



高等院校管理科学与工程规划教材

应用统计学： 经济与管理中的数据分析

赠送电子课件及
其他立体化资源

徐 哲

石晓军
王惠文

杨继平
柏满迎

编著

清华大学出版社

高等院校管理科学与工程规划教材

应用统计学：经济与管理中的数据分析

徐 哲 石晓军 杨继平 编 著
王惠文 柏满迎

清华大学出版社

北 京

内 容 简 介

统计学是一门研究如何搜集、整理、描述、显示和分析数据,帮助人们认识和探索系统的主要特征和内在数量规律性的学科。作为数据分析的一种重要工具,统计方法已经成为经济与管理学科领域的研究者和实际工作者的必备知识,是经济与管理类各专业学生必修的核心课程。本书从经济与管理学科对统计分析的要求出发,重点介绍应用统计学中的基础理论方法及其分析技术。与以往同类教材相比,本书增加和强化了经济管理统计分析方法及应用部分的章节,如抽样调查、非参数统计、多元统计分析,并且增加了案例分析和统计软件的应用。希望可以帮助读者更全面地掌握应用统计学的基本原理、方法和应用技巧,使得应用统计学方法真正成为一个实用的工具。

本书可作为高等院校经济管理专业本科生“应用统计学”课程的教材,也可作为 MBA 的教材和参考书,同时可供广大实际工作者参考。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

应用统计学:经济与管理中的数据分析/徐哲,石晓军,杨继平,王惠文,柏满迎等编著. —北京:清华大学出版社,2011.11

(高等院校管理科学与工程规划教材)

ISBN 978-7-302-26816-1

I. ①应… II. ①徐… ②石… ③杨… ④王… ⑤柏… III. ①统计分析—高等学校—教材 IV. ①C812

中国版本图书馆 CIP 数据核字(2011)第 186830 号

责任编辑:章忆文 杨作梅 郑期彤

装帧设计:杨玉兰

责任校对:李玉萍

责任印制:何 芊

出版发行:清华大学出版社

地 址:北京清华大学学研大厦 A 座

<http://www.tup.com.cn>

邮 编:100084

社 总 机:010-62770175

邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

印 刷 者:北京密云胶印厂

装 订 者:北京市密云县京文制本装订厂

经 销:全国新华书店

开 本:185×260 印 张:30.5 字 数:734 千字

版 次:2011 年 11 月第 1 版 印 次:2011 年 11 月第 1 次印刷

印 数:1~4000

定 价:49.00 元

产品编号:038703-01



前 言

本书主要面向经济管理学院的本科生，可作为其基础课程“应用统计学”的教材。对于经济管理学院的本科学生，也许第一个期待了解的问题就是：什么是管理？关于这个问题曾经有过很多细致而又全面的定义。从管理科学的发展历史来看，管理概念的发展主要经历了两个大的阶段。在泰勒、法约尔、梅育或马斯洛等学者的研究时期，关于管理科学的研究主要集中在如何提高工厂的生产效率上。而在第二次世界大战后，由于市场竞争日趋激烈，外部环境对企业发展的影响也越来越突出。因此，西蒙提出一个更加简洁和深刻的定义，即“管理就是决策”。与传统的管理科学相比，这是一个全新理念。它把企业的管理工作从如何正确地做事提升到如何做正确的事。现代的许多管理理论和方法研究都是围绕着如何进行科学有效的决策展开的。而所谓决策就是要在多个可能的方案中进行择优。最典型的譬如运筹学，就是研究如何在资源有限的情况下选择最佳的决策方案的。

在本书中我们要重点思考的问题是：在决策过程中，统计学又扮演什么样的角色。事实上，在做任何决策判断之前，一个最重要的工作就是要对系统及其环境的未来发展有十分准确的判断。古人云：预知三日事，富贵一千年。也就是说，预测在决策过程中往往起着决定性作用。然而，怎样才能够做好预测工作呢？关键是要对系统及其环境的状态、特性以及发展趋势有非常深刻及全面的认识。而这些认识则源于决策者或辅助分析人员能够很好地收集、处理和运用相关的资料与数据。其实，这正是应用统计学要完成的科学任务。换句话说，作为决策分析的前期基础，统计学中关于数据收集、描述、分析和解释的技术在管理科学工作中是不可或缺的。而且读者在亲身体验实际的数据分析工作后一定会感悟到：数据是会说话的！

