



普通高等教育“十三五”规划教材



◎ 刘小平 李 忆 段 俊 编著

# 统计学

## ——理论、案例、实训



中国工信出版集团



电子工业出版社  
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY  
<http://www.phei.com.cn>



统计分析系列

# 统计学

## ——理论、案例、实训

刘小平 李 忆 段 俊 编著

電子工業出版社

Publishing House of Electronics Industry

北京 · BEIJING

## 内 容 简 介

本书是一本集理论与方法、案例与实训为一体的统计教材。全书共9章,主要介绍统计基本概念、数据收集与显示、数据描述、抽样分布、参数估计、假设检验、相关与回归、时间序列、指数。在本书编写过程中,作者广泛吸纳国内外优秀统计学教材成果,大胆融入现代教学理念,充分结合作者二十多年的统计学教学实践和教材编写经验,各章采取“案例导入—经典理论—本章小结—案例实验—思考题—练习题—实训题”的编写路径,体系完整,结构新颖。全书将统计理论、统计方法、实际案例、实验实训、综合练习融为一体,具有知识性、应用性、趣味性和易读性等特点。电子课件和习题参考答案可登录华信教育资源网 [www.hxedu.com.cn](http://www.hxedu.com.cn) 免费下载。

本书内容丰富,很少涉及统计学理论的数学证明,每章均配有与实际应用紧密联系的案例、练习和实训,运用一至两种软件对案例进行实验操作,适合作为高等院校经济、金融、管理、文学、法学等专业的本科生统计学课程教材,也是广大统计爱好者和统计工作者的参考读物。

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。

### 图书在版编目(CIP)数据

统计学:理论、案例、实训 / 刘小平, 李忆, 段俊编著. —北京: 电子工业出版社, 2017.7

(统计分析系列)

ISBN 978-7-121-31719-4

I. ①统… II. ①刘… ②李… ③段… 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(2017)第 120979 号

策划编辑: 秦淑灵

责任编辑: 徐 萍

印 刷: 三河市良远印务有限公司

装 订: 三河市良远印务有限公司

出版发行: 电子工业出版社

北京市海淀区万寿路 173 信箱 邮编: 100036

开 本: 787×1092 1/16 印张: 15 字数: 384 千字

版 次: 2017 年 7 月第 1 版

印 次: 2017 年 7 月第 1 次印刷

印 数: 2000 册 定价: 35.00 元

凡所购买电子工业出版社图书有缺损问题, 请向购买书店调换。若书店售缺, 请与本社发行部联系, 联系及邮购电话: (010) 88254888, 88258888。

质量投诉请发邮件至 [zltz@phei.com.cn](mailto:zltz@phei.com.cn), 盗版侵权举报请发邮件至 [dbqq@phei.com.cn](mailto:dbqq@phei.com.cn)。

本书咨询联系方式: (010) 88254531。

# 前 言

本书是为高等院校非统计学专业的学生编写的一本统计学教材,主要适用于经济、金融、管理、文学、法学等专业的学生。由于统计学是一门方法论性质的学科,应用性很强,尤其在数据分析方面,有着其他学科所不具备的独有特性,因此,学生对统计学掌握的好坏,对人才培养的质量有着重要影响。

本书集理论与方法、案例与实训为一体,各章采取“案例导入—经典理论—本章小结—案例实验—思考题—练习题—实训题”的编写路径,体系完整,结构新颖。全书将统计理论、统计方法、实际案例、实验实训、综合练习融为一体,内容丰富。书中很少涉及统计学理论的数学证明,每章均配有与实际应用紧密联系的案例、练习和实验,并运用一至两种软件(SPSS、R 或 Excel)对案例进行实验操作,具有知识性、应用性、趣味性和易读性等特点。为方便教师教学和学生学习,本书提供电子课件和习题参考答案,读者可登录华信教育资源网 [www.hxedu.com.cn](http://www.hxedu.com.cn) 免费下载。

本书主要包括描述统计、推断统计和经济管理中常用统计方法三个部分。第 1 章至第 3 章为描述统计部分,第 4 章至第 6 章为推断统计部分,第 7 章至第 9 章为经济管理中常用统计方法部分。本书按一学期 48 学时教授,每章后的案例实验操作及实验题的练习,教师可以根据实际教学进度和效果,适当调整内容。

本书由重庆邮电大学刘小平教授、李忆教授和重庆师范大学段俊老师联合编写。刘小平编写了第 1、2、3、8 章,李忆编写了第 5、6、7 章,段俊编写了第 4、9 章。全书由刘小平总纂定稿。本书的撰写得到了重庆邮电大学和重庆师范大学的大力支持和帮助。此外,重庆邮电大学经济管理学院硕士研究生毛立静、文瑞、田晓颖、张津铭、王青、官银平、郑慧对本书的撰写提供了协助,在此一并深表感谢!

