



21世纪高等院校公共课精品教材



# 统计学概论

李庆东 战颂 主编

林则宏 副主编

TONGJIXUE GAILUN

**FE** 东北财经大学出版社  
Dongbei University of Finance & Economics Press



013028378



C8-43

195

21世纪高等院校公共课精品教材



# 统计学概论

李庆东 战颂 主编

林则宏 副主编

## TONGJIXUE GAILUN



北航

C1635112



东北财经大学出版社

Dongbei University of Finance & Economics Press

大连

C8-43  
195

878850817

© 李庆东 战 颂 2013

图书在版编目 (CIP) 数据

统计学概论 / 李庆东, 战颂主编. —大连: 东北财经大学出版社, 2013. 4

(21 世纪高等院校公共课精品教材)

ISBN 978-7-5654-1106-9

I. 统… II. ①李… ②战… III. 统计学-高等学校-教材 IV. C8

中国版本图书馆 CIP 数据核字 (2013) 第 025102 号

东北财经大学出版社出版

(大连市黑石礁尖山街 217 号 邮政编码 116025)

教学支持: (0411) 84710309

营 销 部: (0411) 84710711

总 编 室: (0411) 84710523

网 址: <http://www.dufep.cn>

读者信箱: [dufep@dufe.edu.cn](mailto:dufep@dufe.edu.cn)

大连美跃彩色印刷有限公司印刷

东北财经大学出版社发行

幅面尺寸: 185mm×260mm

字数: 346 千字

印张: 15

2013 年 4 月第 1 版

2013 年 4 月第 1 次印刷

责任编辑: 田玉海 尹惠 刘佳 赵晓冬

责任校对: 贺 欣

封面设计: 冀贵收

版式设计: 钟福建

ISBN 978-7-5654-1106-9

定价: 30.00 元

# 前 言

统计学是一门关于数据收集、处理、分析和解释的理论与方法的科学。随着社会的发展,统计学的应用领域也变得越来越广泛。无论是在经济管理领域,还是在军事、医学、生物、物理、化学等科学领域,只要有数据的地方就都需要用到统计学的分析方法。为了更好地适应知识经济时代的发展,就要掌握统计学的有关知识,使当代大学生能满足各主要应用领域对统计人才的需求。

“统计学”是大学本科教育经济管理类专业的核心基础课程,本书的编写是根据最新的统计学教学大纲的要求,在总结多年教学经验并参阅大量的国内外相关资料的基础上完成的。参与本教材编写的全体成员,长期致力于统计学的教学与研究工作,总结了一些对经济管理类学生进行统计基础教育的比较成功的经验。

我们编写这本教材的主要思想是,既要遵循统计研究基本规律,保留统计学的主要研究内容,又要做到简洁明了,以使学生易于学习和掌握。本书具体内容如下:第1章绪论,介绍统计学的含义、应用领域、数据类型和一些基本概念;第2章统计调查,介绍统计数据的来源、统计调查的组织方式、方法、方案设计等;第3章统计数据的整理与图示,介绍统计数据的预处理、整理、显示方法和统计表;第4章数据的概括性度量,介绍统计数据分布特征的3个方面的测度方法,包括集中趋势、离散程度和偏态与峰态;第5章抽样分布与参数估计,介绍常用的抽样方法、样本统计量的抽样分布、参数估计的基本原理以及总体参数的区间估计;第6章假设检验,介绍假设检验的基本思想,总体均值、总体比例和总体方差的假设检验过程;第7章相关分析与回归分析,介绍相关分析的基本问题、一元线性回归分析方法等;第8章时间序列分析与预测,介绍时间序列基本问题、平稳时间序列预测、趋势序列的预测等;第9章指数,介绍指数的基本概念、综合指数的编制方法、平均指数编制方法、指数体系等。

全书编写的具体分工如下:第1章和第6章由林则宏编写;第2章、第3章和第7章由战颂编写;第4章和第5章由李庆东编写;第8章和第9章由丁洪福编写。本书由李庆东、战颂担任主编,林则宏担任副主编,负责全书的设计、审定、修改、总纂和定稿工作。

本书在编写出版过程中,得到了东北财经大学出版社的大力支持和帮助,在编写过程中,参阅、引用了许多统计学方面的著作、教材,在此对有关作者一并表示感谢!

