

21世纪应用型本科经管类规划教材
Shiji Yingyongxing Benke Jingguanlei Guihua Jiaocai



应用统计学

YINGYONG TONGJIXUE

● 主 编 余 勃



华南理工大学出版社
SOUTH CHINA UNIVERSITY OF TECHNOLOGY PRESS

014060568

21世纪应用型本科经管类规划教材

Shiji Yingyongxing Benke Jingguanlei Guihua Jiaocai

内容简介

C8-43
232



应用统计学

YINGYONG TONGJIXUE

● 主 编 余 勃



华南理工大学出版社
SOUTH CHINA UNIVERSITY OF TECHNOLOGY PRESS



北航

C1748181

C8-43
232

832030210

内 容 简 介

本书编写思路定位是：以统计学学科知识为基础内容，以综合素质培养和实际能力训练为基础目标，争取形成内容丰富、功能齐全、形式多样、系统完整、使用方便的综合性应用型人才培养统计学教材内容体系。内容涵盖统计概述、数据图表展示、数据的概括性度量、概率和概率分布、抽样与参数估计、假设检验、相关与回归分析、时间序列、指数、数据分析应用。本书可供应用型本科院校经管类专业师生使用。

图书在版编目 (CIP) 数据

应用统计学/余勍主编. —广州：华南理工大学出版社，2014. 8
21 世纪应用型本科经管类规划教材
ISBN 978 - 7 - 5623 - 4316 - 5

I. ①应… II. ①余… III. ①应用统计学 - 高等学校 - 教材 IV. ①C8

中国版本图书馆 CIP 数据核字 (2014) 第 151207 号

应用统计学

余 勍 主 编

出 版 人：韩中伟

出版发行：华南理工大学出版社

(广州五山华南理工大学 17 号楼，邮编 510640)

http: //www. scutpress. com. cn E-mail: scutc13@scut. edu. cn

营销部电话：020 - 87113487 87110964 22236378 87111048 (传真)

策划编辑：朱彩翩

责任编辑：朱彩翩

印 刷 者：广州市穗彩印务有限公司

开 本：787mm × 1092mm 1/16 印张：12.75 字数：295 千

版 次：2014 年 8 月第 1 版 2014 年 8 月第 1 次印刷

印 数：1 ~ 3 000 册

定 价：26.00 元

版权所有 盗版必究 印装差错 负责调换



前言

应用统计学是研究如何科学地搜集、整理、分析、解释数据，并通过对数据的解释探求客观事物内在变化规律性的一门学科。在实践中，无论对经济现象、社会现象，还是自然现象的研究，都离不开对有关数据的分析与解释，以便帮助人们进行决策。因此，它的应用范围十分广泛。学习统计学的重要意义不仅在于可以掌握一种收集数据和解释数据的手段，还可以掌握一种探究现象发展变化规律的分析方法，使人们的决策建立在科学的基础之上，在激烈的市场竞争中立于不败之地。

“应用统计学”是普通高等院校经济、管理类专业本科生一门重要的公共基础课程，一直被列为教育部核心课程之一，本教材为适应经济、管理类非专业统计学教学的需要而编写。该教材能使经济、管理类学生掌握系统的统计学基础知识，在体系及内容上规范详实，做到由浅入深、循序渐进，注重统计学基本原理、基本知识的阐述和基本统计方法的训练，力求把社会经济统计与数理统计方法相结合，使读者更容易接受统计学的理论与方法，并运用计算机软件 Excel 结合实例进行统计运算，可增强学生学习的兴趣、提高学习效率和实践操作能力。通过本书的学习，可培养学生的统计意识和能力，能够运用基本的统计理论与方法，分析和解决经济、管理中出现的实际问题。本教材在体例上，每章开篇设有学习目的，提出教学要点和对学生的教学要求，同时设有案例导入，引导学生进入章节内容并思考，每章末附有本章小结，对教学内容的重点进行了扼要的概括，并在每章后附有课后练习，以便学生理解和掌握所学内容。

