



统计学理论与实务

主 编 王雪秋 董小刚

21 世纪全国高等院校财经管理系列实用规划教材

统计学理论与实务

主 编	王雪秋	董小刚	
副主编	王明明	赵天宇	
参 编	张纯荣	赵 晖	杜春晶



北京大学出版社
PEKING UNIVERSITY PRESS

内 容 简 介

本书以统计学基本步骤为主线,从最为基础的收集和整理数据开始,渐次介绍数据分析的两种方法:描述性统计分析和推断性统计分析。对变量的分析也由研究单一变量间的关系逐步过渡到对二维变量及多元变量的分析,其间适时配以案例。考虑到学生的向量代数的基础,回归分析中主要是以单个方程为主。

本书共分8章,具体内容包括:总论、统计数据的收集与处理、统计数据的整理与图形展示、统计数据的指标度量、参数估计、假设检验、方差分析、相关与一元回归分析。

本书既可作为高等院校财经管理类专业的本、专科生教材,也可作为相关工作人员自学参考用书。

图书在版编目(CIP)数据

统计学理论与实务/王雪秋,董小刚主编. —北京:北京大学出版社,20158
(21世纪全国高等院校财经管理系列实用规划教材)

ISBN 978-7-301-24467-8

I. ①统… II. ①王…②董… III. ①统计学—高等学校—教材 IV. ①C8

中国版本图书馆 CIP 数据核字(2014)第 147742 号

书 名	统计学理论与实务
著作责任者	王雪秋 董小刚 主编
责任编辑	王显超
标准书号	ISBN 978-7-301-24467-8
出版发行	北京大学出版社
地 址	北京市海淀区成府路 205 号 100871
网 址	http://www.pup.cn 新浪微博: @北京大学出版社
电子信箱	pup_6@163.com
电 话	邮购部 62752015 发行部 62750672 编辑部 62750667
印 刷 者	三河市北燕印装有限公司
经 销 者	新华书店
	787 毫米×1092 毫米 16 开本 13.75 印张 312 千字
	2015 年 8 月第 1 版 2015 年 8 月第 1 次印刷
定 价	30.00 元

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,侵权必究

举报电话:010-62752024 电子信箱: fd@pup.pku.edu.cn

图书如有印装质量问题,请与出版部联系,电话:010-62756370

21 世纪全国高等院校财经管理系列实用规划教材

专家编审委员会

主任委员 刘诗白

副主任委员 (按拼音排序)

韩传模

李全喜

王宗萍

颜爱民

曾 旗

朱廷珺

朱淑珍

顾问 (按拼音排序)

高俊山

郭复初

胡运权

万后芬

张 强

委员 (按拼音排序)

程春梅

邓德胜

范 徵

冯根尧

冯雷鸣

黄解宇

李柏生

李定珍

李相合

李小红

刘志超

沈爱华

王富华

吴宝华

张淑敏

赵邦宏

赵 宏

赵秀玲

法律顾问 杨士富

丛 书 序

我国越来越多的高等院校设置了经济管理类学科专业,这是一个包括理论经济学、应用经济学、管理科学与工程、工商管理、公共管理、农林经济管理、图书馆、情报与档案管理 7 个一级学科门类和 31 个专业的庞大学科体系。2006 年教育部的数据表明,在全国普通高校中,经济类专业布点 1518 个,管理类专业布点 4328 个。其中除少量院校设置的经济管理专业偏重理论教学外,绝大部分属于应用型专业。经济管理类应用型专业主要着眼于培养社会主义国民经济发展所需要的德智体全面发展的高素质专门人才,要求既具有比较扎实的理论功底和良好的发展后劲,又具有较强的职业技能,并且又要求具有较好的创新精神和实践能力。

在当前开拓新型工业化道路,推进全面小康社会建设的新时期,进一步加强经济管理人才的培养,注重经济理论的系统化学习,特别是现代财经管理理论的学习,提高学生的专业理论素质和应用实践能力,培养出一大批高水平、高素质的经济管理人才,越来越成为提升我国经济竞争力、保证国民经济持续健康发展的重要前提。这就要求高等财经教育要更加注重依据国内外社会经济条件的变化,适时变革和调整教育目标和教学内容;要求经济管理学科专业更加注重应用、注重实践、注重规范、注重国际交流;要求经济管理学科专业与其他学科专业相互交融与协调发展;要求高等财经教育培养的人才具有更加丰富的社会知识和较强的人文素质及创新精神。要完成上述任务,各所高等院校需要进行深入的教学改革和创新,特别是要搞好有较高质量的教材的编写和创新工作。