由于编者的水平有限,书中难免有疏漏和错误之处,敬请专家、同行和读者批评指正,以便进一步修改和完善。

编者

2013年元月

# 目 录

|                |    |
|----------------|----|
| 第1章 绪 论        | 1  |
| 学习指南           | 1  |
| 走近统计学          | 1  |
| 1.1 统计与统计学     | 2  |
| 1.2 统计数据类型     | 6  |
| 1.3 统计学的基本概念   | 8  |
| 本章小结           | 10 |
| 综合训练           | 11 |
| 第2章 统计调查       | 14 |
| 学习指南           | 14 |
| 走近统计学          | 14 |
| 2.1 统计数据的来源    | 15 |
| 2.2 统计调查的组织方式  | 17 |
| 2.3 统计调查的方法    | 21 |
| 2.4 调查方案设计     | 24 |
| 2.5 调查问卷设计     | 27 |
| 2.6 数据的质量      | 30 |
| 本章小结           | 31 |
| 综合训练           | 32 |
| 第3章 统计数据的整理与图示 | 34 |
| 学习指南           | 34 |
| 走近统计学          | 34 |
| 3.1 统计数据预处理    | 36 |
| 3.2 统计数据的整理    | 37 |
| 3.3 统计数据的图示    | 42 |
| 3.4 统计表        | 50 |
| 本章小结           | 55 |
| 综合训练           | 56 |
| 第4章 数据的概括性度量   | 60 |
| 学习指南           | 60 |

|                            |     |
|----------------------------|-----|
| 走近统计学 .....                | 60  |
| 4.1 集中趋势的度量 .....          | 61  |
| 4.2 离散程度的度量 .....          | 73  |
| 4.3 偏态与峰态的度量 .....         | 80  |
| 4.4 EXCEL 描述统计功能 .....     | 83  |
| 本章小结 .....                 | 85  |
| 综合训练 .....                 | 86  |
| <b>第5章 抽样分布与参数估计</b> ..... | 92  |
| 学习指南 .....                 | 92  |
| 走近统计学 .....                | 92  |
| 5.1 常用的抽样方法 .....          | 93  |
| 5.2 抽样分布的基本概念 .....        | 95  |
| 5.3 样本统计量的抽样分布 .....       | 98  |
| 5.4 参数估计的基本原理 .....        | 103 |
| 5.5 一个总体参数的区间估计 .....      | 106 |
| 5.6 两个总体参数的区间估计 .....      | 111 |
| 5.7 样本容量的确定 .....          | 114 |
| 本章小结 .....                 | 117 |
| 综合训练 .....                 | 118 |
| <b>第6章 假设检验</b> .....      | 124 |
| 学习指南 .....                 | 124 |
| 走近统计学 .....                | 124 |
| 6.1 假设检验概述 .....           | 125 |
| 6.2 总体均值的假设检验 .....        | 128 |
| 6.3 总体比例的假设检验 .....        | 133 |
| 6.4 总体方差的假设检验 .....        | 135 |
| 本章小结 .....                 | 136 |
| 综合训练 .....                 | 136 |
| <b>第7章 相关分析与回归分析</b> ..... | 142 |
| 学习指南 .....                 | 142 |
| 走近统计学 .....                | 142 |
| 7.1 相关分析的基本问题 .....        | 143 |
| 7.2 相关关系的描述与测度 .....       | 146 |
| 7.3 一元线性回归分析 .....         | 149 |
| 7.4 利用回归方程进行估计与预测 .....    | 157 |
| 本章小结 .....                 | 161 |
| 综合训练 .....                 | 161 |
| <b>第8章 时间序列分析与预测</b> ..... | 166 |
| 学习指南 .....                 | 166 |

|                |     |
|----------------|-----|
| 走近统计学          | 166 |
| 8.1 时间序列分析概述   | 167 |
| 8.2 时间序列描述性分析  | 170 |
| 8.3 平稳时间序列预测   | 179 |
| 8.4 趋势型时间序列预测  | 183 |
| 8.5 复合型时间序列预测  | 191 |
| 本章小结           | 195 |
| 综合训练           | 195 |
| 第9章 指 数        | 199 |
| 学习指南           | 199 |
| 走近统计学          | 199 |
| 9.1 统计指数的概念及分类 | 199 |
| 9.2 综合指数编制     | 202 |
| 9.3 平均指数编制     | 205 |
| 9.4 指数体系和因素分析  | 209 |
| 9.5 统计指数应用     | 215 |
| 本章小结           | 220 |
| 综合训练           | 221 |
| 附 录            | 226 |
| 参考文献           | 231 |

# 第1章 绪 论

## ◆学习指南

### 1. 学习目标

- 了解统计学的产生与发展
- 掌握统计学的含义及分科
- 理解统计学的研究方法
- 熟练掌握统计数据的各种类型、特征
- 熟练掌握统计学中常用的几对基本概念

### 2. 关键概念

- 统计学 (statistics)
- 描述统计 (descriptive statistics)
- 推断统计 (inferential statistics)
- 分类数据 (categorical data)
- 顺序数据 (rank data)
- 数值型数据 (metric data)
- 总体 (population)
- 样本 (sample)
- 变量 (variable)

### 【走近统计学】

一提到“统计”一词，很多人的第一反应是脑海里浮现出一大堆枯燥的数字，或者一些表格和图形。其实，这些仅是统计学的一小部分内容，统计工作的内容远远不止这些。随着社会信息技术的发展与进步，统计学被应用到许多领域，变得越来越重要。

我国著名的经济学家马寅初指出：“学者不能离开统计而治学，政治家不能离开统计而施政，事业家不能离开统计而执业。”