本教材第1、3、4、10章由余勍编写，第2、9章由管青编写，第5、6章由史永平编写，第7章由陈小慧编写，第8章由吴锦桂编写。在教材的编写过程中，参阅并吸收了大量国内外最新出版的同类教材及统计学方面的研究成果。虽然执笔人员有多年从事经济、管理类统计学课程教学的经验，在写作和修改的过程中也竭尽全力，但由于编者水平有限，书中不足之处在所难免，敬请同行、读者不吝赐教，以便我们做进一步的修正和完善。

编者
2014年6月1日

目 录

第一章 统计概述	1
1.1 统计与统计数据	2
1.2 数据的类型	4
1.3 数据来源	4
1.4 统计学的基本概念	7
1.5 常用的统计软件	7
本章小结	8
课后习题	8
第二章 数据的图表展示	10
2.1 数据的预处理	11
2.2 品质型数据的整理与展示	20
2.3 顺序型数据的整理与展示	25
2.4 数值型数据的整理与展示	27
2.5 合理使用图表	36
本章小结	38
课后习题	38
第三章 数据的概括性度量	40
3.1 集中趋势的测度	41
3.2 离散趋势的测度	46
3.3 偏态与峰态	50
本章小结	52
课后习题	53
第四章 概率和概率分布	55
4.1 随机事件与概率	56
4.2 概率的性质和运算法则	58
4.3 离散型随机变量的概率分布	60
4.4 连续型随机变量的概率分布	65

本章小结	71
课后习题	71
第五章 抽样与参数估计	73
5.1 抽样调查	73
5.2 抽样分布	76
5.3 参数估计基本原理	81
5.4 总体均值的区间估计	84
5.5 总体比例的区间估计	87
5.6 样本容量的确定	88
本章小结	89
课后习题	90
第六章 假设检验	92
6.1 假设检验的基本原理	93
6.2 单样本总体均值的检验	99
6.3 单样本总体比例的检验	101
6.4 用 P 值做决策	102
本章小结	104
课后习题	104
第七章 相关与回归分析	105
7.1 相关关系	106
7.2 回归分析	113
7.3 一元线性回归分析	114
本章小结	126
课后习题	127
第八章 时间序列	131
8.1 时间序列的相关概念	132
8.2 增长率	135
8.3 平稳序列预测方法	138
8.4 线性趋势序列预测方法	143
8.5 非线性趋势序列预测方法	145
本章小结	148
课后习题	148

第九章 指数	152
9.1 指数的概念与分类	153
9.2 总指数编制的基本问题	154
9.3 加权总指数的编制方法	158
9.4 价值指数与指数体系	161
9.5 几种常用的价格指数	163
本章小结	165
课后习题	165
第十章 数据分析应用	168
10.1 因子分析	169
10.2 方差分析	173
本章小结	180
课后习题	180
总复习题	182
总复习题一	182
总复习题二	184
附表1 标准正态分布表	189
附表2 t 分布表	191
参考文献	193

第一章 统计概述

学习目标

- 掌握统计学的基本概念，理解为什么要学习统计学；
- 懂得如何运用统计学；
- 能区分什么是描述统计学和推断统计学；
- 了解如何收集数据；
- 知道常用的统计软件。

本章将介绍统计学的一些基本问题，包括统计学的含义、统计数据及其分类、统计中常用的一些基本概念等。

案例导入

一生一世的统计数字

《今日文摘》2009年第1期有篇文章叫《一生一世的统计数字!》，文中有这样一段话：“一个人赤条条来到世界，离开的时候，也带不走多少东西。但是一个人的生给地球留下了什么？又创造了什么？以下是人活一世的一组数据。数据以英国人的生活方式为标准，兼顾了世界各地的人。这组数据可以给我们一个参照，也可以给我们很多思考。英国人平均寿命78.5岁，共24亿零750万秒……人一生吃掉的东西：4头牛、15头猪、21只羊、1200只鸡、13000枚鸡蛋（未出生的鸡）、5000多只苹果、10000多个胡萝卜、3吨面包、630公斤巧克力、2吨葡萄酒、11吨啤酒、18吨牛奶、75000杯茶，相当于装满一个浴缸的罐头豆子，一生总共吃下约50吨食物。当然，这是人的平均数，也考虑到穷人吃得少些，富人吃得多些。”