出版社的领导和编辑通过对国内大学经济管理学科教材实际情况的调研,在与众多专家学者讨论的基础上,决定编写和出版一套面向经济管理学科专业的应用型系列教材,这是一项有利于促进高校教学改革发展的重要措施。

本系列教材是按照高等学校经济类和管理类学科本科专业规范、培养方案,以及课程教学大纲的要求,合理定位,由长期在教学第一线从事教学工作的教师编写,立足于 21 世纪经济管理类学科发展的需要,深入分析经济管理类专业本科学生现状及存在的问题,探索经济管理类专业本科学生综合素质培养的途径,以科学性、先进性、系统性和实用性为目标,其编写的特色主要体现在以下几个方面:

(1) 关注经济管理学科发展的大背景,拓宽理论基础和专业知识,着眼于增强教学内容与实际的联系和应用性,突出创造能力和创新意识。

(2) 体系完整、严密。系列涵盖经济类、管理类相关专业以及与经管相关的部分法律类课程,并把握相关课程之间的关系,整个系列丛书形成一套完整、严密的知识结构体系。

(3) 内容新颖。借鉴国外最新的教材,融会当前有关经济管理学科的最新理论和实践经验,用最新知识充实教材内容。

(4) 合作交流的果。本系列教材是由全国上百所高校教师共同编写而成,在相互进行学术交流、经验借鉴、取长补短、集思广益的基础上,形成编写大纲。最终融合了各地特点,具有较强的适应性。



(5) 案例教学。教材融入了大量案例研究分析内容,让学生在 学习过程中理论联系实际,特别列举了我国经济管理工作中 的大量实际案例,这可大大增强学生的实际操作能力。

(6) 注重能力培养。力求做到不断强化自我学习能力、思维能力、创造性解决问题的能力以及不断自我更新知识的能力,促进学生向着富有鲜明个性的方向发展。

作为高要求,经济管理类教材应在基本理论上做到以马克思主义为指导,结合我国财经工作的新实践,充分汲取中华民族优秀文化和西方科学管理思想,形成具有中国特色的创新教材。这一目标不可能一蹴而就,需要作者通过长期艰苦的学术劳动和不断地进行教材内容的更新才能达成。我希望这一系列教材的编写,将是我国拥有较高质量的高校财经管理学科应用型教材建设工程的新尝试和新起点。

我要感谢参加本系列教材编写和审稿的各位老师所付出的大量卓有成效的辛勤劳动。由于编写时间紧、相互协调难度大等原因,本系列教材肯定还存在一些不足和错漏。我相信,在各位老师的关心和帮助下,本系列教材一定能不断地改进和完善,并在我国大学经济管理类学科专业的教学改革和课程体系建设中起到应有的促进作用。

刘诗白

刘诗白 现任西南财经大学名誉校长、教授,博士生导师,四川省社会科学联合会主席,《经济学家》杂志主编,全国高等财经院校《资本论》研究会会长,学术团体“新知研究院”院长。

前言

统计信息是社会经济信息的主体,统计信息的搜集、整理、分析是国家乃至企业科学决策和科学管理的一项重要基础工作,是制定政策、编制长短期规划的重要依据。经济越发展,越需要有健全的、发达的现代化统计工作。作为经济工作者必须学会统计资料的搜集、整理的方法,学会利用统计数据进行分析,发现问题并解决问题。

现有统计学教材主要为培养研究型人才服务,因而理论性较强。而作为应用型本科院校的统计学任课教师,教学实践使我们清晰地认识到:学生绝大部分对理论接受能力偏弱,加之现有教材体系的枯燥乏味,造成学生学习效果整体不甚理想,随之带来的是就业困难。为满足市场对应用型人才的需求,迫切需要能将理论与实践有机结合起来的教学参考资料。因此,组织编写一本能使应用型院校师生在教与学过程中得心应手的教材已迫在眉睫。