本书的出版得到重庆邮电大学教育教学改革项目的资助(项目编号: XJG1509, 负责人: 刘小平),还得到重庆邮电大学校级教材立项资助(项目编号: JC2015-06, 负责人: 刘小平)。重庆邮电大学已在前期将统计学课程作为学校重点课程建设。

在本书付印之际,谨向所有帮助和支持本书编写和出版的朋友表示衷心的感谢!由于编者水平有限,书中难免有不妥或谬误之处,恳请各位专家和读者指出,不胜感激(作者电子邮箱: [liuxiaoping@cqupt.edu.cn](mailto:liuxiaoping@cqupt.edu.cn))。

刘小平

2017 年 5 月于重庆

# 目 录

|                    |    |
|--------------------|----|
| 第 1 章 绪论           | 1  |
| 1.1 统计的含义          | 1  |
| 1.1.1 统计与统计学       | 2  |
| 1.1.2 统计数据的内在数量规律性 | 2  |
| 1.2 统计学的基本概念       | 4  |
| 1.2.1 总体和样本        | 4  |
| 1.2.2 参数和统计量       | 4  |
| 1.3 统计软件简介         | 4  |
| 1.3.1 SPSS         | 5  |
| 1.3.2 SAS          | 5  |
| 1.3.3 R            | 5  |
| 1.3.4 Excel        | 6  |
| 本章小结               | 6  |
| 思考题                | 6  |
| 练习题                | 7  |
| 第 2 章 数据收集与显示      | 8  |
| 2.1 数据的计量          | 8  |
| 2.1.1 数据的计量尺度      | 8  |
| 2.1.2 数据的类型        | 10 |
| 2.2 数据的收集          | 11 |
| 2.2.1 数据的直接获取      | 11 |
| 2.2.2 数据的间接获取      | 12 |
| 2.2.3 数据的质量        | 13 |
| 2.3 数据的显示          | 15 |
| 2.3.1 数据的审核        | 15 |
| 2.3.2 统计分组与频数分布    | 15 |
| 2.3.3 数据的显示        | 20 |
| 本章小结               | 30 |
| 案例实验               | 31 |
| 思考题                | 39 |
| 练习题                | 39 |
| 实训题                | 42 |
| 第 3 章 数据描述         | 44 |
| 3.1 总量指标与相对指标      | 45 |

|       |                   |    |
|-------|-------------------|----|
| 3.1.1 | 总量指标              | 45 |
| 3.1.2 | 相对指标              | 45 |
| 3.2   | 集中趋势的测度           | 49 |
| 3.2.1 | 众数                | 49 |
| 3.2.2 | 中位数               | 51 |
| 3.2.3 | 分位数               | 53 |
| 3.2.4 | 均值                | 54 |
| 3.2.5 | 几何均值              | 56 |
| 3.2.6 | 众数、中位数和均值的比较      | 57 |
| 3.3   | 离散程度的测度           | 58 |
| 3.3.1 | 异众比率              | 58 |
| 3.3.2 | 极差                | 59 |
| 3.3.3 | 四分位差              | 59 |
| 3.3.4 | 方差和标准差            | 60 |
| 3.3.5 | 离散系数              | 61 |
| 3.4   | 偏态与峰度的测度          | 62 |
| 3.4.1 | 偏态                | 62 |
| 3.4.2 | 峰度                | 63 |
| 3.5   | 数据的标准化            | 64 |
| 3.5.1 | 标准化值的计算           | 64 |
| 3.5.2 | 标准化值的作用           | 64 |
| 3.5.3 | 契比雪夫定理            | 65 |
|       | 本章小结              | 66 |
|       | 案例实验              | 66 |
|       | 思考题               | 69 |
|       | 练习题               | 70 |
|       | 实训题               | 72 |
| 第4章   | 抽样分布              | 74 |
| 4.1   | 抽样方法              | 74 |
| 4.1.1 | 简单随机抽样            | 74 |
| 4.1.2 | 分层抽样              | 74 |
| 4.1.3 | 机械抽样              | 75 |
| 4.1.4 | 整群抽样              | 75 |
| 4.2   | 抽样分布              | 75 |
| 4.2.1 | $\chi^2$ 分布       | 75 |
| 4.2.2 | $t$ 分布            | 76 |
| 4.2.3 | $F$ 分布            | 77 |
| 4.2.4 | 正态总体的样本均值和样本方差的分布 | 78 |
| 4.2.5 | 样本比率的分布           | 81 |