这些数据未必都准确，但信息很清晰。这里既有无法改变的，也有可以改变的；既有真实的生活，也有背后的贫富差距。

我们如何看待这些统计数据，其实就是如何看待自己，也是如何看待人类。是增加这些数据，还是减少这些数据，不光对我们自己，对人类都会产生影响。这就是统计科学的力量。

1.1 统计与统计数据

在日常工作和生活中，到处都有统计数据。例如，学生考试后非常关心自己的考试成绩和名次；企业管理人员要掌握生产销售情况和利润额；人们可以通过报刊和电视等获得 GDP（国内生产总值）、CPI（消费者价格指数）和经济增长率等数据；等等。日常工作与生活中的这些数字都是人们关心的统计数据。

什么是统计学？以下从三个方面进行定义：

定义 1 统计学是一门收集、处理、分析、解释数据并从数据中得出结论的科学，其目的是探索数据内在的数量规律性。统计学的定义告诉我们，统计学总是在和数据打交道，因而也可称为“数据的科学”。从统计方法来看，统计学提供了一系列的方法，专门用来收集数据、整理数据、展示数据的特征，进而分析和探索（或推断）出事物总体的数量规律性。

统计学的英文是“Statistics”。它通常有两个含义：当它以单数名词出现时，表示一门科学的名称——“统计学”；当它以复数名词出现时，表示“统计数据”或“统计资料”。“Statistics”一词的英文解释至少可以说明两件事情。

第一，由于统计数据在英文中是以复数形式出现的，表明统计数据不是指个别的单个数字，而是指同类的较多数据。因为单个数据如果不和其他数据进行比较，是不能说明问题的。例如，某个学生在某门课程的考试中得了 85 分，如果仅凭这一数字，我们很难对这位学生的知识和能力水平作出判断和评价。因为这个 85 分可能是班上的最高分，也可能是中等水平的分数，还可能是较低的分数。如果还知道这次考试的平均分数，就可以对这位学生的成绩是高于还是低于平均分数，以及高多少或者低多少进行评价了。在生产和生活实际问题中，通常可以收集到较多的数据，进而利用统计方法对数据进行加工整理，从而发现数据中的内在联系及数量规律。

第二，作为单数的“统计学”和作为复数的“统计数据”在英文中使用同一名词，显示出两者之间的密切关系。统计学是由收集、处理、分析、解释统计数据的方法组成的，这些方法来源于对统计数据的研究，目的也在于对统计数据的研究。离开了统计数据，统计方法乃至统计学就失去了其存在的意义。

定义 2 研究数据收集、处理和描述的统计学方法，称为描述统计（Descriptive Statistics）。描述统计即用图形、表格和概括性的数字对所研究现象的一般性特征进行展示。描述统计是对数据进行处理的第一阶段。

定义 3 研究如何利用样本数据来推断总体特征的统计学方法，称为推断统计（Inferential Statistics）。推断统计是根据样本信息对总体进行参数估计、假设检验、预测或其他推断的统计方法。

统计学分为描述统计和推断统计，一方面反映了统计发展的前后两个阶段，另一方面也反映了统计方法研究和探索客观事物内在数量规律性的先后两个过程。统计研究过程的起点是数据，终点是探索到客观事物总体内在的数量规律性。要达到统计研究的目

的,如果收集得到的是总体数据,则经过描述统计之后就可以达到探索内在数量规律的目的了;但如果所获得的数据只是研究总体的一部分数据,要得到总体的数量规律就必须根据样本整理出来的信息对总体做出科学的推断。显然,描述统计是整个统计学的基础和统计研究工作的第一步。它包括对客观现象的度量,调查方案的设计,科学、及时、快速、经济地收集与整理数据,用图表显示数据,分析和提取数据中的有用信息以最终推断总体;推断统计是现代统计学的核心和统计研究工作的关键环节,因为统计最终能否科学准确地探索到总体内在的数量规律性与选用何种统计量、选用什么推断方法、如何进行推断有着直接的联系。

