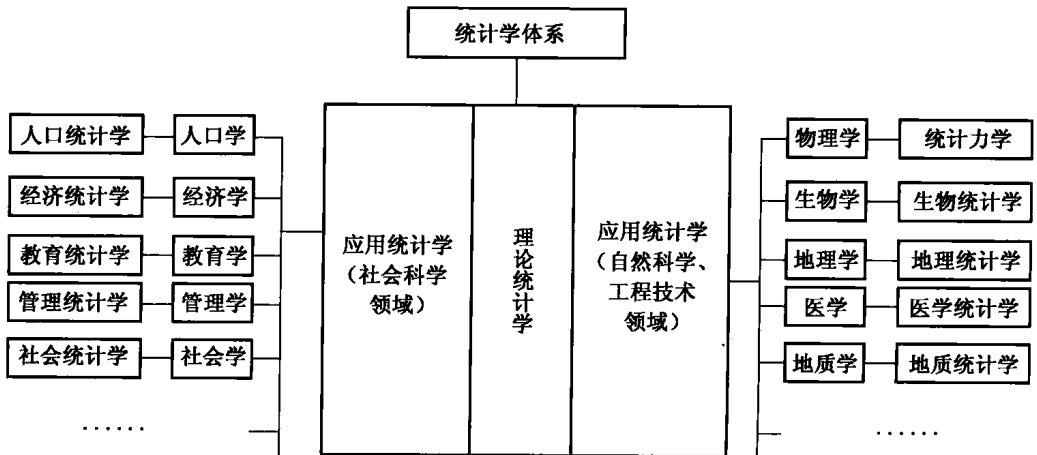


图 1.1 统计学的整体结构

第二节 统计学的概念

什么是统计学，不同的人可能会给出不同的回答。尽管统计学作为一门科学的地位已经得到了确立，但在一些重要问题的认识上仍然存在着分歧，这或多或少影响到人们对统计学范畴的表述。在此，我们侧重于统计应用，把统计学解释为关于资料搜集、整理、分析和解释的科学。

统计学研究客观现象数量方面，要达到这个认识目的，首先要搜集能够反映或说明客观现象的数字资料，这是统计活动首要的、基本的环节。搜集资料，就是计数和获取量数的过程。统计资料的质量如何，关系到能否得出客观的、正确的结论。因此要搞好统计调查活动，设法提高统计调查的水平。

统计整理在统计活动过程中，处于承上启下的地位。一方面它是统计搜集资料工作的继续；另一方面它又是统计分析的前提，能够为统计分析准备系统的资料形式。统计整理就是运用各种合适的形式展示和表述统计资料，其内容包括资料的审核、分组分类、编制频数分布、绘制资料分布图等。

统计分析是从已有的数据资料出发，通过有针对性的对比、计算提取有关综合信息的过程，其目的是使一个概括的、全面的数量描述能够形成。

在统计分析的基础上，导出一般性结论的过程就是统计解释，它往往涉及如何依据样本观察对同类大量现象作出估计、检验、预测等问题。

完整的统计活动过程由资料搜集、整理、分析和解释四个部分组成。它们虽然各有各的任务和作用，但又存在着密切的联系，只有同时做好各项工作，才能圆满地完成统计活动的任务。

第三节 现代统计学的性质与特点

历经 300 多年，特别是 19 世纪中后期以来的快速发展，人们对统计科学的看法逐渐形成了一定的共识。总结起来，现代统计学具有如下几个性质：

第一，数量性。统计学是研究数量问题的学问，统计的语言是数字，没有数字，不是从数量方面入手进行认识，就谈不上统计。无论是社会科学还是自然科学，只要出现大量数据的地方，都用得着统计也需要统计。根据辩证唯物主义的认识论，任何事物都是由数量和质量两方面组成的，是二者的统一体。统计研究客观事物数量，如果数字资料准确，统计方法运用得当，同样可以达到正确认识和反映客观事物的目的。统计研究客观事物的数量，主要包括数量状态、数量关系和数量变化规律。

第二，总体性。数量有个体数量与总体数量之别，统计学主要研究后者，它要对大量同类现象的数量方面进行综合反映。单个数字不能称为统计，也不可能指望从它身上发现什么有价值的东西，只有对大量的现象，或对某一现象进行多次重复的观察，才有可能找到统计关系和统计规律。统计虽然研究总体数量，但必须从个体数量的调查入手，遵循由个体数量到总体数量的认识逻辑。

第三，不确定性。统计学主要研究不确定性现象。所谓不确定性是指，由于受到偶然的、随机因素的作用，使得客观事物的实际数量表现存在一定程度的“不可确知性”。在现代统计学中，处理不确定性问题，是统计学的主要课题和任务。

第四，归纳推断性。统计方法带有归纳推断的特点，统计对总体的认识有两条途径，一是对构成总体的全部事物逐一进行调查，取得全面资料；二是从总体中抽取部分事物组成样本，然后依据样本观察结果对总体进行推断。至于前者，运用算术方法和统计描述手段就可达到目的，而后者相对比较复杂，需要运用概率论知识和统计数学方法。实际中，全面调查与非全面的抽样调查都会用到，但由于全面调查受到诸多因素的约束，从经济性、时效性、实用性和可行性方面考虑，利用样本资料进行推断的优势比较明显。统计方法的归纳推断性质，主要是相对推断统计而言的，同逻辑学意义上的归纳推断有着明显的区别，统计推断不是从假设、命题出发，按严格的逻辑推理程序进行推断，而是基于观察到的样本数据，对总体的可能情况作出判断。