本书具有以下特点。

- (1) 由生活中熟知的小案例作为引导,以增强趣味性。
- (2) 将统计术语通俗化。
- (3) 案例贯穿始终,既有统领全书的整体性案例,又有加深每一章学习的分散化案例分析。
- (4) 理论学习与软件应用适时结合,做到“学完就练,举一反三”。
- (5) 体例编排新颖别致。
- (6) 可作为工作人员的必备工具书。

本书由王雪秋和董小刚担任主编,王明明和赵天宇担任副主编,张纯荣、赵晖和杜春晶担任参编。其中,第1章由王明明编写;第2章由赵天宇编写;第3章、第5章、第6章和第7章由王雪秋编写;第4章和第8章由董小刚编写。张纯荣、赵晖和杜春晶负责本书的校对。

在本书的编写过程中,参编作者在将统计理论的最新研究成果和自己多年教学经验体现在教材内容中的同时,参考了许多统计学研究论文及统计学教材,这些资料对本书的形成提供了极大的帮助,在此向这些统计学文献的作者表示诚挚的感谢!

由于编者水平有限,书中难免有不妥之处,敬请各教学单位和读者在使用本书的过程中给予批评和指正,以便再版时修正。

编者
2015年5月

目 录

第 1 章 总论.....	1	第 3 章 统计数据的整理与图形展示.....	39
1.1 统计学概述.....	1	3.1 定性数据的整理与图形展示.....	40
1.1.1 统计及统计学的含义.....	1	3.1.1 定性数据的整理.....	40
1.1.2 统计学研究的对象及特点.....	2	3.1.2 定性数据的图形展示.....	41
1.1.3 统计学的应用领域.....	2	3.2 定量数据的整理与图形展示.....	43
1.2 统计学的分类.....	3	3.2.1 未分组的定量数据的整理与 图形展示.....	43
1.2.1 描述统计学和推断统计学.....	3	3.2.2 分组的定量数据的整理与 图形展示.....	44
1.2.2 理论统计学和应用统计学.....	4	3.2.3 多维定量数据的整理与 图形展示.....	48
1.3 统计学的基本内容.....	4	3.3 统计表的使用.....	48
1.3.1 统计数据类型.....	4	3.4 案例分析:啤酒市场的调查与分析及 Excel 上机应用——样本组成分析.....	49
1.3.2 统计学中的几个基本概念.....	6	3.4.1 性别结构的分析.....	49
1.4 案例分析:啤酒市场的调查与分析.....	7	3.4.2 年龄结构的分析.....	52
习题.....	8	习题.....	54
第 2 章 统计数据的收集与处理.....	9	第 4 章 统计数据的指标度量.....	57
2.1 统计数据的来源.....	9	4.1 集中趋势的指标.....	57
2.1.1 统计数据的间接来源与处理.....	10	4.1.1 分类数据——众数.....	58
2.1.2 统计数据的直接来源与处理.....	10	4.1.2 顺序数据:中位数和 四分位数.....	59
2.2 抽样调查数据的收集.....	11	4.1.3 数值数据——平均数.....	62
2.2.1 调查方案的设计.....	11	4.1.4 众数、中位数和平均数的 关系.....	63
2.2.2 问卷调查的设计.....	15	4.1.5 众数、中位数和平均数应用的 注意事项.....	64
2.3 统计数据质量.....	19	4.2 离散程度的绝对指标.....	64
2.3.1 统计数据的误差.....	19	4.2.1 分类数据——异众比率.....	64
2.3.2 统计数据的质量要求.....	19	4.2.2 顺序数据——四分位差.....	66
2.3.3 降低统计数据误差的措施.....	19	4.2.3 数值数据——方差和标准差.....	66
2.4 案例分析:啤酒市场的调查与分析及 Excel 上机应用——数据的收集.....	20	4.2.4 相对离散程度——离散系数.....	68
2.4.1 调查问卷的设计.....	20		
2.4.2 自动接收问卷结果的设置.....	24		
2.4.3 “自动统计调查结果”工作表 的隐藏和问卷邮件的发送.....	26		
2.4.4 调查结果资料库的创建.....	27		
习题.....	37		