统计学的应用领域非常广泛,是一门适用于几乎所有学科的通用数据分析方法。无论是学术研究、政府管理,还是公司或企业的生产经营管理,都离不开统计方法的应用。表1-1列出了统计的一些应用领域,从中可以看到,统计无处不在。

表1-1 统计的应用领域

精算	水文学
农业	工业
动物学	语言学
人类学	文学
考古学	劳动力计划
审计学	管理科学
晶体学	市场营销学
人口统计学	医学诊断
牙医学	气象学
生态学	军事科学
经济计量学	核材料安全管理
教育学	眼科学
选举预测和策划	制药学
工程学	物理学
流行病学	政治学
金融	心理学
水产渔业研究	心理物理学
博彩业	质量控制
遗传学	宗教研究
地理学	社会学
地质学	抽样调查
历史研究	分类学
人类遗传学	气象改善

资料来源:贾俊平,何晓群,金勇进.统计学[M].北京:中国人民大学出版社,2007.

1.2 数据的类型

统计数据按照所采用计量尺度的不同可以分为三种类型。一是数值型数据，是指用数字尺度测量的观察值，如 GDP 等宏观经济运行数据、每天进出海关的旅游人数、某地流动人口的数量等。数值型数据表现为具体的数值，统计处理中的大多数数据都是数值型数据。二是分类型数据，是指对数字进行分类的结果，如人口按性别分为男、女两类；受教育程度也可以按照不同类别来区分。这种数据通常用频数（率）来表示分类的结果，如某地区男性常住人口占 52%，女性占 48%。三是顺序型数据，是指数据不仅是分类的，而且其类别是有序的。例如，满意度调查中的选项有“非常满意”“比较满意”“比较不满意”“非常不满意”等。在这三类数据中，数值型数据由于说明了事物的数量特征，因此可归为定量数据或数量数据，分类型数据和顺序型数据由于定义了事物所属的类别，说明了事物的品质特征，因而可统称为定性数据或品质数据。区分数据的类型非常重要，因为不同类型的数据在一些情况下需要用不同的统计方法进行处理。

此外，还可以从其他角度对统计数据进行划分：

(1) 按照收集方法的不同，可将统计数据分为观测数据和试验数据两类。观测数据在没有对事物进行人为控制的条件下得到，是通过调查或观测收集到的数据，主要集中在社会经济领域。例如，对商品零售价格水平的测量可以得到商品零售价格指数，对股票价格变动水平的测量可以得到股票价格指数。试验数据是指通过在试验中控制实验对象而收集得到的数据，主要集中在自然科学领域，如某种新型电池的使用寿命等。

(2) 按照是否与时间相联系，还可以将统计数据分成截面数据和时间序列数据。截面数据描述了现象在某一时刻的变化情况，时间序列数据描述了现象随时间变化的情况。

区分数据的类型非常重要。因为对不同类型的数据，我们需要采用不同的统计方法来处理和分析。比如，对于分类型数据，我们通常计算出各组的频数或频率，计算其众数和异众比率，进行列联表分析和卡方检验等；对于顺序型数据，我们可以计算其中位数和四分位差等；对于数值型数据，我们可以用更多的统计方法进行分析，如计算各种统计量、进行参数估计和检验等。

1.3 数据来源

应用统计方法分析问题首先要取得数据。从使用者的角度来看，数据主要来源于两种渠道：一是来源于直接的调查和科学实验的数据，对使用者来说，这是数据的直接来源，称为第一手数据或直接数据；二是来源于已有的数据，对使用者来说，这是数据的间接来源，称为第二手数据或间接数据。

1.3.1 直接获取的数据

在进行科学研究和管理决策时,若没有现成的数据可以利用,就需要专门组织调查或进行科学实验。对于社会经济管理和决策而言,主要是通过统计调查的方式获取数据,如客户满意度调查、电视收视率调查、家庭收支调查等。由于抽样调查是一项技术含量相当高的工作,从制订调查方案到抽取样本,从调查到数据整理,都需要有专门的技能和培训,因此调查公司和调查业因市场的需求而发展迅速。统计调查的方法主要有以下几种:

1. 普查

普查是为某一特定目的,专门组织的一次性全面调查。这是一种摸清国情、国力的重要调查方法。世界各国都定期地进行各种普查。普查是适合特定目的、特定对象的一种调查方式,它主要用于收集处于某一时点状态的社会经济现象的数量,目的是掌握特定社会经济现象的基本全貌,为国家制定有关政策或措施提供依据。

目前,我国进行的普查主要有人口普查、农业普查和经济普查。其中,经济普查的内容包括工业普查、第三产业普查、基本单位普查和建筑业普查等。由于普查涉及面广、调查单位多,需要耗费大量的人力、物力、财力和时间,通常需要间隔较长的时间进行一次。在我国的普查中,经济普查每5年进行一次,其他普查每10年进行一次。每逢年份的末位数字为“0”的年份进行人口普查,每逢“6”的年份进行农业普查,每逢“3”和“8”的年份进行经济普查。

2. 抽样调查

抽样调查是统计调查中应用最广、最为重要的调查方法,它是通过随机样本对总体数量规律性进行推断的调查研究方法。它具有以下几个特点:

(1) 经济性。这是抽样调查的一个最显著特点。由于调查的样本单位通常是总体单位中的很小一部分,调查的工作量小,因而可以节省大量的人力、物力、财力和时间,其调查费用较低。

(2) 时效性强。抽样调查可以迅速、及时地获取所需要的信息。由于工作量小,调查的准备时间、调查时间、数据处理时间等都可以大大缩减,从而提高数据的时效性。与普查等全面调查相比,抽样调查可以频繁地进行,并随事物的发生和发展及时取得有关信息,以弥补普查等全面调查的不足。比如在两次人口普查之间各年份的人口数据都是通过抽样调查取得的。

(3) 适应面广。抽样调查可以获得更广泛的信息,它适用于对各个领域、各种问题的调查。从适用的范围和问题来看,抽样调查可用于调查全面调查能够调查的现象,也能调查全面调查所不能调查的现象。抽样调查特别适合对一些特殊现象的调查,如产品质量检验、农产品实验、医药的临床实验等。从调查的项目和指标来看,抽样调查的内容和指标可以更详细、深入,能获得更全面、更广泛和更深入的数据。

(4) 准确性高。抽样调查的数据质量有时比全面调查更高,因为全面调查的工作量大、环节多,登记性误差往往很大;而抽样调查由于工作量小,可使各环节的工作做

得更细致,误差往往较小。当然,用样本数据去推断总体时,不可避免地会有推断误差,但这种误差的大小是可以计算并加以控制的,因此推断的结果通常是可靠的。

由于上述特点,抽样调查已经成为科学研究及管理决策最重要的方法之一。

3. 统计报表

按照国家有关规定,自上而下地统一布置、自下而上地逐级提供基本统计数据的方法为统计报表。统计报表是收集统计数据的一种重要方式,在我国几十年的政府统计工作中,已形成了一套比较完备的统计报表制度,它已成为国家和地方政府部门统计数据的主要来源。

1.3.2 间接获取的数据

对大多数使用者来说,亲自去做调查往往是不可能的或不必要的。他们所使用的数据大多是别人调查或科学实验的数据,这些数据对使用者来说是二手数据。在科学研究和管理决策中,要善于利用各种现成的数据。这种数据既可以从报纸、图书、杂志、统计年鉴、网络等渠道获得,也可以从调查公司或数据库公司等处购买。近年来,互联网已成为数据来源的重要渠道,几乎所有的政府机构和大公司都有自己的网站并提供公共访问端口,访问者可以从中获得有用的数据。表1-2给出部分政府重要网站,每个网站都建有可供公众访问的数据库。