本书作者长期从事统计教学和以统计为主要方法的科学研究工作，具有丰富的教学经验和科研经历。本书从经济与管理学科对统计分析的要求出发，重点介绍应用统计学中的基础理论方法及其分析技术。与以往同类教材比较，本书增加和强化了经济管理统计分析方法及应用部分的章节，如抽样调查、非参数统计、多元统计分析，并且增加了案例分析和统计软件的应用。在教材写作风格上力求做到对于统计理论和方法的讨论简洁明了、通俗易懂，注意直观描述和逻辑推理，尽量避免复杂烦琐的数学推导与证明，而以方法论为主线，同时注意理论、方法和实用并行。

本书是由北京航空航天大学经济管理学院承担“应用统计学”精品课程建设项目的五位教师协作完成的，他们分别是：徐哲(编写第一章、第二章、第三章、第五章、第七章、附录)，石晓军(编写第六章、第八章)，杨继平(编写第十二章、第十三章、第十四章)，王惠文(编写第十章、第十一章、前言)，柏满迎(编写第四章、第九章)。本书由徐哲、石晓军负责全书的统稿。

本书的习题答案和模板文件可以在网站 www.wenyuan.com.cn 上下载。

本书在相关的研究工作中，受到北京航空航天大学精品课程建设项目的支持，在本书



的写作中，曾多次得到北京航空航天大学精品课程建设评估小组专家们的指导和帮助，并承蒙评审委员会对本书的写作计划和部分初稿提出十分重要的建议。作者愿借此机会，向他们表示衷心的感谢。

作者还要特别感谢清华大学出版社的编辑们为本书出版所付出的心血。没有他们的帮助和细致严谨的工作，本书是不会以这样的形式面世的。

由于作者的水平有限，书中难免存在缺点和错误，敬请读者批评指正。

编 者



作者简介

徐 哲，北京航空航天大学经济管理学院企业管理系教授、博士生导师。现为中国质量协会学术教育工作委员会委员、中国商业统计学会理事。

石晓军，北京航空航天大学经济管理学院保险与风险管理系教授、博士生导师。

杨继平，北京航空航天大学经济管理学院金融系副教授，硕士生导师。现为全国统计方法应用标准化技术委员会委员、西方金融学会会员。

王惠文，北京航空航天大学经济管理学院管理科学与工程系教授、博士生导师，经济管理学院院长，北航复杂数据分析研究中心主任，并兼任北京市政协第十一届委员会委员以及市政协提案委员会副主任。现为国际统计学学会(ISI)会员、国际统计计算学会(IASC)会员、全国统计教育委员会常务理事、中国市场学会常务理事、中国商业统计学会常务理事、全国商业自动化标准技术委员会委员、中科院虚拟经济与数据科学研究中心客座研究员、《北京航空航天大学学报(社科版)》副主任、《数理统计与管理》编委、《中国非公经济统计年鉴》编委、*Journal of Symbolic Data Analysis* 编委。