|                     |            |
|---------------------|------------|
| 4.2.6 两个样本统计量的分布    | 81         |
| 本章小结                | 82         |
| 案例实验                | 82         |
| 思考题                 | 84         |
| 练习题                 | 85         |
| 实训题                 | 85         |
| <b>第5章 参数估计</b>     | <b>86</b>  |
| 5.1 参数估计的基本原理       | 86         |
| 5.1.1 估计量与估计值       | 86         |
| 5.1.2 参数估计的基本方法     | 86         |
| 5.1.3 点估计评价标准       | 87         |
| 5.2 一个总体参数的区间估计     | 88         |
| 5.2.1 总体均值的区间估计     | 88         |
| 5.2.2 总体比率的区间估计     | 91         |
| 5.2.3 总体方差的区间估计     | 92         |
| 5.3 两个总体参数的区间估计     | 93         |
| 5.3.1 两个总体均值差的区间估计  | 93         |
| 5.3.2 两个总体比率差的区间估计  | 96         |
| 5.3.3 两个总体方差比的区间估计  | 97         |
| 5.4 样本量的确定          | 98         |
| 5.4.1 确定样本量的一般问题    | 98         |
| 5.4.2 估计总体均值时样本量的确定 | 99         |
| 5.4.3 估计总体比率时样本量的确定 | 100        |
| 本章小结                | 101        |
| 案例实验                | 101        |
| 思考题                 | 104        |
| 练习题                 | 104        |
| 实训题                 | 108        |
| <b>第6章 假设检验</b>     | <b>109</b> |
| 6.1 假设检验的基本原理       | 109        |
| 6.1.1 假设检验的基本思想     | 109        |
| 6.1.2 假设检验的基本概念     | 110        |
| 6.1.3 假设检验的基本步骤     | 112        |
| 6.1.4 关于 $p$ 值      | 113        |
| 6.2 一个总体参数的假设检验     | 114        |
| 6.2.1 总体均值的假设检验     | 114        |
| 6.2.2 总体比率的假设检验     | 117        |
| 6.2.3 总体方差的假设检验     | 118        |
| 6.3 两个总体参数的假设检验     | 120        |

|                  |                 |     |
|------------------|-----------------|-----|
| 6.3.1            | 两个总体均值差的假设检验    | 121 |
| 6.3.2            | 两个总体比率差的假设检验    | 126 |
| 6.3.3            | 两个总体方差比的假设检验    | 127 |
| 本章小结             |                 | 129 |
| 案例实验             |                 | 129 |
| 思考题              |                 | 133 |
| 练习题              |                 | 133 |
| 实训题              |                 | 136 |
| <b>第7章 相关与回归</b> |                 | 137 |
| 7.1              | 相关分析            | 137 |
| 7.1.1            | 变量间的关系          | 137 |
| 7.1.2            | 相关关系的类型         | 138 |
| 7.1.3            | 相关系数            | 140 |
| 7.2              | 一元线性回归          | 142 |
| 7.2.1            | 回归分析的基本概念       | 142 |
| 7.2.2            | 一元线性回归模型的设定     | 143 |
| 7.2.3            | 参数的最小二乘估计       | 144 |
| 7.2.4            | 一元线性回归分析中的显著性检验 | 146 |
| 7.2.5            | 一元线性回归方程预测      | 150 |
| 7.3              | 多元线性回归          | 152 |
| 7.3.1            | 多元线性回归模型的设定     | 152 |
| 7.3.2            | 参数的最小二乘估计       | 152 |
| 7.3.3            | 多元线性回归分析中的检验问题  | 155 |
| 本章小结             |                 | 158 |
| 案例实验             |                 | 158 |
| 思考题              |                 | 164 |
| 练习题              |                 | 164 |
| 实训题              |                 | 167 |
| <b>第8章 时间序列</b>  |                 | 168 |
| 8.1              | 时间序列概述          | 169 |
| 8.1.1            | 绝对数时间序列         | 170 |
| 8.1.2            | 相对数时间序列         | 170 |
| 8.1.3            | 平均数时间序列         | 170 |
| 8.2              | 时间序列描述性分析       | 170 |
| 8.2.1            | 时间序列的水平描述       | 170 |
| 8.2.2            | 时间序列的动态描述       | 173 |
| 8.3              | 时间序列的分解与模型      | 176 |
| 8.3.1            | 时间序列的分解         | 176 |
| 8.3.2            | 时间序列组合模型        | 177 |