表1-2 提供统计数据的部分网站

	相关网站	网 址	数据内容
中国	国家统计局网站	www.stats.gov.cn	统计年鉴、统计月报等
	中国经济信息网	www.cei.gov.cn	经济信息及各类网站
	华通数据中心	Data.acmr.com.cn	国家统计局授权的数据中心
	中国决策信息网	www.china-policy.com	决策知识及案例
美国	人口普查局网站	www.census.gov	人口和家庭等
	联邦政府数据网站	www.fedstats.gov	美国政府100多个部门数据
联合国	各国际组织数据网站	Data.un.org	联合国各国际组织数据
联合国	各国际组织数据网站	Data.un.org	联合国各国际组织数据

资料来源:贾俊平,何晓群,金勇进.统计学[M].北京:中国人民大学出版社,2007.

利用二手数据对使用者来说既经济又方便,但使用时应注意统计数据的含义、计算口径和计算方法,以避免误用或滥用。同时,使用者在引用二手数据时,一定要注明数据的来源,以尊重他人的劳动成果。

1.4 统计学的基本概念

1. 总体与参数

为研究某一问题,通常需要对研究对象进行界定。统计学将所要研究的全部个体称为总体(Population),将其特征的一些概括性数字度量称为参数(Parameter)。

总体范围的确定有时比较容易。比如,要检验一批灯泡的使用寿命,那么这一批灯泡构成的集合就是总体,其中的每一个灯泡就是一个个体,该总体的范围很清楚。但在有些场合,总体范围的确定则比较困难。比如,对于新推出的一款饮料,要想知道消费者是否喜欢,首先必须弄清楚哪些人是消费者,也就是要确定构成该饮料的消费者这一总体。但事实上,我们很难确定哪些消费者消费该饮料,所以该总体范围的确定十分复杂。当总体的范围难以确定时,我们可以根据研究的目的来定义总体。

参数是研究者想要了解的总体的某种特征值,其种类可以有很多,但研究者关心的主要有以下几种:总体平均数(μ)、总体方差(σ^2)、总体比例(π)等。由于总体数据通常是不知道的,因此参数是一个未知的常数。比如,我们不知道某一班级所有学生的平均年龄,不知道一个城市所有家庭的收入差异,不知道一批产品的合格率等,正因如此,我们才进行抽样,然后根据样本计算出某些值,进而估计出总体参数。

2. 样本与统计量

样本(Sample)是总体的一部分单位。例如,一家公司在接受审计,审计人员没有必要对该公司年度内的所有发票全部审查,只需随机抽查100张发票的样本即可,审计人员通过这100张发票样本数据计算的差错率就可以对全部发票的差错率进行推断。

用来描述样本特征的概括性数字度量称为统计量(Statistics)。统计量是根据样本数据计算出来的一个量,它是样本的函数。统计量具有不确定性,根据所抽取的样本数据有所不同。和总体参数相对应,样本统计量主要有以下几种:样本平均数($\bar{x}$)、样本方差(s^2)、样本比例(p)等。

3. 变量

在研究总体时,重点关注的是总体单位具有哪些特征或属性,我们把这些特征称为变量(Variable)。变量是总体中个体单位所具有的特征或特性。“变量”的名称是针对总体中每一基本单位的属性都存在着差异而言的。例如,被调查的每位失业者的年龄、性别和收入等都不能事先确定,并且存在着差异。

事实上,所有随机取值的数据都归属于某个变量,是变量的某一次具体实现。数据可划分为定性数据和定量数据两类。类似的变量的类型也划分为定性变量和定量变量。性别、受教育程度等都是定性变量,而商品销售额、职工工资等都属于定量变量。

1.5 常用的统计软件

统计软件种类很多,这里简要介绍一下最常见的几种统计软件。

1. SPSS

SPSS 的全称是 Statistical Product and Service Solution, 即统计产品与服务解决方案, 是目前非常受欢迎的一款统计软件。SPSS 囊括了各种成熟的统计方法与模型, 同时提供了各种数据准备与整理技术, 其非常突出的一个特点就是易用性强, 是众多统计软件中为数不多的不需要编程的软件之一。人机界面的友好、简单的操作, 使得 SPSS 拥有广大的用户群。另外, SPSS 在处理抽样调查数据上也拥有其他统计软件不可比拟的优势。