柏满迎，北京航空航天大学经济管理学院保险与风险管理系教授、系主任、硕士生导师。现为中国企业运筹学会理事，中国保险学会理事。

目 录

第一章 绪论..... 1

第一节 统计与统计学..... 1

一、统计的含义..... 1

二、学习统计学的意义..... 2

三、统计学研究对象的特点..... 2

四、客观现象内在数量规律性..... 3

第二节 统计数据的计量与类型..... 4

一、统计数据的计量尺度..... 4

二、统计数据的类型..... 6

第三节 统计分析的基本概念..... 7

一、总体与样本..... 7

二、总体参数与样本统计量..... 8

三、变量..... 8

第四节 统计学分科..... 9

一、描述统计学与推断统计学..... 9

二、理论统计学与应用统计学..... 10

第五节 统计学的产生和发展..... 11

一、古典统计学时期..... 11

二、近代统计学时期..... 12

三、现代统计学时期..... 12

第六节 统计学在经济管理中的

应用举例..... 13

本章小结..... 14

思考题..... 14

第二章 统计数据的收集、整理 与展示..... 16

第一节 统计资料的收集..... 16

一、统计数据的来源..... 16

二、统计资料的收集方法..... 17

三、统计资料收集的技术工具..... 18

四、统计调查的组织方式..... 21

五、统计资料收集方案设计..... 23

第二节 统计数据的整理..... 24

一、统计数据的审核..... 24

二、统计分组..... 25

三、频数分布..... 27

第三节 统计表与统计图..... 32

一、统计表..... 32

二、统计图..... 33

第四节 品质数据的图形展示..... 34

一、条形图..... 34

二、饼图..... 35

三、累积频数(频率)分布图..... 35

第五节 数量数据的图形展示..... 36

一、直方图..... 36

二、折线图..... 37

三、累积折线图..... 38

四、茎叶图..... 39

五、频数(频率)分布曲线..... 40

第六节 统计软件的应用..... 42

一、品质数据的图形展示..... 42

二、数量数据的图形展示..... 46

本章小结..... 56

思考与练习题..... 56

第三章 统计数据分布特征的描述..... 59

第一节 数据分布的集中趋势..... 59

一、集中趋势与平均指标..... 59

二、数值平均数..... 60

三、位置平均数..... 63

四、众数、中位数和平均数的比较..... 67

五、相对位置：百分位数..... 68

第二节 数据分布的离散程度..... 70

一、测定离散程度的意义..... 70

二、品质数据离散程度的度量..... 70

三、数量数据离散程度的度量..... 70



第三节 集中趋势与离散程度的综合运用	74	一、总体均值的区间估计	109
一、 z 分数	74	二、总体比例的区间估计	113
二、经验规则与切比雪夫定理	75	三、正态总体方差的区间估计	114
三、异常值检测	77	第四节 两个总体参数的区间估计	115
第四节 数据分布形状的测度	77	一、两个总体均值之差的区间估计： 独立样本	115
一、测定分布形状的意义	77	二、两个总体均值之差的区间估计： 配对样本	118
二、数据分布的偏态	77	三、两个总体比例之差的区间估计： 独立样本	119
三、数据分布的峰度	78	四、两个总体方差之比的区间估计： 独立样本	120
第五节 统计软件应用	80	第五节 样本容量的确定	122
本章小结	81	一、估计总体均值时样本容量的 确定	122
思考与练习题	82	二、估计总体比例时样本容量的 确定	123
案例分析	84	三、估计两个总体均值之差时样本 容量的确定	124
第四章 正态总体的抽样分布	86	四、估计两个总体比例之差时样本 容量的确定	124
第一节 抽样分布的概念	86	第六节 统计软件应用	125
一、样本均值的抽样分布	86	一、一个正态总体均值的 区间估计(大样本)	125
二、样本比例的抽样分布	88	二、一个正态总体均值的 区间估计(小样本)	125
第二节 正态分布及其三大抽样统计量 分布	88	三、一个总体比例的区间估计	126
一、正态分布	88	四、一个正态总体方差的区间 估计	126
二、 χ^2 分布	91	五、两个正态总体均值之差的区间 估计(独立样本、大样本)	127
三、 t 分布	92	六、两个正态总体均值之差的区间估计 (独立样本、小样本、等方差)	127
四、 F 分布	94	七、两个正态总体均值之差的区间估计 (独立样本、小样本、异方差)	128
本章小结	96	八、两个正态总体均值之差的区间估计 (配对样本、大样本)	129
思考与练习题	96	九、两个正态总体均值之差的区间估计 (配对样本、小样本)	129
案例分析	99		
第五章 参数估计	101		
第一节 参数估计概述	101		
一、参数估计的基本概念	101		
二、点估计的基本概念	101		
三、区间估计的基本概念	102		
四、置信区间的性质	102		
第二节 点估计	103		
一、矩估计法	103		
二、极大似然估计法	105		
三、点估计的优良性准则	106		
第三节 一个总体参数的区间估计	109		