|       |                  |     |
|-------|------------------|-----|
| 8.4   | 时间序列趋势分析         | 178 |
| 8.4.1 | 移动平均法            | 178 |
| 8.4.2 | 趋势模型法            | 180 |
| 8.5   | 时间序列季节变动分析       | 184 |
| 8.5.1 | 按季(或月)平均法        | 185 |
| 8.5.2 | 移动平均趋势剔除法        | 186 |
| 8.5.3 | 季节变动的调整          | 187 |
| 8.6   | 时间序列循环波动与不规则变动分析 | 188 |
| 8.6.1 | 循环波动分析           | 188 |
| 8.6.2 | 不规则变动分析          | 191 |
|       | 本章小结             | 191 |
|       | 案例实验             | 192 |
|       | 思考题              | 201 |
|       | 练习题              | 201 |
|       | 实训题              | 203 |
| 第9章   | 指数               | 204 |
| 9.1   | 指数概述             | 204 |
| 9.1.1 | 指数概念             | 204 |
| 9.1.2 | 指数分类             | 205 |
| 9.1.3 | 个体指数编制           | 205 |
| 9.2   | 总指数编制            | 206 |
| 9.2.1 | 综合指数法            | 206 |
| 9.2.2 | 平均数指数法           | 208 |
| 9.3   | 指数体系             | 209 |
| 9.3.1 | 基本概念             | 209 |
| 9.3.2 | 指数体系分析           | 210 |
| 9.4   | 指数的应用            | 213 |
| 9.4.1 | 工业生产指数           | 214 |
| 9.4.2 | 零售价格指数           | 214 |
| 9.4.3 | 消费价格指数           | 216 |
| 9.4.4 | 股票价格指数           | 217 |
|       | 本章小结             | 218 |
|       | 思考题              | 218 |
|       | 练习题              | 218 |
|       | 实训题              | 221 |
| 附录A   | 常用统计表            | 222 |
|       | 参考文献             | 232 |

# 第1章 绪 论

## 案例导入:

### 毕业生起薪你给多少?

据国内某招聘网站 2016 年 8 月发布的《2016 年应届毕业生就业力调研报告》显示, 2016 年应届毕业生期望月薪平均值为 4985 元, 而实际签约月薪为 4765 元, 近三成毕业生进入互联网行业。而且, IT/互联网/通信/电子行业的平均月薪最高, 应届生平均薪资水平为 5693 元。与 2015 年相比, 2016 年选择创业的应届毕业生比例明显下降, 由 6.3% 降至 3.1%。同时, 2016 年应届毕业生选择就业的比例有所上升。

#### 1. 从行业看

毕业生期望就业的三大行业依次为 IT/互联网/通信/电子、金融/银行/投资/基金/证券/保险、政府/公共事业/非营利机构。而从实际就业行业来看, 排名前三的是: IT/互联网/通信/电子行业占 29.5%; 加工/生产/制造/汽车行业占 16.1%; 金融/银行/投资/基金/证券/保险行业占 13.7%。

#### 2. 从类型看

排名靠前的学校仍是以理工类和财经类为主, 前十名中, 清华大学、上海大学、同济大学都是典型的理工类大学; 而上海财经大学、上海外国语大学、对外经济贸易大学和中央财经大学都是典型的财经类大学。

#### 3. 从薪资看

应届生实际签约薪资水平较高行业: IT/互联网/通信/电子行业, 5693 元; 金融行业, 4685 元; 文化/体育/娱乐/传媒, 4552 元。

应届生实际签约薪资水平较低行业: 农/林/牧/渔/其他, 3500 元; 服务业, 3873 元; 文教体育/工艺美术, 3900 元。

应届生实际签约薪资水平较高专业: 法学, 6060 元; 医学, 5722 元; 工学, 5222 元。

应届生实际签约薪资水平较低专业: 农学, 3425 元; 艺术学, 3395 元。

## 1.1 统计的含义

案例导入中报告了国内某招聘网站对中国 2016 年应届毕业生就业力的调研结果, 其中与数据有关的事实很多。在当今全球化经济环境中, 大量的数据处处存在, 它们都是可以利用的。案例中的数据就被用来反映 2016 年应届毕业生起薪情况。在当今的管理界, 最成功的管理者是那些能够理解数据并能够有效利用数据的人。我们几乎无法想象, 案例导入中如果不存在数据将会怎样, 那将会是一份价值渺小的报告。

数据无处不在, 又是如此有价值, 研究数据的科学必将产生, 并将具有极其广泛的应用, 实用性极其强大。统计学就是一门研究数据的科学。

### 1.1.1 统计与统计学

“统计”在现今人们的工作、生活中绝不是一个陌生的词。人们经常说“这是上月的产量统计”、“我国去年的GDP值统计出来了吗？”“你学过统计吗？”从中可以看出，“统计”一词使用频繁但含义却不单一，它被人们赋予了不同的含义。因此，很难给“统计”下一个简单的定义。概括起来，“统计”有三个方面的含义。一是指统计工作，统计工作是收集、整理和分析统计数据的活动和过程；二是指统计数据，统计数据是收集、整理和分析统计数据活动的成果；三是指统计学，统计学是进行统计数据收集、整理和分析的方法和科学。这三个方面的含义各不相同，但互相之间却是有联系的。统计学是指导统计工作的方法和理论，统计数据是统计工作的成果，而统计工作对统计学的发展不断提出需求。可见，统计实际上是统计工作、统计数据和统计学三者相互区别又相互联系的有机统一体。