2. Excel

严格来说, Excel 不是一款统计软件, 但它自带了一些统计计算功能, Excel 设计了种类十分齐全的统计函数, 并且通过加载宏安装数据分析的功能, 能够实现一些诸如方差分析、线性回归等简单的统计功能。

3. SAS

SAS 系统的全称为 Statistics Analysis System, 在数据处理和统计分析领域, SAS 系统也是一款权威的统计软件。由于采用模块化设计, SAS 的功能十分强大, 几乎囊括了所有的统计方法与模型。SAS 现已成为一套完整的计算机语言和使用程序方式, 用户可以完成所有需要做的工作, 包括统计分析、预测、建模和模拟抽样等。

4. Eviews

在时间序列数据的分析和处理上, Eviews 是一款非常专业的软件, 可用于多种常用的计量经济模型。它通过建立序列间的统计关系式, 实现预测和模拟等功能。该软件在科学数据分析与评价、金融分析、经济预测、销售预测和成本分析等领域应用广泛。

当然还有许多其他的统计软件, 在此不一一罗列。值得注意的是, 统计软件的使用必须建立在用户熟悉相关统计理论和方法的基础上, 否则容易导致误用。

■ 本章小结

本章通过介绍统计数据与统计学、统计数据的来源及统计数据的种类和统计学的几个基本概念, 目的是在一开始就给读者留下一些初步的印象, 自始至终都要体会到统计学就是“数据的科学”, 它的目的就是“探索数据内在的数量规律性”。人们研究问题的数据有些是原始数据, 有些只能得到经过加工的间接数据。对不同类型的数据, 加工整理的方法不同。

■ 课后习题

1. 举出你所知道的统计应用的例子。
2. 解释样本和总体的区别。
3. 举出一些观测数据和试验数据的例子。
4. 《福布斯》杂志每年都会公布该年度的公司首席执行官 (CEO) 收入排行榜, 主要包括以下变量: ①CEO 所在公司所属行业 (如银行、零售业等); ②上一年全年

CEO 总收入 (百万美元); ③过去 5 年中 CEO 的总收入; ④CEO 在公司所持有的股票份额 (百万); ⑤CEO 的年龄 (岁); ⑥CEO 的效率评分。问:

(1) 2005 年 500 名 CEO 的收入排行榜上的数据是样本还是总体? 为什么?

(2) 确定测量的变量的类型是定性还是定量?

5. 随机抽取一个有 500 名顾客的样本, 这些顾客被邀请试用一种新型的药膏。在这 500 人中, 有 400 人认为非常好, 32 人认为一般, 剩下的顾客没有发表意见。根据这一样本的结果, 对全体顾客对这种新型药膏的反应做出推断。

6. 指出下面的变量类型。

(1) 年龄;

(2) 性别;

(3) 汽车销量;