4.3 数据的相对位置测量——标准分数 ... 68	第6章 假设检验 105
4.4 偏态与峰态的指标度量 69	6.1 假设检验的基本理论 106
4.4.1 偏态及偏态系数 69	6.1.1 假设检验的定义 106
4.4.2 峰态及峰态系数 70	6.1.2 假设检验的基本步骤 106
4.5 案例分析：啤酒市场的调查与分析及 Excel 上机应用——描述性统计 指标 70	6.2 一个总体参数的假设检验 113
4.5.1 不同性别的啤酒印象分数 分布情况 71	6.2.1 一个总体均值的假设检验 113
4.5.2 不同学历的啤酒的印象分数 分布情况 76	6.2.2 一个总体比例的假设检验 119
习题 78	6.2.3 一个总体方差的假设检验 120
第5章 参数估计 82	6.3 两个总体参数的假设检验 122
5.1 几个重要的统计分布 82	6.3.1 两个总体均值之差的假设 检验 122
5.1.1 正态分布 83	6.3.2 两个总体比例之差的假设 检验 128
5.1.2 标准正态分布 83	6.3.3 两个总体方差之比的假设 检验 130
5.1.3 χ^2 (卡方) 分布 83	6.4 案例分析：啤酒市场的调查与分析及 Excel 上机应用——啤酒印象与性别 的相关性分析 132
5.1.4 t 分布 83	习题 135
5.1.5 F 分布 84	第7章 方差分析 140
5.2 样本抽样分布 84	7.1 方差分析的基本理论 141
5.2.1 样本均值的抽样分布 84	7.1.1 方差分析的定义 141
5.2.2 样本比例的抽样分布 86	7.1.2 方差分析中的几个基本 概念 142
5.2.3 样本方差的抽样分布 87	7.1.3 方差分析的基本思路 142
5.3 参数估计的基本理论 87	7.1.4 方差分析的条件 143
5.3.1 参数估计的含义 87	7.2 单因素方差分析 144
5.3.2 参数估计的几个基本概念 87	7.2.1 数据结构 144
5.3.3 评价估计量的标准 89	7.2.2 单因素方差分析的基本 步骤 144
5.3.4 参数估计的思路 91	7.2.3 方差分析表 148
5.4 一个总体的参数区间估计 91	7.2.4 关系强度的测量 149
5.4.1 总体均值的区间估计 91	7.2.5 多重分析比较 149
5.4.2 总体比例的区间估计 96	7.3 案例分析：啤酒市场的调查与分析及 Excel 上机应用——啤酒印象与 学历的相关性分析 151
5.4.3 总体方差的区间估计 98	习题 152
5.5 样本容量的确定 99	
5.5.1 估计总体均值时样本容量的 确定 99	
5.5.2 估计总体比例时样本容量的 确定 100	
习题 101	

第 8 章 相关与一元回归分析	156
8.1 相关分析的基本理论	157
8.1.1 变量间的关系	157
8.1.2 相关分析	157
8.2 一元线性回归分析	162
8.2.1 回归分析的含义	162
8.2.2 一元线性回归模型	162
8.2.3 参数的最小二乘估计	163
8.2.4 样本回归方程的评价	165
8.2.5 一元线性回归方程的统计 检验	168
8.3 一元线性回归模型的预测	174
8.3.1 点估计	174
8.3.2 区间估计	174
8.4 多元线性回归分析	176
8.4.1 多元线性回归模型的含义	176

8.4.2 最小二乘法	177
8.4.3 样本回归方程的评价	177
8.4.4 显著性检验	178
8.5 案例分析：啤酒市场的调查与分析及 Excel 上机应用——啤酒销售量 预测	178
习题	185

附录 用 Excel 生成概率分布表 189

附表 1 标准正态分布表	189
附表 2 标准正态分布临界值表	191
附表 3 t 分布临界值表	192
附表 4 χ^2 分布表	194
附表 5 F 分布临界值表	197

习题答案 199

参考文献 208

第 1 章 总 论

教学目标

1. 掌握统计学的含义及研究对象。
2. 掌握统计学的分类。
3. 掌握统计数据类型。
4. 掌握统计学中的几个基本概念。

1.1 统计学概述

日常生活中，人们经常会使用“统计”这一专业术语，同时也会在有关的媒体中经常看见一些资料使用统计数据、图表等形式。本章将详细讲解统计学的基本原理，包括统计及统计学的含义、统计学研究的对象及特点、统计学的应用领域、统计学的分类、统计数据的类型及统计学中几个基本的概念。