十、两个总体比例之差的区间估计	130	一、假设的陈述	170
十一、两个正态总体方差之比的区间估计	130	二、假设检验中的两类错误	171
本章小结	131	三、检验统计量与拒绝域	173
思考与练习题	132	四、假设检验的基本思想	175
案例分析	134	五、假设检验的基本步骤	176
第六章 抽样调查基础	135	第二节 一个总体参数的假设检验	177
第一节 基本概念	135	一、总体均值的检验	177
一、抽样单元和抽样框	135	二、总体比例的检验	179
二、误差	136	三、总体方差的检验	180
三、精度、信度与效度	136	第三节 两个总体参数的假设检验	181
第二节 简单随机抽样	137	一、两个总体均值之差的检验： 独立样本	181
一、样本容量的确定	137	二、两个总体均值之差的检验： 配对样本	184
二、估计方法	138	三、两个总体比例之差的检验： 独立样本	186
三、方差与区间估计	141	四、两个总体方差之比的检验： 独立样本	188
第三节 分层随机抽样	142	第四节 假设检验中的其他问题	189
一、层数确定与样本量分配	143	一、 P 值(P -value)	189
二、估计量及其性质	145	二、检验功效曲线	190
第四节 多阶段抽样	147	第五节 统计软件应用	194
一、多阶段抽样概述	147	一、一个正态总体均值的 检验(大样本)	194
二、估计量及其性质	148	二、一个正态总体均值的 检验(小样本)	195
第五节 整群随机抽样	151	三、一个总体比例的检验	195
一、整群随机抽样概述	151	四、一个正态总体方差的检验	196
二、群划分的原则	152	五、两个正态总体均值比较的 检验(独立样本, 大样本)	197
三、群规模大小相等时的估计量 及其性质	152	六、两个正态总体均值比较的检验 (独立样本, 小样本, 等方差) ..	197
第六节 系统抽样	154	七、两个正态总体均值比较的检验 (独立样本, 小样本, 异方差) ..	198
一、系统抽样的实施方法	155	八、两个正态总体均值比较的检验 (配对样本, 大样本)	199
二、估计量及其性质	156	九、两个正态总体均值比较的检验 (配对样本, 小样本)	199
第七节 统计软件应用与案例研究	159	十、两个总体比例比较的检验	200
一、统计软件应用	159		
二、案例研究	161		
本章小结	165		
思考与练习题	166		
案例分析	169		
第七章 假设检验	170		
第一节 假设检验基本原理	170		



十一、两个正态总体方差比较的 检验	201	案例分析	233
本章小结	201	第九章 非参数统计初步	236
思考与练习题	202	第一节 卡方检验	236
案例分析	205	一、多项总体拟合优度检验	237
第八章 试验设计与方差分析初步	208	二、泊松分布拟合优度检验	239
第一节 试验设计与方差分析概述	208	三、正态分布拟合优度检验	239
一、基本术语	208	四、列联表独立性检验	241
二、方差分析的基本思想	209	第二节 单个总体的位置检验： 符号检验	244
第二节 单因素试验设计与方差分析	210	第三节 两个总体的比较	246
一、单因素设计的方差分析	211	一、两个总体的比较：独立随机 样本	246
二、多重比较	212	二、两个总体的比较：配对样本	248
三、影响最大的处理	212	第四节 多个总体的比较	251
第三节 配对比较试验设计与方差 分析	213	一、多个总体的比较：独立样本	251
一、配对比较试验设计	213	二、多个总体的比较：配对样本	252
二、配对比较的方差分析	214	第五节 其他非参数统计方法	253
第四节 随机化区组设计与方差分析	214	一、Spearman 秩相关检验	253
一、随机化区组试验设计	214	二、Kendall 一致性检验	255
二、随机化区组设计的方差分析	215	三、随机性游程检验	257
三、随机化区组设计的多重比较	216	本章小结	258
第五节 二因素无交互作用试验设计 与方差分析	217	思考与练习题	258
一、二因素无交互试验设计	217	案例分析	265
二、二因素无交互试验方差分析	217	第十章 线性回归分析	267
第六节 二因素且有交互作用的试验 设计与方差分析	218	第一节 相关系数的概念	267
一、二因素且有交互试验设计	218	一、函数关系与随机关系	267
二、二因素有交互试验的方差 分析	220	二、Pearson 相关系数	269
第七节 统计软件应用	221	三、相关关系与因果关系之间的 联系	270
一、单因素方差分析	222	第二节 线性回归模型	270
二、无交互作用的双因素方差 分析	223	一、回归分析所研究的问题	270
三、有交互作用的双因素方差 分析	225	二、线性回归的总体模型	272
本章小结	227	第三节 最小二乘估计方法	274
思考与练习题	227	一、最小二乘估计方法的推导	274
		二、总体参数估计量的性质	276
		第四节 模型效果分析	276
		一、为什么要进行模型效果分析	276