美国著名科幻作家威尔斯说：“对于追求效率的公民而言，统计思维总有一天会和读写能力一样重要。”

美国总统的年薪已经达到 40 万美元，在各国元首中名列前茅，但根据美国《工作等级年鉴》一书的排名，总统一职并未进入最好的职业之列。在美国，最好的职业是统计学家。

在日常生活中,我们也会经常接触到各种统计数据,例如下面是统计研究得出的一些结论:吸烟的男性减少寿命 2 250 天;身材高的父亲,其子女的身高也较高;学生们在听了莫扎特钢琴曲 10 分钟后进行的推理测试会比他们听 10 分钟娱乐音频或其他曲目做得更好;怕老婆的丈夫得心脏病的几率较大,等等。

由以上信息可以看出,统计已经渗透到社会经济活动和科学研究的方方面面,统计无处不在,并且正在发挥着越来越重要的作用。统计已经成为现代人类事业不可或缺的一个组成部分,与我们的生活息息相关。

在终极的分析中,一切知识都是历史。

在抽象的意义下,一切科学都是数学。

在理性的基础上,所有的判断都是统计学。

——C. R. 劳

## 1.1 统计与统计学

随着人类社会信息化进程的加快,对于各种社会经济活动如何进行量化的测度,已经成为社会发展科学分析与决策的重要基础。在对自然科学进行深入的的研究中,也越来越依赖各种科学的量化分析方法,获得最前沿的科学成就,如生物技术、基因研究等。而这些都需要统计科学的支持。统计学已经成为认识社会发展规律、经济规律以及认识各门科学的重要手段和工具。

### 1.1.1 统计的含义

要了解什么是统计学,首先要了解什么是统计。

“统计”一词由来已久,最早源自中世纪拉丁语“status”,意思是指各种现象的状态和状况。由这一词根组成的意大利语“stato”,意思为国家,作为各国的国家结构和国情知识的总称。“统计”最早作为学名使用是在 1749 年,德国哥丁根大学政治学教授阿亨瓦尔(G. Achenwall)将课程“国势学”定为“statistik”(统计)。此后,各国相继沿用“统计”一词,并将其译为各国文字,如英国将其译为“statistics”。至此,“统计”一词成为记述国家和社会发展状况数量关系的总称。

统计随着人类社会的发展和国家管理的需要而不断发展完善。统计一词得到了广泛的运用,人们对于统计这一术语的理解主要有 3 种:统计工作、统计资料和统计学。统计工作就是收集、整理和分析客观现象总体数量方面的活动过程。如企业每年要“统计”产量和产值,这里就将其作为一项工作来看待。统计资料就是统计实践活动所取得的各种信息,其中主要是反映统计对象总体数量特征的数字资料。如了解股票的交易状况要看有关成交额和股票指数的统计,这时又将其作为资料来运用。统计学是一门收集、整理与显示、分析、解释统计数据,并从数据中得出结论的科学,它是一门关于数据的科学。如学生们所说的我们正在学习统计,这里所说的统计则是指一门科学,即统计学。统计工作、统计资料和统计学三者之间既有区别,又密切联系。三者都是以统计工作为基础和核心,统计资料是统计工作的成果,统计学是统计工作的经验总结和理论概括;反过来,统计学

又是指导统计工作的原理、原则和方法,能推动统计工作水平和质量的不断提高,使统计资料更加准确、及时、全面。

### 1.1.2 统计学的产生与发展

统计理论来源于统计实践,是统计实践活动的经验总结和理论概括。统计学随着统计的产生而产生,统计的起源很早,但统计学却只有300多年的历史。从统计学发展的过程来看,可以将其划分为古典、近代和现代3个时期。

#### 1. 古典统计学时期

古典统计学时期是指17世纪中叶至18世纪中叶的统计学萌芽时期,其代表学派有国势学派和政治算术学派。

国势学派亦称记述学派,产生于17世纪的德国,主要代表人物是康令(H. Goning, 1606—1681)和阿亨瓦尔(G. Achenwall, 1791—1772)。国势学派认为统计学是关于国家显著事项的学问,主要通过国家对国家组织、人口、军队、领土、居民职业和资源财产等事项的记述,从而对国情国力进行研究,由于该学派在进行国势比较分析中偏重事物性质的解释而不注重数量分析,因此尽管它首先提出了“统计学”之名,但无统计学之实。

政治算术学派产生于17世纪的英国,主要代表人物是威廉·配第(W. Petty, 1623—1687)和约翰·格朗特(J. Graunt, 1620—1674)。威廉·配第的代表作是《政治算术》,该书用数量分析的方法对比了英国、法国、荷兰3国的财富和力量,以此批驳了当时英国国内的悲观论调。马克思对其有极高的评价,称其为“政治经济学之父,在某种程度上也可以说是统计学的创始人”。约翰·格朗特的代表作为《关于死亡表的自然和政治观察》,书中运用大量观察的方法,对伦敦人口的出生率、死亡率、性别比例和人口发展趋势,做了分类计算和预测,证明没有悲观的必要;并且第一次编制了初具革命性的“生命表”,对年龄、性别、死亡率及人口寿命作了具体分析,他被后人称为“人口统计之父”。由于政治算术学派运用大量观察法、分类分析法和对比分析法来综合研究社会经济问题,因此被认为虽无统计学之名,但有统计学之实的学派。