统计学是一门收集、整理和分析大量统计数据的方法科学，其目的是探索数据的内在数量规律性，从而科学地认识客观事物。

统计数据的收集是指获取统计数据的过程，统计学研究获取统计数据的来源以及获取的方式和方法。

统计数据的整理是指对原始统计数据的处理、分组和显示，统计学研究不同类别的统计数据的不同整理方法。

统计数据的分析是统计学的核心内容，统计学研究探索数据的内在数量规律性的各种方法。

既然统计学是研究统计数据的收集、整理和分析的科学，可见，统计学与统计数据的关系密不可分。

离开统计数据，统计学将失去研究对象，也就无须存在。另一方面，对于统计数据来说，如果没有统计学的科学方法进行整理和分析，从而获取有用的信息，统计数据只能是一堆枯燥的数字而已，将毫无价值。

### 1.1.2 统计数据的内在数量规律性

如前所述，统计学的研究目的是探索数据的内在数量规律性。那么统计数据的内在数量规律性是什么？统计数据的内在数量规律性是否存在？为什么统计学的方法能够通过分析统计数据来找到统计数据的内在数量规律性呢？下面通过几个简单的例子来看看。

图 1.1 是新中国人口发展趋势图，从图的走势可以初步看出，新中国人口发展大致经历了六个阶段，分别是：第一个人口高速增长阶段（1949—1957 年）、人口低增长阶段（1958—1961 年）、第二个高速增长阶段（1962—1970 年）、人口有控制增长阶段（1971—1980 年）、第三个高速增长阶段（1981—1990 年）、人口平稳增长阶段（1991 年至今）。

图 1.2 是我国 2000 年以来各年移动电话年末用户数趋势图，从图的形状可以看出，我国从 2000—2014 年期间，移动电话年末用户数的发展趋势非常明显，几乎为直线趋势。

表 1.1 是 10 个品牌啤酒的广告费用和销售量数据表，通过表中数据可以大致看出，随着广告费用的增加，啤酒的销售量总体上也在增加。这是否说明啤酒销售量与广告费用投入之间存在某种联系？

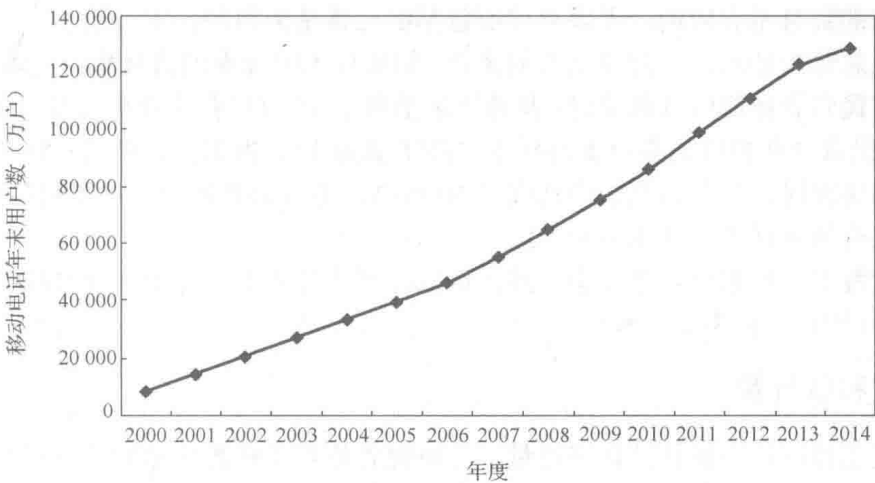
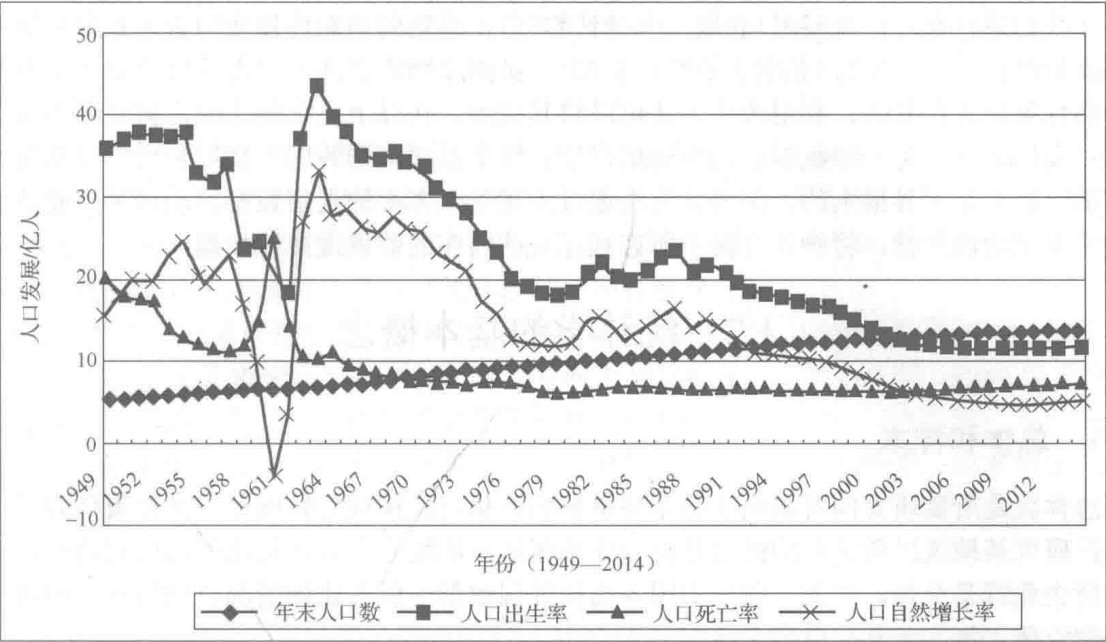