二、残差的样本标准差.....	277
三、测定系数.....	278
第五节 显著性检验.....	281
一、回归模型的线性关系检验:	
F 检验.....	281
二、回归参数的检验: t 检验.....	282
第六节 变量筛选方法.....	283
一、向后删除变量法.....	283
二、向前选择变量法.....	284
三、逐步回归法.....	284
第七节 残差分析.....	285
一、残差分析的基本原理.....	285
二、残差分析的几个主要内容.....	285
第八节 案例分析与统计软件应用.....	286
一、一元线性回归建模与 Excel 软件	
应用.....	286
二、多元线性回归建模与 SPSS 软件	
应用.....	289
第九节 虚拟自变量的回归.....	292
第十节 拟线性回归模型.....	293
一、非线性回归问题.....	293
二、拟线性回归模型的一般形式.....	294
第十一节 异常值的问题.....	295
一、离群点.....	295
二、高杠杆率点.....	296
第十二节 多重相关性问题.....	297
一、多重相关性的含义.....	297
二、多重相关性的危害.....	299
三、多重相关性的经验诊断与处理	
方法.....	302
四、多重相关性的补救方法简介.....	302
本章小结.....	304
思考与练习题.....	305
案例分析.....	309
第十一章 主成分分析.....	311
第一节 工作目标和基本思路.....	311
一、主成分分析的工作目的.....	311
二、主成分分析的基本原理.....	312

第二节 主成分分析的计算方法.....	313
一、算法推导.....	313
二、主成分分析的计算步骤.....	315
三、主成分的基本性质.....	315
第三节 辅助分析技术.....	316
一、精度分析.....	316
二、解释主成分.....	317
三、特异点的发现.....	318
四、样本点在主超平面上的表现	
质量.....	319
五、数据重构.....	320
第四节 因子分析方法.....	321
一、因子分析模型.....	321
二、因子分析的基本原理.....	322
第五节 统计软件应用与案例研究.....	322
一、SPSS 软件应用.....	322
二、案例研究.....	327
本章小结.....	333
思考与练习题.....	334
案例分析.....	336
第十二章 聚类分析和判别分析.....	338
第一节 聚类分析.....	338
一、聚类分析概述.....	338
二、距离和相似系数.....	339
三、系统聚类法.....	342
四、动态聚类法.....	350
第二节 判别分析.....	353
一、判别分析概述.....	353
二、距离判别法.....	353
三、Bayes 判别法.....	357
四、Fisher 判别法.....	360
第三节 统计软件应用.....	362
一、聚类分析.....	362
二、判别分析.....	369
本章小结.....	373
思考与练习题.....	373
案例分析.....	374



第十三章 时间序列分析和预测375

第一节 时间序列的概念和种类375

一、时间序列的概念375

二、时间序列的种类376

第二节 时间序列的水平与速度分析377

一、时间序列的水平分析378

二、时间序列的速度分析381

第三节 时间序列的构成和预测模型383

一、时间序列的构成383

二、时间序列预测模型385

第四节 长期趋势的测定386

一、移动平均法386

二、指数平滑法390

三、模型拟合法393

第五节 季节变动的测定396

一、季节指数396

二、测定季节变动396

三、季节调整399

第六节 循环波动的测定400

第七节 统计软件应用403

一、时间序列数据格式的产生403

二、时间序列序列图的绘制403

三、时间序列的分解404

本章小结411

思考与练习题412

案例分析414

第十四章 指数416

第一节 指数概述416

一、指数的概念416

二、指数的作用417

三、指数的分类417

第二节 总指数的编制方法419

一、综合指数419

二、平均指数423

第三节 指数体系和因素分析427

一、指数体系428

二、因素分析429

三、总量指标因素分析429

第四节 几种常用的价格指数431

一、商品零售价格指数431

二、居民消费价格指数435

三、工业品出厂价格指数442

四、股票价格指数443

本章小结445

思考与练习题446

案例分析448

附录 常用统计表450

中英文专业术语对照464

参考文献469





第一章 绪 论

【本章导读及学习目标】

统计学是一门收集、分析、表现和解释数据的学科(大英百科全书, Encyclopaedia Britannica, 2009)。在日常的工作和生活中,人们经常会接触各种各样的统计数据,如报刊、电视和广播中经常报道的国内生产总值、物价指数和证券指数等数据;球类比赛得分、进攻次数和成功次数等数据;企业的财务和销售部门定期统计的生产和销售情况数据;学生的考试分数和平均分等数据;等等,这些都是统计数据。对这些数据进行整理和分析的工作,以作出种种结论和预测,所依赖的主要方法之一就是统计方法。统计学在自然科学和社会科学领域发挥着日益重要的作用。通过本章的学习,要求了解统计学的一些基本问题,包括统计的含义、统计数据计量与类型、统计分析的基本概念、统计学的基本分科等。