1.1.1 统计及统计学的含义

统计指对某一现象有关数据的收集、整理、计算和分析等的活动。在统计的应用中，人们对“统计”一词的理解一般有 3 种含义：统计工作、统计资料和统计学。

定义 1.1 统计工作是指利用科学的方法收集资料、整理资料、分析资料 and 提供关于社会经济现象数量关系的工作总称。

统计工作是统计资料和统计学的基础。

定义 1.2 统计资料是指通过上面所说的统计工作取得的、用来反映社会经济现象的数据资料总称。

统计工作的成果是统计资料，统计工作所取得的各项数字资料及有关文字资料，通常反映在统计表、统计图、统计手册、统计年鉴、统计资料汇编和统计分析报告。

定义 1.3 统计学是指研究如何收集数据、如何整理数据、如何分析数据和最后解释数据，并从数据中得出规律的一门科学。

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。

统计工作、统计资料、统计学三者之间的关系：利用统计学的理论，指导统计工作，

最后得出统计资料。其中统计学是关于数据的科学,它提供了有关数据收集、数据处理、数据分析、数据解释并从数据中得出结论的方法。

1.1.2 统计学研究的对象及特点

1. 统计学研究的对象

统计学研究的内容是数据,而数据是社会经济现象的数量特征和数量关系的表现,从数据中找出经济规律性。因此,统计学的研究对象为大量社会经济现象的数量特征和数量关系,以揭示其规律性。

2. 统计学研究对象的特点

1) 总体性

一般情况下,统计学研究对象是社会经济现象总体或自然现象总体的数量特征,研究的方法是利用样本的信息推导出总体的数量特征。

例如,要研究某地区居民的收入水平,目的不在于了解个别居民的收入状况,而是要通过对很多个别居民收入状况的了解,达到对全区居民总体收入水平的认识。再如,某企业对其生产的一批日光灯管的平均使用寿命进行研究,不可能把这批日光灯管全部点亮来研究其平均寿命,因为该种研究属于破坏性研究,只能从这批日光灯管中抽取一组样本,对样本进行实验,通过样本的平均寿命来得知总体的平均寿命。

2) 数量性

统计学研究对象的数量性。

对象的数量性包含3个方面的内容:研究对象数量的多少、研究现象间的数量关系和研究对象的质与量间的关系。即通过各种不同的统计指标和指标体系来反映研究对象总体的规模、水平、速度、比例、效益和趋势等。

例如,我国的人口数量构成及其发展趋势、人口结构的构成及发展趋势、国民生产总值的总量构成等。再如,某投资者筹备一超市,筹备前就要研究选址是否会影响超市的营业利润。

1.1.3 统计学的应用领域

统计学的研究内容为数据,只要有数据的存在就会用到统计学。随着定量研究重要性的提高,现代统计学的应用日益广泛,理、工、农、医、文、经,都要用到统计学的。例如,政府部门利用统计学进行宏观调控和管理;日常生活中,统计学是企业管理与决策的依据等。下面主要介绍统计学在经济管理中的一些应用。

1. 企业开发新产品

企业为了在激烈的市场竞争中求得平稳的发展,必然要根据市场的变化在适当的时候引入新产品。引入新产品前,企业要对新产品进行市场定位,包括市场分析、对新产品的价格进行市场定位、对消费群体进行定位等,只有全面分析,新产品开发才能获胜。同时,新产品进入市场后,企业要不断进行产品跟踪调查,不断完善新产品。而这些离不开统计学,它们需要统计学提供可靠的数据,对数据进行分析,得出决策的信息。

2. 财务分析

上市公司的财务报表中的数据是投资者进行决策的重要参考依据。投资者分析上市公司的财务报表数据的定量关系，最后进行决策。企业自身的投资也离不开财务数据的分析。

3. 竞争对手的研究

随着社会经济的发展，各个行业内部之间的竞争也日益加剧。企业要想取得优势，抢占时机，就要不断地了解竞争对手，摸清对手的竞争策略，发现其弱势，利用自己的优势，制定获胜的策略，而做这些的前提是要进行数据分析。