#### 2. 近代统计学时期

从18世纪末到19世纪末,是近代统计学时期,这一时期的一个重大贡献是大数定律和概率论被引入统计学。之后,最小平方法、误差理论和正态分布理论等相继成为统计学的重要内容。其代表学派有数理统计学派和社会统计学派。

数理统计学派产生于19世纪中叶,代表人物为比利时的凯特勒(A. Quetelet, 1796—1874),代表作有《关于概率论》、《社会物理学》,他主张用研究自然科学的方法来研究社会现象,正式把概率论引入统计学,并最先用大数定律论证了社会生活中随机现象的规律性,还提出了误差理论和“平均人”思想。凯特勒使统计学的发展进入了一个新的阶段,因此,凯特勒被称为“统计学之父”。

社会统计学派始于19世纪末,主要代表人物是恩格尔(C. L. E. Engel, 1821—1896)和梅尔(C. V. Magr, 1841—1925)。恩格尔的主要贡献有:在人口调查中提倡用个人调查卡片法、在消费计量上首创消费权数、发明恩格尔法则和恩格尔系数、建设社会统计学体系;梅尔是第一个对社会现象表现出来的规律性形式及其性质进行系统阐述的统计学家。他认为统计学是社会科学中的一门独立科学,是研究规律性的实质性科学。

### 3. 现代统计学时期

从19世纪末到现在,是现代统计学时期。这一时期的显著特点是数理统计学由于同自然科学、工程技术科学紧密结合及被广泛应用于各个领域而获得迅速发展,各种新的统计理论与方法,尤其是推断统计理论与方法得以大量涌现,如英国统计学家卡尔·皮尔逊(K. Pearson, 1857—1936)的卡方分布理论,戈赛特(W. S. Gosset, 1876—1937)的小样本t分布理论,费希尔(R. A. Fisher, 1890—1962)的F分布理论和实验设计方法,波兰统计学家尼曼(J. Neyman, 1894—1981)和英国统计学家皮尔逊(E. S. Pearson, 1895—1980)的置信区间估计理论和假设检验理论,以及非参数统计法、序贯抽样法、多元统计分析法、时间数列跟踪预测法等都应运而生,并逐步成为现代统计学的主要内容。现代统计学时期是统计学发展最辉煌的时期。

#### 1.1.3 统计学的分科

统计学的内容十分丰富,研究与应用的领域非常广泛。从统计方法的构成来看,统计学可以分为描述统计学和推断统计学;从统计方法的研究和应用角度来看,统计学可以分为理论统计学和应用统计学。

##### 1. 描述统计学和推断统计学

描述统计学(descriptive statistics)是研究如何取得反映客观现象的数据,并通过图表形式对所收集到的数据进行整理与显示,进而通过综合概括与分析得出反映客观现象的规律性数量特征。描述统计学的内容包括统计数据的搜集方法、加工处理方法、显示方法、分布特征的概括与分析方法等。

推断统计学(inferential statistics)研究如何根据样本数据去推断总体数量特征的方法,它是在对样本数据进行描述的基础上,对统计总体的未知数量特征做出以概率形式表述的推断。推断统计学的内容包括参数估计和假设检验两大类。参数估计是利用样本信息推断样本总体特征,假设检验则是利用样本信息判断对总体的某个假设是否成立。例如,美国盖洛普(Gallup)调查公司在美国总统大选前通常会从全美国的选民中随机抽取1500人左右,对大选结果进行调查与预测,并会给出不超过3%的预测误差。这就是利用样本信息和概率论原理进行统计推断的过程。

描述统计和推断统计是统计学的两个组成部分。描述统计是统计学的基础,推断统计则是现代统计学的核心。没有描述统计提供有效的样本信息,推断统计方法就只能成为无的放矢,没有推断统计学,就不可能使统计学方法论更为完善,也不可能使统计学建立在更为科学的基础之上。

##### 2. 理论统计学和应用统计学

理论统计学(theoretical statistics)把研究对象一般化、抽象化,以数学中的概率论为基础,从纯理论的角度,对统计方法加以推导论证,其中心内容是以归纳方法研究随机变量的一般规律,例如统计分布理论、统计估计与假设检验理论、相关与回归分析、方差分析、时间序列分析、随机过程理论等。

应用统计学(applied statistics)是研究如何应用统计理论和方法去解决实际问题的统计学。由于在各种科学研究中都需要通过数据分析来解决问题,因而统计学的应用几乎扩展到了所有的科学研究领域,从而形成了一个庞大的应用统计科学体系。例如,统计方法在经济和社会科学研究领域的应用形成了经济统计学、管理统计学、社会统计学、人