第一节 统计与统计学

一、统计的含义

“统计”的英文是“Statistics”,这个词在英文字典中有两个含义:①当它以单数名词出现时,表示作为一门学科的“统计学”;②当它以复数名词出现时,表示“统计数据”。

人们常说的“统计”,一般可以从三个角度来理解:一是统计工作,二是统计数据,三是统计学。

(一)统计工作

统计工作是指人们运用科学方法对调查研究的对象进行数据收集、整理和分析的全部工作,其结果形成一系列的统计数据资料。统计工作一般包括统计设计、统计资料的收集、统计整理和统计分析这四个不同的环节。

(二)统计数据

统计数据是指通过统计工作取得的各种数据,是统计工作的结果。统计数据必须具有客观性、准确性、及时性、系统性和连续性等特点,它是进行科学管理和决策的基本依据。

(三)统计学

统计学是一门收集、分析、表现和解释数据的学科(大英百科全书, Encyclopaedia



Britannica, 2009)^①。统计学是对统计实践加以总结升华而产生的理论，是关于统计方法如何建立，及其正确性和有效性的数学论证。

统计工作、统计数据和统计学三者之间有着密切的联系：统计数据是统计工作的成果，是对客观现象进行统计研究的基础；统计学是统计工作经验的理论概括和总结，又是指导统计工作的原理和方法。

二、学习统计学的意义

从现代管理的观点来看，决策是一项至关重要的工作。美国管理学家、诺贝尔奖获得者西蒙(H.A.Simon)曾对“管理”一词给过最著名和简洁的定义，即“管理就是决策”。在他看来，企业是一个与社会相关联的开放系统，始终处于复杂的、时刻变化的内外部环境之中。企业经营的成败，不仅取决于其内部的作业效率，更取决于它在投资、计划和销售等方面的决策。一个关键性的创意、决策，可以使企业迅速走向发展的高潮；而一项重大决策的失误，也可以导致企业经营的失败。而决策成败的关键就在于决策者能否对系统的状况、规律有充分的认识和准确的判断。

广义地说，系统分析和决策的进程可以划分为四个阶段：首先，是对系统的描述性分析，它是运用所掌握的信息，对系统进行尽可能充分和全面的认识；其次，是对系统的解析性分析，它往往通过数学模型，识辨和刻画系统的解析结构，确定系统中各要素之间的内在联系；第三，是有关系统的预测性研究，其目的是掌握系统运行和动态变化规律，对系统的未来做出准确的预见。准确的预见是正确决策的先决条件；第四，是决策阶段，所谓决策，就是指对系统状况进行充分的观察和认识，对系统构造及其要素的内在联系进行识辨和深入的分析，并且，在对未来做出正确判断的基础上，在多种可行的方案中进行优选。显而易见，这四个阶段是递进的，而作为决策基础的系统认识、分析和预测性研究，均属于数据分析的范畴。基于以上分析，可以认为：统计数据分析是决策的基础和前导工作，是经济管理工作中一个十分重要的实用分析工具。

三、统计学研究对象的特点

统计学的研究对象是统计研究要认识的客体，它决定着统计学的研究领域以及相应的研究方法。一般地说，统计学的研究对象是客观事物的数量特征和数量关系，其研究目的是认识现象的发展规律。人们要认识客观事物，认识现象的数量关系和变动规律，就必须从观测个别现象入手，去收集、加工、整理这些观测值，对大量的个别现象的数量表现进行归纳，概括出它们的共同特征，即现象总体的数量特征，从而从数量上对现象总体的变动规律进行解释。统计学的研究对象具有以下明显的特点。

^① Statistics is the science of collecting, analyzing, presenting, and interpreting data. 参见 Encyclopaedia Britannica 的“statistics”条目。Encyclopaedia Britannica 2009 Student and Home Edition. Chicago: Encyclopaedia Britannica.