1.2 统计学的分类

1.2.1 描述统计学和推断统计学

1. 描述统计学和推断统计学的定义

统计学按照统计数据分析的方法不同，分为描述统计学和推断统计学。

定义 1.4 描述统计是指研究数据收集、处理数据和描述数据的统计学分支。

描述统计学的内容包括如何取得要研究的数据、用什么样的图表对数据进行处理和显示，进而通过综合、概括与分析，得出反映所研究现象的一般性特征。

例如，反映长春大学光华学院经济系某年级的统计学成绩的条形图即属于描述统计学，其统计学成绩数据见表 1-1 所示。

表 1-1 统计学成绩

成绩结果	频 数
优秀	7
良好	15
中等	16
及格	14
不及格	3
合计	55

该组数据采用条形图来描述，利用 Excel 软件来绘制，绘制结果如图 1.1 所示。

定义 1.5 推断统计学是指通过研究如何根据统计样本的资料计算样本的特征信息，来推断总体相关的特征的方法。

例如，从一个果园中采摘 40 个橘子，利用这 40 个橘子的平均重量来估计整个果园所有橘子的平均重量。

2. 描述统计学与推断统计学的关系

一般来讲，描述统计学是现代统计学的前期工作，是推断统计学的基础；推断统计学是现代统计学的核心，是描述统计学的发展。这两部分是统计学的两个不可缺的组成内容。

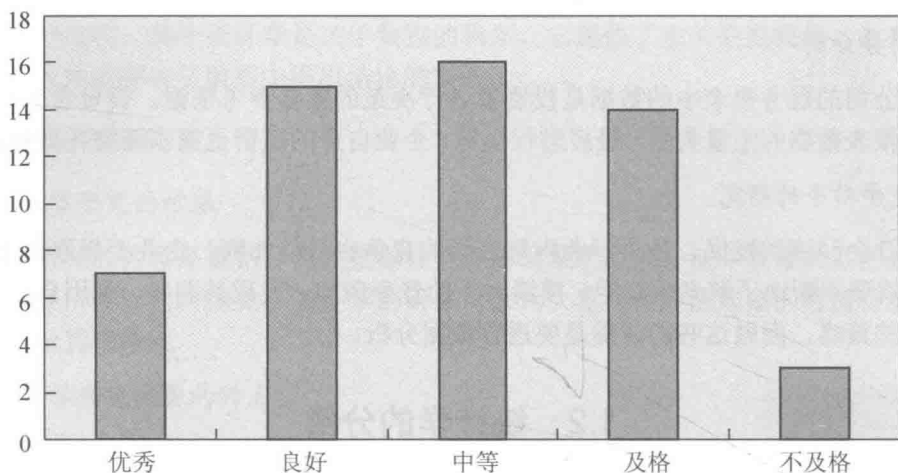


图 1.1 统计学成绩条形图

例如，想研究某城市居民的消费水平，而往往这个城市的所有居民消费水平的数据是很难收集的，这时，需要对这个城市的一小部分居民的消费状况数据进行收集，通过描述统计学，了解样本居民的消费状况，而后用到推断统计学，在对样本居民消费状况的了解情况下，达到对全市居民总体消费水平的认识。

1.2.2 理论统计学和应用统计学

1. 理论统计学和应用统计学的定义

统计学按照研究的目的不同，可分为理论统计学和应用统计学。

理论统计学把研究对象一般化、抽象化，以概率论为基础，从纯理论的角度，对统计方法加以推导论证，以归纳方法研究随机变量的一般规律。

应用统计学侧重于统计学的应用，它研究如何应用统计学理论和方法，对实际的问题进行研究，以揭示其各种经济现象的规律性。其目的是解决经济存在的问题，对经济增长进行预测。