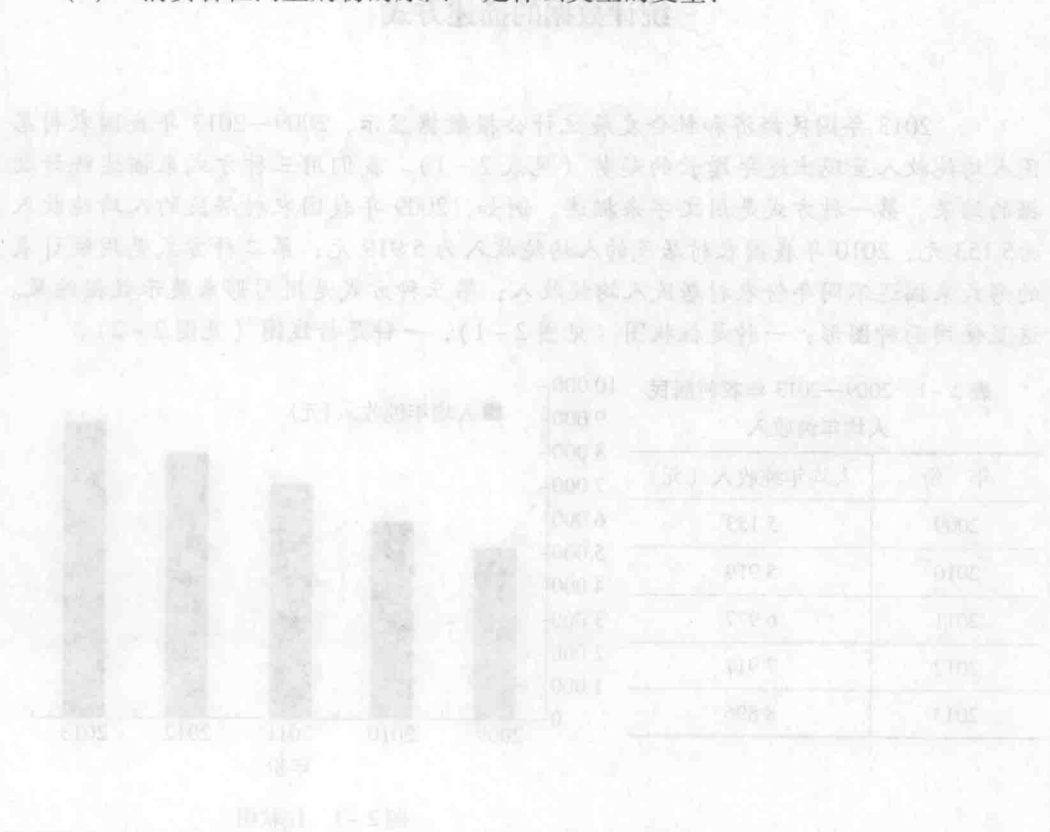
(4) 员工对企业某项制度的态度 (赞成、中立、反对);

(5) 购买商品时的支付方式 (现金、刷卡、支票)。

7. 一项调查表明, 消费者每月在网上购物的平均花费是 500 元, 他们选择在网上购物的主要原因是价格便宜。

(1) 这一研究的总体是什么?

(2) “消费者在网上购物的原因”是什么类型的变量?



第二章 数据的图表展示

学习目标

- 熟悉数据预处理的内容和目的；
- 掌握分类数据和顺序数据的整理与图表展示方法；
- 掌握数值型数据的整理与图表展示方法；
- 能熟练使用 Excel 作频数分布表和图形并合理使用图表。

案例导入

统计数据的描述方式

2013 年国民经济和社会发展统计公报数据显示, 2009—2013 年我国农村居民人均纯收入呈现出逐年增长的趋势 (见表 2-1)。我们用三种方式来描述统计数据的结果。第一种方式是用文字来描述。例如, 2009 年我国农村居民的人均纯收入为 5 153 元, 2010 年我国农村居民的人均纯收入为 5 919 元; 第二种方式是用统计表的形式来描述不同年份农村居民人均纯收入; 第三种方式是用图形来展示数据结果。这里使用两种图形: 一种是柱状图 (见图 2-1), 一种是折线图 (见图 2-2)。

表 2-1 2009—2013 年农村居民
人均年纯收入

年 份	人均年纯收入 (元)
2009	5 153
2010	5 919
2011	6 977
2012	7 917
2013	8 896

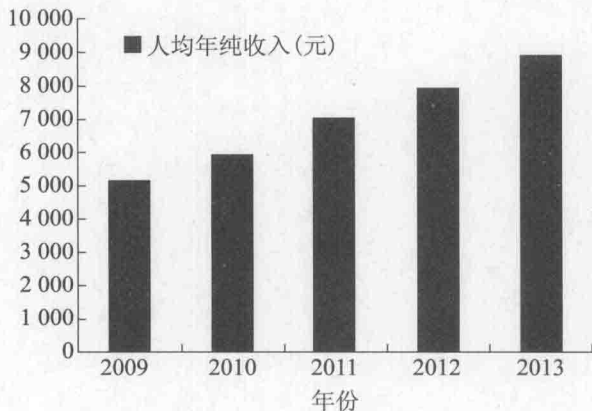


图 2-1 柱状图