(一)数量性

数量性是统计学研究对象的基本特点。对客观现象的认识,无论是社会现象还是自然现象,既有“质”的方面又有“量”的方面。而统计则侧重于对客观现象数量的研究。统计学研究对象的数量包括:数量多少;客观现象数量之间的关系;质量互变的数量界限;客观现象量变的趋势及其规律等。

应当注意的是,统计学作为一种工具和方法,侧重于现象的“量”,它并不解决科学研究的许多“为什么”的问题。但并不意味着它和现象的“质”的方面没有关系。任何事物都是“质”与“量”的统一体,统计学上的量往往都是具体的、有一定的质的规定性的量,统计学如果脱离现象的质的规定性,不重视或不了解现象的质的方面,就不可能正确反映客观现象和揭示其规律性。

(二)总体性

统计学研究的客观现象的数量方面,不是个别现象的数量方面,而是现象总体的数量方面。所以,统计学是研究总体的数量特征及其分布规律性。总体是由具有某种共同性质的许多个体组成的,但个体在数量上存在着差异,需要通过这些个体的差异来描述或推断总体的数量特征和分布规律,这时就产生了统计的推断问题。例如,人口统计不是要了解和研究个别的人,而是要反映一个国家或地区的人口多少、男女比例、年龄结构、出生率、死亡率及其变化的规律性。个别现象的表现是复杂多样的,甚至是杂乱无章的,但现象总体却会表现出某种共同的趋势,从而有规律可循。

(三)变异性

统计学研究同类现象总体的数量特征,它的前提是总体各单位特征表现存在着差异,而且这些差异不是由某些特定的原因事先确定的。统计上把总体各单位由于随机因素引起的某一标志表现的差异称为变异。例如,某企业的职工,其年龄、文化水平、技术能力、工资报酬等存在着差异。这就需要我们研究该企业职工的平均年龄、文化结构、技术等级分布、平均工资水平等统计指标。试想,如果总体各单位不存在差异,就不需要收集大量的统计数据,也就不需要进行统计了。

四、客观现象内在数量规律性

从统计学研究对象的特点可知,统计研究是从总体各单位的变异中归纳概括出总体的共同特征,各总体单位的变异表现出个别现象的特殊性和偶然性,而它们的共同特征则表现出现象的普遍性、规律性和必然性。

(一)客观现象内在数量规律性举例

1. 男女性别比例的数量规律性

从表面看,新生儿的性别比例似乎没有什么规律可循。但如果对新生儿的性别进行大量观察,就会发现男孩与女孩的比例为 107:100,即大致为每生育 100 个女孩,就会生育 107 个男孩,这就是新生儿性别的规律性。虽然,新生儿时男多于女不平衡,但男孩的



死亡率高于女孩，到了中青年时，男女比例就大致相同了。进入中老年后，男性的死亡率仍然高于女性，导致男性的平均预期寿命比女性短，老年男性反而少于女性。所以，从一个国家乃至全人类来看，婴幼儿时男多于女，中青年时大致平衡，老年时女多于男，总人口上性别是大致平衡的，保证了人类社会的进化和发展。对人类性别比例的规律性研究是统计学方法探索的最早的数量规律性之一。

2. 赌博游戏

掷硬币和掷骰子是大家最为熟悉的游戏，随机地掷一次硬币或骰子不能事先确定出现正面、反面或某个点数，也就是说个别的游戏或试验充满了不确定性或偶然性。赌徒们正是利用了这种偶然性进行赌博。但当我们进行大量观察，即不断做重复试验时，就会发现掷一枚均匀的硬币出现正面或反面的次数接近相等，即比值为 $1/2$ ，试验次数越多，就越接近 $1/2$ 这一稳定的数值。同样，在掷骰子(六面体)时，出现 $1\sim 6$ 每个点的比率也逐渐接近 $1/6$ 。这里的 $1/2$ 和 $1/6$ 就是掷硬币和掷骰子时出现某一特定结果的概率，也就是我们探索的数量规律性。

3. 医学研究

为探索吸烟与患肺癌之间是否有关联，采取随机调查的方法，调查了一大批人，按是否吸烟和是否患肺癌分成四类：不吸烟也不患肺癌、吸烟患肺癌、不吸烟患肺癌和吸烟不患肺癌。在抽选的这批被调查者中，统计不吸烟患肺癌和吸烟患肺癌的比率，比方说分别为十万分之三和十万分之十二，用一定的统计方法可以推断：吸烟患肺癌的危险性是不吸烟患肺癌的四倍。这个结论是否可信，取决于我们如何设计这个调查方案、如何进行调查的以及采用的统计整理和推断方法是否可靠。