口统计学等；在农业试验方面的应用形成了农业统计学；在医学中的应用形成了医疗卫生统计学；在生物学中的应用形成了生物统计学等等。

理论统计学和应用统计学相互促进、共同提高。理论统计学的研究为应用统计的数量分析提供方法论基础，大大提高了统计分析的认识能力；而应用统计学在对统计方法的实际应用中，又常常会对理论统计学提出新的问题，开拓理论统计学的研究领域。

#### 1.1.4 统计学的基本方法

统计研究的目的和研究对象的性质决定着统计研究的方法，正确的研究方法又是完成统计工作的重要条件。统计研究方法问题在统计学中居于重要地位，没有一整套科学的统计方法便不可能准确、及时、全面、系统地掌握现象总体的数量特征，更不能对现象总体有更深刻的认识。统计数据研究的基本方法有大量观察法、统计分组法、综合指标法、归纳推断法、统计模型法。

##### 1. 大量观察法

大量观察法是统计学特有的研究方法，也是统计调查阶段的基本方法。所谓大量观察法是指在统计研究的过程中，对所研究的社会经济现象总体的全部或足够多的单位进行调查，借以掌握现象的综合特征，认识社会经济现象发展规律的一种方法。统计调查中的许多方法，如统计报表、普查、抽样调查、重点调查等，都是通过对研究对象的大量观察来认识现象总体的现状和发展情况的。

统计研究之所以要运用大量观察法是由研究现象的大量性和复杂性决定的。在现象总体中，个别单位往往受偶然因素的影响，如果任选其中之一进行观察，其结果不足以代表总体的一般特征，只有观察全部或足够多的单位并加以综合，影响个别单位的偶然因素才会相互抵消，现象的一般特征才会显示出来。大量观察法的意义在于可使个体与总体之间在数量上的误差相互抵消。例如，要了解辽宁省“十一五”期间家庭居民的消费水平由于收入、消费观念及消费习惯等差别，导致各个家庭各不相同。有的家庭的人均消费支出偏高，消费结构以高档消费品为主，有些家庭则恰恰相反。如果不进行大量观察，调查推断的结果就可能出现偏差，也许偏差还会很大。这就不能正确认识辽宁省居民实际的生活消费水平，由此可见大量观察法的意义。

##### 2. 统计分组法

统计分组法是在一定的研究目的下，将总体各单位按照一定的标准，分成若干组成部分，发现所研究现象规律的一种方法。例如，将全国人口按性别划分为男性人口和女性人口两组；将企业按所有制形式划分为国有、集体、个体、私营等若干组。统计分组法是统计研究的基本方法，主要用于统计整理阶段。

统计分组法是研究总体内部差异的重要方法，通过分组可以研究总体中不同类型的性质以及它们的分布情况，如产业的经济类型及其行业分布情况；也可以研究总体中的构成和比例关系，如三大产业的构成、生产要素的比例等；还可以研究总体中现象之间的相互依存关系，如企业经营规模与利润率之间的关系等。

由于将总体单位分组是在取得资料后整理资料时进行的，因此，分组容易被认为只是整理方法，这是一种误解。其实，无论从实际工作过程讲，还是从作用上讲，它是贯穿统计活动全过程的一种重要的方法。

##### 3. 综合指标法

统计研究的对象具有数量性和总体性的特点，要综合说明大量现象的数量关系，概括地

表明其一般特征,必须采用综合指标法。综合指标法是指运用各种统计指标来反映和研究客观总体现象的一般数量特征和数量关系的方法。例如,某市2010年国内生产总值为300亿元,国有企业职工的平均年工资为20 050元等,都是运用了综合指标法。大量原始资料经过分组整理汇总,得出综合指标数值。统计必须在此基础上,按照分析的要求,进一步计算各种分析指标,对现象的数量关系进行对比分析。综合指标法是统计分析的基本方法。

#### 4. 统计推断法

统计推断法是指在一定的置信程度下,根据样本资料的特征,对总体的特征作出估计和预测的方法。统计在研究现象的总体数量关系的时候,需要了解总体对象的范围往往是很大的,有时甚至是无限的,而由于经费、时间和精力等各种原因,以致有时在客观上只能从中观察部分单位或有限单位进行计算和分析,根据结果来对总体进行推断。例如,要说明一批灯泡的平均使用寿命,只能从该批灯泡中随机抽取一小部分进行检验,推断这一批灯泡的平均使用寿命,并给出这种推断的置信程度。统计推断是现代统计学的基本方法,在统计研究中得到了极为广泛的应用,它既可以用于对总体参数的估计,也可以用作对总体的某些假设检验。除了最常见的总体参数的推断外,统计模型参数的估计和检验,根据时间序列所作的外推预测等,也都属于统计推断的范畴,都存在着误差和置信度的问题。