2. 理论统计学和应用统计学的关系

理论统计学和应用统计学的关系十分密切。理论统计学为应用统计学提供了统计的理论和方法；而应用统计学是理论统计学的发展或延伸。

1.3 统计学的基本内容

从统计学的定义中可以看出统计学的核心是数据，所以在学习统计学内容之前要掌握数据的类型。

1.3.1 统计数据的类型

统计学在研究事物的数量方面是离不开数据的，如研究股票价格变动水平就要收集数据，计算出股票价格指数的指标来分析等，但数据不同，分析的方法不同。

1. 分类数据、顺序数据、数值型数据

按照计量尺度的不同,统计数据可分为分类数据、顺序数据和数值数据。

定义 1.6 只归于某一类别的非数字型数据,称为分类数据。

分类数据是对事物进行分类,该数据表现的是文字叙述,如人口按性别可分为男、女两类。

定义 1.7 只能归于某一有序类别的非数字型数据,称为顺序数据。

顺序数据也是对事物进行分类的结果,如某大学的选修课以优、良、中、及格和不及格分成 5 级。

定义 1.8 按数字尺度测量的观察值,称为数值数据。

数值数据是对事物进行了准确的测量,表现具体的数值,如某大学的统计学成绩。

在这里要注意以下两点。

(1) 分类数据是指归于某一类别非数字型数据,而顺序数据是归于某一有序的类别,即在分类数据的基础上加一个条件“有序”。从这里可以得出,顺序数据比分类数据高一层。

(2) 数值数据是对事物进行了准确的测量,表现具体的数值。例如,某专业学生的“英语成绩”为数值数据,但有时为了了解成绩的状况,要对数据进行分析,即 90 分以上的学生归于优秀,80~89 分的学生归于良好,70~79 分的学生归于中等,60~69 分的学生归于及格,最后 60 分以下的归于不及格。所以说,从这个例子可以得知 3 种数据,最高级数据为数值数据,其次是顺序数据,最低级的数据为分类数据。

2. 观测数据和实验数据

按照统计数据的收集方法不同,统计数据可分为观察数据和实验数据。

定义 1.9 通过调查或者观测而收集到的数据,称为观察数据。

观察数据是在没有对事物进行人为控制的条件下得到的数据,有关社会经济现象的统计数据几乎都是观察数据。

例如,某一时间段,某市交通广播电台发布信息:延安大街由南向北车流量比较大,请各位司机避让。此数据“车流量大”就是观察数据。再如,某人对寄居在家中的一窝燕子很感兴趣,他每天观察并记录燕子飞出飞入的时间,以及它们喂养小燕子的习惯,那么他得到的燕子飞出飞入的时间数据即为观察数据。

定义 1.10 在实验中控制实验对象而收集到的数据,称为实验数据。

实验数据是对事物进行了人为控制而收集到的数据。在自然科学领域中所使用的统计数据大多是实验数据。

例如,某医药企业研发新药,数据通常是通过对小白鼠进行实验而得到的实验数据。

3. 截面数据和时间序列数据

按照被描述的对象与时间之间的关系,统计数据分为截面数据和时间序列数据。

定义 1.11 在相同或近似相同的时间点上所收集的数据,描述现象在某一时刻的变化情况,称为截面数据。

定义 1.12 在不同时间上所收集的数据,用来描述现象随时间而变化情况的数据,称为时间序列数据。

例如, 1994—1999 年我国国内生产总值见表 1-2 所示。

表 1-2 1994—1999 年我国国内生产总值

单位: 亿元人民币

年 份	第一产业	第二产业	第三产业
1994	9 457.2	22 372.2	14 930
1995	11 993	28 537.9	17 947.2
1996	13 844.2	33 612.9	20 427.5
1997	14 211.2	37 222.7	23 028.7
1998	14 552.4	38 619.3	25 173.5
1999	14 457.2	40 417.9	27 035.8

其中行数据为截面数据, 列数据为时间序列数据, 即有 3 组时间序列数据, 6 组截面数据。

1.3.2 统计学中的几个基本概念

统计学主要是想从样本的信息推断出总体的特征值。在进行研究时, 涉及很多概念, 其中有几个概念是经常用到的。

1. 总体和样本

定义 1.13 包含所研究的全部个体(数据)的集合, 称为总体。

例如, 要检验一批奶粉是否合格, 那么这批奶粉构成的集合就是总体, 其中每一袋奶粉就是总体中的个体。

总体根据包含的个体数目是否可数, 分为有限总体和无限总体, 即总体的范围确定有时比较容易, 有时较难。例如, 要检验一批灯泡的使用寿命, 那么这批灯泡构成的集合就是总体, 较易确定。再如, 某企业推出一种新产品, 想了解消费者是否喜欢, 这时它就需要先确定消费对象, 即要确定消费这种新产品的消费者总体, 这时该总体的范围确定就较难。所以说, 在实务中, 我们可根据自己的研究目的来定义总体。