表 1.1 10 个品牌啤酒的广告费用和销售量数据表

| 啤酒品牌     | A     | B    | C     | D    | E   | F   | G    | H   | I   | J   |
|----------|-------|------|-------|------|-----|-----|------|-----|-----|-----|
| 广告费用(万元) | 122.0 | 69.7 | 105.1 | 78.6 | 9.7 | 1.5 | 23.5 | 1.8 | 5.8 | 1.8 |
| 销售量(万箱)  | 39.8  | 25.7 | 19.8  | 15.5 | 9.8 | 8.6 | 8.9  | 5.6 | 5.7 | 5.5 |

以上例子简单地说明了一个道理，统计数据的数量规律性是客观存在的，统计学通过研究大量统计数据，能够找到其数量规律性，从而科学地认识客观事物，人们利用它解决各种各样的问题。

统计学之所以能够通过研究大量统计数据找到统计数据的数量规律性，是因为统计数据所反映的客观事物是必然性和偶然性的对立统一体。必然性反映事物内在的本质特征，偶然性则反映事物表现上的差异。

在我们进行某一次观察时(获取一个统计数据),事物的偶然性使我们看不到数量规律,而在我们进行多次观察时(获取大量统计数据),虽然事物的偶然性使得统计数据千差万别,但必然性就隐含在其中,利用统计方法可以将其找到。在以上几个例子中,虽然新中国每一年人口发展数据、每一年我国移动电话用户数、每个品牌啤酒的广告费用和销售量都是不尽相同的,甚至是差异很大的,但统计方法通过大量的、多次的观察数据,能够尽可能地去除数据所表现的偶然性,将统计数据中所客观存在的内在的数量规律性找到。

## 1.2 统计学的基本概念

### 1.2.1 总体和样本

总体就是所要研究的对象的全部个体所构成的集合。构成总体的每一个元素称为个体。例如,研究某地区民营企业的经营状况,该地区所有的民营企业就是我们的研究总体,每一个民营企业就是个体。再如,研究中国各高校外国留学生所占比例情况,全国所有的高校就是研究总体,每一所高校就是个体。

总体根据所包含个体是否有限可以分为有限总体和无限总体。有限总体的范围能够明确确定,且元素的数目是有限的。无限总体所包括的元素是无限的,不可数。

样本是从总体中抽取的一部分元素的集合,构成样本的元素的数目称为样本容量。例如,在研究某地区民营企业的经营状况时,从该地区抽取了 15 家民营企业进行调查研究,这里所抽取的 15 家民营企业构成的集合就是样本,样本量为 15。再如,在研究中国各高校外国留学生所占比例情况时,从全国高校中抽取了 50 所高校进行调查研究,这里被抽中的 50 所高校所构成的集合就是样本,样本量为 50。

在抽样过程中,根据样本量大小不同,可以将样本量大于或等于 30 的样本称为大样本,样本量小于 30 的样本称为小样本。

### 1.2.2 参数和统计量

参数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。总体参数的种类很多,但研究者常关心的主要有:总体均值(记为  $\mu$ ),总体方差(记为  $\sigma^2$ ),总体比率(记为  $\pi$ )等。总体参数通常用希腊字母表示。

与总体参数相对应,统计量用来描述样本特征的概括性数字度量,它是根据样本数据计算出来的一些量,是样本的函数。常用的样本统计量有:样本均值(记为  $\bar{x}$ ),样本方差(记为  $s^2$ ),样本比率(记为  $p$ )等。样本统计量通常用英文字母表示。

统计学中的绝大多数问题都是研究如何根据样本统计量去推断总体参数的问题。例如,用样本均值  $\bar{x}$  去推断总体均值  $\mu$ ,用样本方差  $s^2$  去推断总体方差  $\sigma^2$ ,用样本比率  $p$  去推断总体比率  $\pi$  等。