#### 5. 统计模型法

在统计研究中,我们经常需要对相关现象之间的数量变动关系进行定量研究,以了解某一(些)现象数量变动与另一(些)现象数量变动之间的关系及变动的影响程度。在研究这种数量变动关系时,需要根据具体的研究对象和一定的假设条件,用合适的数学方程来进行模拟,这种方法我们称之为统计模型法。如相关分析法、回归分析法和统计预测法。运用统计模型法,可以对客观现象和过程中存在的数量关系进行比较完整和全面的描述,可以使统计研究更具广度和深度,提高统计的认识能力。

上述各种方法之间是相互联系、互相配合的,共同组成了统计学方法体系。

## 1.2 统计数据的类型

当今世界属于信息爆炸的知识经济时代,人们经常从各种媒体上听到或看到各种各样的数据,统计数据不仅真实再现了经济发展的状况,而且也成为影响经济发展的一个重要因素。每到公布经济数据之际,万众瞩目,参与各种经济活动的人们都要根据这些数据来进行判断与决策。

在日常工作和生活中,随处可见统计数据。例如,教师上课之前统计班级的出勤人数,计算学生期末考试的平均成绩和名次;体育比赛时解说员统计比赛双方的得分率、失误率等等。可以说,统计数据随处可见,与人们的生活息息相关。统计数据是对现象进行测量的结果,统计数据不仅仅是数字,也可以是文字。下面从不同的角度对统计数据进行分类。

### 1.2.1 分类数据、顺序数据、数值型数据

按照所采用计量尺度的不同,将统计数据分为分类数据、顺序数据和数值型数据。

#### 1. 分类数据

分类数据(categorical data)是指只能归于某一类别的非数字型数据,表现为类别,

用文字来表述事物分类的结果。例如,人的消费按照支出去向分为衣、食、住、用、行、烧、医、文、娱、健等类别;人口按性别分为男和女两种类别。这些都是分类数据。

在分类数据中,各类别之间是平等、并列关系,无法区分优劣或大小,各类别之间的顺序可以任意改变。因此分类数据只有 $=$ 或 $\neq$ 的数学特性。在实际工作中,为了便于统计处理,常用数字代码来表示各个类别,例如分别用1、2、3、4、5、6、7、8、9、10表示衣、食、住、用、行、烧、医、文、娱、健,分别用0、1表示男性和女性。这时的数字只是代表不同类别的一个代码,没有任何程度上的大小之分,也没有进行任何数学运算。

## 2. 顺序数据

顺序数据(rank data)是指只能归于某一有序类别的非数字型数据,表现为有顺序的类别,用文字来表述。例如,产品的等级分为一等品、二等品、三等品、等外品;学生考查课的成绩分为优、良、及格、不及格;某小区居民对物业公司服务态度的满意度分为非常满意、满意、一般、不满意、非常不满意等等。这些都属于顺序数据。

在顺序数据中,虽然结果也表现为类别,但是各个类别之间是可以比较顺序的。很显然,顺序数据比分类数据更精确一些,它除了具有 $=$ 或 $\neq$ 的数学特性外,还具有 $>$ 或 $<$ 的数学特性。同样,顺序数据也可以用数字代码来表示,例如,学生考查课的成绩可以分别用4、3、2、1来分别表示优、良、及格、不及格。这时,数字代码体现一种顺序或程度的不同,但还不能体现事物之间或不同结果之间(例如及格和不及格,一等品和二等品之间)的具体数量差别。

## 3. 数值型数据

数值型数据(metric data)是指按数字尺度测量的观察值,是使用自然或度量衡单位对事物进行计量的结果,其结果表现为具体的数值。按照“0”是否有意义,又可以分为定距数据和定比数据。

### (1) 定距数据

定距数据是一种不仅能反映事物所属的类别和顺序,还能反映事物类别或顺序之间数量差距的数据,表现结果为数值。例如,两名学生的统计学考试成绩分别为95分和56分,不仅能说明第一名学生成绩优秀,第二名学生不及格,前者成绩高于后者,还可以说明两者之间具有39分的差距;再如,一个地区的温度为 $30^{\circ}\text{C}$ 与另一个地区的 $35^{\circ}\text{C}$ 相差 $5^{\circ}\text{C}$ 等等,这些都属于定距数据的类型。由于定距数据表现为数值,因此,其结果除了具有 $=$ 、 $\neq$ 、 $>$ 、 $<$ 的数学特性外,还具有 $+$ 、 $-$ 的数学特性。但是定距数据不适合进行乘除运算,例如气温 $30^{\circ}\text{C}$ 和 $10^{\circ}\text{C}$ 相比,不能说明前者的温暖程度是后者的3倍,因为气温可以是 $0^{\circ}\text{C}$ 或 $0^{\circ}\text{C}$ 以下,而 $0^{\circ}\text{C}$ 或 $0^{\circ}\text{C}$ 以下并不代表没有温度。这种情况称为不存在绝对零点的现象,类似的现象还有企业利润等等。