定义 1.14 从总体中抽取一部分个体重新组成一个集合, 称为样本。

从总体中抽取一部分个体作为样本, 目的是要根据样本提供的有关信息去推断总体的信息。

例如, 估计一批灯泡的平均寿命, 不能把所有灯泡点亮, 因为这种试验属于破坏性试验。所以要从这批灯泡中随机抽取一小部分作为样本, 进行测试, 从而得出样本的平均寿命, 通过此信息推断出这批灯泡的平均寿命。

再如, 研究全国大学生平均月消费结构的状况。通常全国大学生这个总体的数据不易得到, 只能抽取样本, 通过对样本的信息研究推断出总体的信息。

定义 1.15 构成样本的个体数目, 称为样本容量, 或称为样本量。

2. 参数和统计量

定义 1.16 用来描述总体特征的概括性数字度量, 称为参数。

参数是研究者想要了解总体信息的特征值。在统计学中, 我们最关心的是总体的均值、

总体的方差、总体的比例等。这些参数通常是未知的。

例如，某企业想了解今天生产的这批灯泡的平均寿命，即总体的均值；某投资企业想了解它的投资组合的风险，即总体方差；某企业想了解今天生产的这批灯泡的次品率，即总体的比例。

在统计中，总体参数通常用希腊字母表示，如总体平均数用 μ 表示，总体标准差用 σ 表示，总体比例用 π 表示等。

定义 1.17 用来描述样本特征的概括性数字度量，称为统计量。

统计量是根据样本数据计算出来的一个已知的量。通常，最关心的是样本的平均数、样本的方差、样本的比例等。样本统计量也用英文字母表示。例如，样本平均数用 $\bar{x}$ 表示，样本标准差用 s 表示，样本比例用 p 表示等。

抽取样本的目的是根据样本的统计量估计总体的参数，即用样本平均数 $\bar{x}$ 估计总体平均数 μ ，用样本标准差 s 估计总体的标准差，用样本比例 p 估计总体比例 π 。

3. 变量

定义 1.18 说明现象某种特征的概念，称为变量。

变量按照不同的划分标准有不同的分类：

- (1) 按数据的类型，变量可分为分类的变量、顺序变量和数值变量。
- (2) 按数值是否可数，变量分为离散型变量和连续型变量。

1.4 案例分析：啤酒市场的调查与分析

随着经济的快速发展，人们的生活水平也日益提高了。在经济不发达的时候，人们对一种产品是否消费，主要是受产品的价格因素影响，但随着经济的发展，人们的生活水平也在不断地提高，人们对一种产品的消费，不单单会受产品的价格因素影响，还有其他因素，如品牌、产品的售后服务、产品的外观、性能等因素的影响。

啤酒已成为一种日常消费品进入千家万户，且啤酒市场的竞争也一直从未停止过，各类啤酒犹如雨后春笋般不断地推陈出新。2003 年，自我国加入 WTO(World Trade Organization, 世界贸易组织)后，开放的中国市场已逐渐融入到了世界经济的均衡游戏中。尤其从 2005 年开始，外资进入中国啤酒业的步伐更为快速，随着外资收购速度的加快，中国啤酒市场的竞争正在发生变化，要想在这激烈的竞争市场中处于不败之地，抓住市场、扩展市场才是唯一的出路。所以有必要进行新一轮的啤酒市场调查与分析，以准确地知道是哪些因素影响消费者的购买行为。

要想准确地知道是哪些因素影响消费者的购买行为，首先要收集数据，其次要整理数据，再次要分析数据，最后解释数据，从数据中得出想要的信息。

收集数据前要确定研究的总体。从上面的分析可知，总体是购买啤酒的消费者，这一总体的范围很难确定，即总体的数据很难收集，所以使用推断性统计得出总体的信息，即收集样本的数据，根据样本的信息来推断总体的信息。

啤酒市场的调查与分析流程如图 1.2 所示。