第二章 数据来源与数据质量

实际工作中，统计资料是统计整理、统计分析和统计解释的原料，没有数据资料，统计整理、分析、解释便无法开展，所以搜集资料是统计活动的重要任务。本章将对统计数据的来源、统计测量尺度、统计数据质量等问题进行说明。

第一节 统计数据来源

一、直接来源

获得统计资料有多种途径，但概括起来不外乎是直接方式和间接方式，其中属于直接方式的有统计调查和统计实验。

(一) 统计调查

比较典型的统计调查主要有普查和抽样调查，下面来介绍它们的含义、特征与用途。

1. 普查

普查是专门组织的、一次性的全面调查，用来搜集某一时点或一定时期内现象的总量资料。普查是一种古老的调查手段，当初主要用于人口统计，目的是为了抽丁征税。由于其在掌握全面情况中的独特作用，后来慢慢地被用到许多问题的调查之中。现在，普查作为获取全面资料的有效途径之一，仍然保持着特殊的地位，并且在科学性、准确性、广泛性方面得到进一步的加强。通过普查所获得的资料，在了解国情国力、制定社会发展规划、确定重大决策方面，发挥着重要的作用。

现代普查活动一般具有下列几个特征：

第一，定期举行。为保证各个普查期资料的可比性，通常要求在同样的时间间隔内举行普查。

第二，依法进行。将普查活动用法律条文明确规定下来，以维护普查的严肃性和合法性，使其能够顺利地进行。

第三，应用广泛。过去的普查，主要是人口调查，现在它几乎在社会经济管理领域的方方面面都有应用，如经济普查、农业普查、工业普查、国土资源普查、教育普查、住房普查、失业普查等。

第四，质量控制。从道理上讲，普查对总体单位逐一进行登记，应该能够得到准确的资料，可是由于涉及面广、经办人员多、工作量浩大等原因，经常会出现差错。现代普查活动十分重视普查的组织工作和技术手段，同过去相比，普查资料的准确性已大有提高。

第五，项目增多。在一次普查活动中，列入调查的项目不再是区区的几项，而是多达十几项，甚至几十项。虽然不能指望将搜集资料的任务集中在一次普查中都解决掉，但是适当地增加一些调查项目，能扩大对调查对象基本情况的了解，从节约费用的角度看有时候可能还是划算的。

第六，计算机介入。运用计算机处理普查资料，给普查注入了新的活力，大大提高了普查效率和普查结果公布的时效性。

举行一次专门的普查活动，要耗费大量的人力、物力和财力，单独的研究机构无力承担。从准备调查方案、设计表格、试点、培训普查员，到实施调查以及后续的资料整理和分析，需要持续较长的时间。另外，大量的社会经济问题研究，并不总是需要通过普查来搜集资料，一味依赖普查，势必得不偿失。因此，应该结合调查对象的特点和统计任务的要求，综合运用各种搜集资料的方法。

2. 抽样调查

在统计调查方法体系中，抽样调查由于具备一般非全面调查的长处，同时又可以通过科学的推算达到对统计总体的认识，因而是一种最常用、最重要的获取统计资料的手段。抽样调查是指，按照随机性原则从调查对象(总体)中抽取一部分单位组成样本，然后根据样本调查的结果，对总体情况进行推断。

抽样调查的重要特征：

第一，按照随机性原则确定观察单位。在抽样调查中，究竟要对研究对象中的哪些单位进行调查，不取决于调查者与被调查者的主观

愿望，原则上要求每一个单位都应有同等的可能性被抽到。一方面随机性原则是抽样调查的基本原则，遵守这个原则可以避免统计估计的系统性偏差；另一方面只有符合抽样随机性原则，才能计算抽样估计误差。

第二，根据部分单位的调查结果，对总体进行科学推断。抽样调查既是搜集资料的方法，同时也是对统计总体进行认识的方法。就这一点来说，抽样调查和普查的作用是一样的，只不过采用的路径有所不同而已。

第三，抽样误差可以计算。用样本资料推断总体，必然会产生误差。但抽样理论表明，样本估计误差的大小可以计算出来，并且还能进行控制。

抽样调查中样本的抽取是关键问题之一，常用的抽样组织方式有以下五种。

(1) 简单随机抽样。简单随机抽样又叫纯随机抽样，是指从总体 N 个单位中抽取 n 个单位组成样本，每一个可能的样本被抽到的可能性或概率完全相等。由简单随机抽样方式所得到的样本，叫做简单随机样本或随机样本。要获得简单随机样本，可以通过抽签法和随机数表法进行。其中，抽签法在抽取样本单位时首先给总体每个单位确定一个唯一的编号，然后再准备同样数量的竹签(纸条或小球等)并分别标上号码，使得竹签(纸条或小球)上的号码对应着总体单位的一个编码，最后将竹签(纸条或小球)混合在一起拌匀，按随机原则逐次抽出或摸出竹签，直到抽出符合样本容量要求的竹签数为止。这时抽到的竹签上面的标号所对应的总体单位，即可作为被抽中的样本单位。随机数表法依据的是随机数表，它是通过随机过程生成的由 $0\sim 9$ 数字构成的一组数。随机数表的使用方法是：给研究对象的每个单位分别进行编号；确定要使用的随机数字的位数；查随机数表。查随机数表时，可以从随机数表中的任何一个地方出发，按照预先定下的查找方向，依次摘录下表中的数据，如果摘录出的数字为对应的总体单位的编号，就把这个单位确定为样本单位，对碰到的数字大于总体单位的最大编号则应舍去或跳过去，直到抽到符合要求的样本单位数为止。