## 1.3 统计软件简介

由于计算机的发明和广泛应用,使得统计学有了更大的发展,计算机为统计应用的普及提供了更好的条件。

目前,处理统计数据的软件很多,比较常用的有 SPSS、SAS、R、Excel 等。

### 1.3.1 SPSS

SPSS 是软件英文名称的首字母缩写,原意为 Statistical Package for the Social Sciences,即“社会科学统计软件包”,是一种集成化的计算机数据处理应用软件。但是随着 SPSS 产品服务领域的扩大和服务深度的增加,SPSS 公司已于 2000 年正式将英文全称更改为 Statistical Product and Service Solutions,意为“统计产品与服务解决方案”。

SPSS 是世界上最早的统计分析软件,由美国斯坦福大学的三位研究生于 20 世纪 60 年代末研制,同时成立了 SPSS 公司,并于 1975 年在芝加哥组建了 SPSS 总部。1984 年 SPSS 总部首先推出了世界上第一个统计分析软件微机版本 SPSS/PC+,开创了 SPSS 微机系列产品的开发方向,极大地扩充了它的应用范围,并使其能很快地应用于自然科学、技术科学、社会科学的各个领域。迄今 SPSS 软件已有 40 余年的成长历史,全球拥有大量的产品用户,广泛分布于通信、医疗、银行、证券、保险、制造、商业、市场研究、科研教育等多个领域和行业,是世界上应用最广泛的专业统计软件。

SPSS 由多个模块构成,模块数量随版本的不同而有所变化,主要模块功能有:Statistic Base、Advanced Statistics、Regression、Categories、Missing Value、Conjoint、Forecasting、Custom Tables、Complex Samples、Bootstrap、Decision Trees、Neural Network、Data Preparation、Statistic Adapter、Direct Market。

SPSS 具有功能强大、兼容性好、易用性强、扩展性高等特点,得到了广大统计分析人员的钟爱,也是非专业统计人员的首选统计软件。

### 1.3.2 SAS

SAS 全称为 Statistics Analysis System,最早由北卡罗来纳大学的两位生物统计学研究生编制,并于 1976 年成立了 SAS 软件研究所,正式推出了 SAS 软件。

SAS 是用于决策支持的大型集成信息系统,但该软件系统最早的功能仅限于统计分析,至今,统计分析功能也仍是它的重要组成部分和核心功能。SAS 广泛应用于金融、医药卫生、生产、运输、通信、政府和教育科研等领域。

SAS 是一个组合软件系统,它由多个功能模块组合而成,其基本部分是 BASE SAS 模块。在 BASE SAS 的基础上,还可以增加如下不同的模块而扩展不同的功能:SAS/STAT(统计分析模块)、SAS/GRAPH(绘图模块)、SAS/QC(质量控制模块)、SAS/ETS(经济计量学和时间序列分析模块)、SAS/OR(运筹学模块)、SAS/IML(交互式矩阵程序设计语言模块)、SAS/FSP(快速数据处理的交互式菜单系统模块)、SAS/AF(交互式全屏幕软件应用系统模块),等等。

它采用 MDI(多文档界面),用户在 PGM 视窗中输入程序,分析结果以文本的形式在 OUTPUT 视窗中输出。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。但是,这使得初学者在使用 SAS 时必须学习 SAS 语言,入门比较困难。

### 1.3.3 R

R 是 S 语言的一个分支。1980 年左右,由 AT&T 贝尔实验室开发的 S 语言是一种用来进行数据探索、统计分析和作图的解释型语言。最初 S 语言的实现版本主要是 S-PLUS。S-PLUS

是一个商业软件,它基于 S 语言,并由 MathSoft 公司的统计科学部进一步完善。后来新西兰奥克兰大学的 Robert Gentleman 和 Ross Ihaka 及其他志愿人员开发了一个 R 系统,由“R 开发核心团队”负责开发。R 可以看作贝尔实验室(AT&T Bell Laboratories)的 Rick Becker, John Chambers 和 Allan Wilks 开发的 S 语言的一种实现。

R 是一套完整的数据处理、计算和制图软件系统,是一个自由、免费、源代码开放的软件。其功能包括:数据存储和处理系统;数组运算工具(其向量、矩阵运算方面的功能尤其强大);统计分析工具;统计制图功能;编程语言——可操纵数据的输入和输出,可实现分支、循环,用户可自定义功能。

### 1.3.4 Excel

Excel 是办公自动化中非常重要的一款软件,可以方便地对数据进行排序、筛选等预处理,并对各种数据进行统计计算和分析,以丰富的图表方式显示数据及分析结果。Excel 被广泛地应用于日常统计工作中。