### (2) 定比数据

定比数据是一种不仅能体现事物之间数量差距,还能通过对比运算,即计算两个测度值之间的比值来体现相对程度的数据,表现结果为数值。只要是反映存在绝对零点的现象(即零代表没有)的数据,都是可以进行对比运算的定比数据。例如,糖果的重量50克,人的身高180cm、体重60kg,水杯的容量为500ml等等,这些都是定比数据。定比数据是包含信息量最多的数据,绝大多数统计数据都属于这一类。因此,定比数据除了具有 $=$ 、 $\neq$ 、 $>$ 、 $<$ 的数学特性外,还具有 $+$ 、 $-$ 、 $\times$ 、 $\div$ 的数学特性。

上述3种数据是由低级到高级、由粗略到精确逐步递进的。适合于低级数据的性质同样适合高级,反过来适合高级数据的并不适合低级数据。如可以将考试成绩的百分制转化为五等级,但是考查课的等级并不能转化成百分制。

分类数据和顺序数据说明的是事物的品质特征,通常是用文字来表述的,其结果均表现为类别,因而可统称为定性数据或品质数据;定距数据和定比数据说明的是现象的数量特征,通常是用数值来表现的,因此也可称为数值型数据或定量数据。我们所处理的数据大多数为数值型数据。

### 1.2.2 观测数据和实验数据

按照统计数据的收集方法,将统计数据分为观测数据和实验数据。

观测数据(observational data)是通过统计调查或观测的方式而获取的反映研究现象客观存在的数量特征的数据,这类数据是在没有人为控制的条件下获得的。有关社会经济现象的统计数据几乎都是观测数据。例如,2012年4月,全国居民消费价格总水平(CPI)同比上涨3.4%,其中,城市上涨3.4%,农村上涨3.3%;全国工业生产者出厂价格同比下降0.7%,环比上涨0.2%。

实验数据(experimental data)是在人为控制的条件下,通过实验的方式而获得的关于实验对象的数据。自然科学研究中的数据大都属于实验数据。例如,一种艾滋病(AIDS)新药的疗效实验数据、试验田亩产实验数据、新产品性能实验数据等等,都属于这类数据。

### 1.2.3 截面数据、时序数据和面板数据

按照对客观现象观察的角度不同,将其分为截面数据、时序数据和面板数据。

截面数据(cross-sectional data)是对一些同类现象在近似相同的时间上收集到的数据(空间状态不同,时间状态相同),描述的是在相同时间状态下同类现象的数量特征在不同空间状态的差异情况。例如,我国2011年各地区国内生产总值数据就是截面数据。

时序数据(time-series data)是时间序列数据的简称,是对同一现象在不同时间上收集到的数据(空间状态相同,时间状态不同),描述的是现象某一方面(或某几方面)的数量特征随时间而变化的情况。例如,1996年—2010年“十一五”期间我国国内生产总值的统计数据,就是时序数据。

面板数据(panel data)是同时在时间和截面空间上取得的二维数据。例如1996年—2005年30个省份的工业生产生产总值数据,由30个省份10年的观测值组成,共有300个数据,这些数据就是面板数据。

## 1.3 统计学的基本概念

### 1.3.1 总体和样本

#### 1. 总体

就任何一个统计问题而言,总体是最基本的要素。统计总体简称总体,是从数理统计中借用来的名称,在数理统计中又称母体。所谓总体(population),就是根据一定目的确定的所要研究的事物的全体,它是由客观存在的、具有某种共同性质的许多个别事物构成的整体。组成总体的每个个别事物称为个体,也称为总体单位。总体中个体数量的多少,

称为总体容量或总体单位总数,用 $N$ 表示。例如,在我国历次人口普查中“具有中华人民共和国国籍并在中华人民共和国境内常住的人”就构成了总体,其中每一个人就是个体;研究辽宁省工业企业的生产情况,那么辽宁省所有的工业企业就构成了总体,其中的每一家工业企业就是个体;研究辽宁石油化工大学在校学生月支出情况,那么辽宁石油化工大学的所有在校学生就构成了总体,其中的每一名学生就是个体。

总体按照总体单位数量是否有限可以分为有限总体和无限总体两类。有限总体是指总体的范围能够明确确定,即总体所包含的个体数目是有限可数的。例如,某市由民营企业构成的总体,人口普查中的统计总体等等都为有限总体。无限总体是指总体所包含的个体数目是无限的、不可数的。例如,流水生产线上大量连续生产的小件产品,因昼夜不停地生产,我们可以认为其产量是无限的,其组成的总体就是无限总体;大海里的鱼资源总体、宇宙中的星球总体、大气中的悬浮颗粒总体等等都是无限总体。区分有限总体和无限总体,有利于根据不同类型的总体采用不同的调查研究方法。

## 2. 样本

样本是与总体相对应的概念,在统计学中占有重要地位,几乎所有的统计理论和方法都是建立在样本之上的,而统计实践研究现象的数量方面也总是由样本开始的。所谓样本(sample)就是从总体中抽取的一部分个体的集合,也称为子样。样本中所包含的个体数,称为样本容量或样本单位数,用 $n$ 表示。从总体中抽取一部分个体作为样本,目的是根据样本提供的信息去推断总体的特征。例如,一家公司正在接受审计,审计人员没有必要对该公司年度内的所有55 600张发票全部审查,只需随机抽查一个100张发票的样本即可,审计人员通过这100张样本发票计算的差错率就可对全部55 600张发票的差错率进行推断。