(二)为什么统计学能够探索客观现象内在数量规律性

客观现象内在数量规律性是由客观事物本身的特点和统计方法的特性共同决定的。

从客观事物方面来看，根据辩证法的基本原理，任何客观事物都是必然性与偶然性的对立统一。必然性反映了事物本质的特征和联系，是比较稳定的，它决定了事物的内在本质是有规律可循的。偶然性反映了该事物每个表现形式的差异，使事物的表现形式与必然性和规律性发生偏移，形成了数据的千差万别。这样，必然性的数量规律性就被掩盖在表面的差异之中。应用统计方法就可以从表面的偶然性中探索到内在的、本质的数量规律。

从统计方法来看，统计学提供了一系列的方法：实验设计法、大量观察法、综合分析法、归纳推断法，这些方法专门用来收集数据、整理数据、显示数据的特征进而分析和探索(或推断)出事物总体的数量规律性。

第二节 统计数据的计量与类型

一、统计数据的计量尺度

统计数据是对客观现象进行计量的结果，在收集数据之前，我们总是要对现象进行计量或测度。但不同事物的特性决定了其可计量或测度的程度是不同的，比如有些事物只能

对它的属性进行分类,有些则可以用比较精确的数据加以计量。按照我们对客观事物测度的程度或精确水平来看,可将所采用的计量尺度由低级到高级、由粗略到精确分为四个层次,即定类尺度、定序尺度、定距尺度和定比尺度,采用不同的计量尺度可以得到不同类型的统计数据。表 1-1 为某社区 5 位居民的基本情况的统计数据。

表 1-1 某社区 5 位居民的基本情况统计

样本号	身高/cm	体重/kg	年龄	性别	工作性质	颜色喜好	体质
No.1	166	56	31	女	工人	红	好
No.2	168	60	45	女	干部	蓝	中
No.3	173	67	38	男	教师	绿	差
No.4	175	62	42	男	军人	黄	好
No.5	169	59	50	男	技术人员	蓝	中

(一)定类尺度

定类尺度(nominal scale)也称为列名尺度,是最粗略、计量层次最低的计量尺度。它是对客观事物依据某种属性进行平行的分类或分组,使用该尺度对事物所作的分类之间既没有数量关系,也没有等级关系。即该计量尺度的不同表现之间没有高低、大小和优劣之分。如表 1-1 中性别、工作性质、颜色喜好等的分类就是一种定类尺度。一般来说各类之间是互相排斥的,也就是说只能属于其中的一类。

(二)定序尺度

定序尺度(ordinal scale)也称顺序尺度,它是对客观现象类别之间的等级差或顺序差的一种测度。定序尺度不仅可以研究对象分成不同的类别,而且还反映各类的优劣、量的大小或顺序。如表 1-1 中统计体质状态时可以分为好、中、差三个等级。又如学生考试的成绩划分为优、良、中、及格、不及格就是一种顺序的计量,即优比良高一级,良比中高一级,以此类推。又如产品等级就是对产品质量好坏的一种顺序测度,它可以把产品分为一等品、二等品、三等品、次品等。很显然,定序尺度较定类尺度要精细一些,但是它只能测度类别之间的顺序,但是不能测度类别之间的准确差异大小。因此不能进行代数运算和其他数学运算,只可以比较大小,也就是说,各类之间只有“大于”或“优于”的概念。

(三)定距尺度

定距尺度(interval scale)也称间隔尺度,它不仅能将事物区分为不同类型并确定顺序,而且可以准确地计量它们之间的差距是多少,定距尺度数据值进行加、减运算是有意义的。例如,“智商”是定距尺度,假设,甲某智商为 110,乙某智商为 90,从这两个智商值可知:① 甲乙两人智商不同(定类尺度的性质);② 甲比乙智商高(定序尺度的性质);③ 甲智商比乙智商高 20(定距尺度的性质)。但定距尺度的数据值不能进行乘除运算,因为定距尺度的数据值表现为“0”时,“0”在这里并不是“无”的含义,智商为 0 的人并不是说他没有智力。当然,我们也不能说智商为 120 的人比智商为 80 的人的智力高 1.5 倍。