Excel 函数共有 11 类,分别是数据库函数、日期与时间函数、工程函数、财务函数、信息函数、逻辑函数、查询和引用函数、数学和三角函数、统计函数、文本函数以及用户自定义函数,统计函数是其中的一类。Excel 中所提的函数其实就是一些定义的公式,它们使用一些称为参数的特定数值按特定的顺序或结构进行计算,用户可以直接用它们对某个区域内的数值进行一系列运算。统计函数就是一系列的统计模型或计算公式,用户选择或输入相关的参数,就可以进行统计计算,如计算均值、方差等。

虽然 Excel 不是单独开发的统计软件,但由于其具有一定的统计功能,操作界面简单,且为广大用户所熟知,因此深受使用者欢迎。

## 本章小结

1. 统计学是一门收集、整理和分析数据的科学,其目的是探索数据内在的数量规律性。统计学和统计数据密不可分。
2. 统计是认识事物的工具,无论是工作、学习,还是日常生活中,都需要运用到统计学的知识,它广泛应用于社会的各个领域,学好统计学非常有用。
3. 统计学中涉及很多的基本概念,本章介绍了最常见的概念:总体和样本,参数和统计量。它们贯穿全书。
4. 处理统计数据的软件很多,比较常用的有 SPSS、SAS、R、Excel 等。对于不同的使用者,可以根据自己的条件,选择学习。学习统计软件的最好方式是在使用中学习,并多看帮助和说明。

## 思考题

1. 什么是统计与统计学?如何理解统计数据与统计学的关系?
2. 什么是总体与样本?它们之间有什么关系?
3. 什么是参数与统计量?它们之间有什么关系?
4. 试举出日常生活中的例子以说明统计数据及其规律性。

## 练习题

1. 一家研究机构从电商从业者中随机抽取 1000 人作为样本进行调查, 其中 70% 的人回答他们的月收入在 6000 元以上, 80% 的人回答他们日常购物会选择网购方式。

(1) 这一研究的总体是什么? 样本是什么? 样本量是多少?

(2) “月收入 6000 元”是参数还是统计量? “选择网购的占比 80%”是参数还是统计量?

(3) 指出该项研究中的参数和统计量。

2. 某大学经济管理学院为了解毕业生的就业意向, 分别从工商管理专业抽取 20 人、市场营销专业抽取 30 人、会计学专业抽取 50 人、经济类专业抽取 60 人进行调查。

(1) 这一研究的总体是什么?

(2) 研究者所关心的参数是什么?

(3) 样本量是多少?

## 第2章 数据收集与显示

案例导入:

### 庞大的中国网民队伍

中国互联网络信息中心(CNNIC)2017年1月22日发布了第三十九次《中国互联网络发展状况统计报告》。《报告》显示,截至2016年12月,我国网民规模达7.31亿,互联网普及率达到53.2%,超过全球平均水平3.1个百分点,超过亚洲平均水平7.6个百分点。

《报告》显示,我国2016年全年共计新增网民4299万人,增长率为6.2%,我国网民规模已经相当于欧洲人口总量。其中,手机网民规模达6.95亿,占比达95.1%,增长率连续3年超过10%。而台式计算机、笔记本电脑的使用率均出现下降,手机不断挤占其他个人上网设备的使用。

《报告》显示,2016年,我国手机网上支付用户规模增长迅速,达到4.69亿,年增长率为31.2%,网民手机网上支付的使用比例由57.7%提升至67.5%。手机支付向线下支付领域的快速渗透,极大丰富了支付场景,有50.3%的网民在线下实体店购物时使用手机支付结算。

《报告》指出,我国网民规模经历近10年的快速增长后,红利逐渐消失,网民规模增长率趋于稳定。2016年,中国互联网行业整体向规范化、价值化发展,同时,移动互联网推动消费模式共享化、设备智能化和场景多元化。

以上资料比较简明扼要地反映了当下中国网民用户的基本情况,问题的关键在于,上述数据是怎么得到的?众多的数据如何显示才会让我们对研究对象的数量特征与规律“一目了然”?

## 2.1 数据的计量

### 2.1.1 数据的计量尺度

统计数据是对各种客观现象的信息进行计量的结果,由于现象的性质不同,予以计量的尺度或测量的程度也是不同的。例如,有的现象只能对或只需对其属性进行分类,如人口的性别和民族、产品的质量等级、服务态度的好坏等;有些则可以或要求必须用比较精确的数字加以计量,如一定时期一定地区的经济活动总量、人口的平均寿命、企业的销售收入、学生考试的平均成绩等。根据计量学的一般分类方法,按照对事物计量的精确程度,可以将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。采用不同计量尺度可以得到不同类型的统计数据,进而需要采用不同的统计分析方法进行分析研究。

#### 1. 定类尺度

定类尺度也称类别尺度或列名尺度,是最粗略、计量层次最低的计量尺度。运用“属于  
此为试读,需要完整PDF请访问: [www.ertongbook.com](http://www.ertongbook.com)