## 1.3.2 参数和统计量

### 1. 参数

总体参数简称参数(parameter),是用来描述总体特征的概括性数字度量,是研究者想要了解的总体的某种特征值。研究者所关心的参数通常有总体平均数、总体标准差、总体比例等。在统计中,参数通常用希腊字母表示,或者用大写的英文字母表示。例如,总体平均数用 $\mu$ 表示,总体标准差用 $\sigma$ 表示,总体比例用 $\pi$ 表示等等。

在统计研究中,总体往往是不知道的,我们需要进行抽样调查,用样本的信息对总体进行进一步的推断,因此,参数往往是未知的。但是一旦总体确定后,总体的特征是唯一不变的,我们称参数为未知的常数。例如,某企业生产一批灯泡,我们并不知道这批灯泡的平均使用寿命和产品合格率,但是它们是一个常数,可以采用抽样的方式,从中随机抽取一部分灯泡作为样本,用样本灯泡的产品合格率和平均寿命来对总体进行推断。

### 2. 统计量

统计量(statistic)是用来描述样本特征的概括性数字度量,它是根据样本信息计算出来的一个量。通常研究者所关心的统计量主要有样本平均值、样本标准差、样本比例等。在统计中,样本统计量用小写的英文字母来表示。例如,样本均值用 $\bar{x}$ 表示、样本标准差用 $S$ 表示、样本比例用 $p$ 表示等等。

由于样本是从总体中抽取出来的,所以统计量是可以得到的。因为样本是按照随机原则抽取的,所以统计量是样本的函数。除了样本平均值、样本标准差、样本比例这类统计

量,还有一些是为统计分析的需要而构造出来的统计量,比如用于统计检验的 $z$ 统计量、 $t$ 统计量、 $F$ 统计量、 $\chi^2$ 统计量等等。

### 1.3.3 变量

变量是统计学中经常涉及的一个概念。我们把说明现象某种特征的概念称为变量(variable)。它的特点是从一次观察到下一次观察呈现出差别或变化。例如,性别、收入、产品的质量等级等都是变量。变量的具体表现称为变量值,如性别是“男”或“女”,“男”、“女”就是变量值。统计数据就是统计变量的具体表现。变量按其反映数据的计量尺度不同,可以分为以下3种类型。

#### 1. 分类变量

分类变量(categorical variable)是指说明事物类别的一个名称,其数值表现就是分类数据。例如,“企业类型”就是一个分类变量,变量值为“国有企业”、“民营企业”、“三资企业”等等;“购买商品的支付方式”也是一个分类变量,变量值为“现金”、“信用卡”、“支票”等等;“性别”也是分类变量,变量值为“男”、“女”。

#### 2. 顺序变量

顺序变量(rank variable)是指说明事物有序类别的一个名称,其数值表现就是顺序数据。例如,“员工对企业某项改革措施所持的态度”就是顺序变量,变量值可以为“赞成”、“中立”、“反对”等;“大学生考查课的成绩”也是一个顺序变量,变量值可以是“优”、“良”、“中”、“不及格”等;“产品等级”也是一个顺序变量,变量值可以表现为“一等品”、“二等品”、“三等品”、“等外品”等。

#### 3. 数值型变量

数值型变量(metric variable)是指说明事物数字特征的一个名称,其数值表现就是数值型数据。例如,“人的身高”、“体重”、“收入”、“产品的成本”等等,这些都属于数值型变量,他们的变量值可以表现为不同的数值。

数值型变量按其取值是否连续可以分为离散型变量和连续型变量。离散型变量(discrete variable)就是其取值可以用整数一一列举的变量。例如,企业个数、学生人数、产品数量等,这些变量值只能用整数表现,且只能取有限个值,取值都以整数位断开。连续型变量(continuous variable)就是其取值可以无限分割而不能一一列举的变量。例如,人的身高、体重、收入等,这类变量的取值在两个整数之间可有无限个数值,且取值是连续不断的。在统计实践中,为了便于计算,可将有的连续性变量按离散型变量来处理,如人的年龄、企业的产值等。

## ◆ 本章小结

本章我们首先介绍了统计与统计学的概念。“统计”一词有3种含义:统计工作、统计资料和统计学。“统计学”是一门收集、整理与显示、分析和解释统计数据,并从数据中得出结论的科学,它是一门关于数据的科学。统计学的历史大体可分为古典统计学时期、近代统计学时期和现代统计学时期,曾经产生过记述学派、政治算术学派、数理统计学派和社会经济统计学派等流派。

统计学从不同的角度有不同的分科。从统计方法的构成来看,统计学可以分为描述统计学和推断统计学;从统计方法的研究和应用角度来看,统计学可以分为理论统计学和应