



高等教育“十三五”规划课程改革创新教材



# S tatistics

# 统计学

刘 飞 © 主编



科学出版社

高等教育“十三五”规划课程改革创新教材

# 统 计 学

刘 飞 主 编

吴培乐 副主编

科 学 出 版 社

北 京

## 内 容 简 介

本书根据经济管理与人文社科专业的特点,系统介绍了统计学的基本统计思想和基本分析方法,既有较宽的理论基础,又突出实用性。同时,还强调时代性、思想性、前沿性,对大数据分析的思想和方法及学科发展前沿进行了介绍。

本书主要内容包括绪论、统计数据搜集、统计数据整理、数据特征描述、参数估计、假设检验、方差分析、相关与回归分析、时间序列分析、统计指数分析、统计综合评价。

本书适合作为应用技术型高等院校经济管理类和人文社科类专业硕士研究生、本科生的教材,也可作为高等职业院校相关专业的教材,还可作为统计工作者的参考书。

### 图书在版编目(CIP)数据

统计学/刘飞主编. —北京:科学出版社, 2018.8

(高等教育“十三五”规划课程改革创新教材)

ISBN 978-7-03-057328-5

I. ①统… II. ①刘… 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 (2018) 第 093269 号

责任编辑:张振华 / 责任校对:陶丽荣

责任印制:吕春珉 / 封面设计:东方人华平面设计部

科学出版社出版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencecp.com>

天津翔远印刷有限公司印刷

科学出版社发行 各地新华书店经销

\*

2018年8月第 一 版 开本:787×1092 1/16

2018年8月第一次印刷 印张:18

字数:400 000

定价:46.00元

(如有印装质量问题,我社负责调换(翔远))

销售部电话 010-62136230 编辑部电话 010-62135120-2005 (VT03)

版权所有,侵权必究

举报电话:010-64030229; 010-64034315; 13501151303

# 前 言

本书为满足我国普通高等院校向应用技术型高校转型发展的教学需要而编写。本书编写以应用为导引，以“科学、实用、新颖”为原则，注重对学生基本统计思想的训练，以及运用统计学知识实际问题能力的培养。

本书由西安邮电大学教学经验丰富、写作能力强的一线教师精心编写而成。编写过程中紧密结合统计学在理论和方法上的最新发展和最新成果，增加了大数据时代统计分析的思想和方法，保证了内容的实用性、时代性、思想性、前沿性。

本书共 11 个单元，包括绪论、统计数据搜集、统计数据整理、数据特征描述、参数估计、假设检验、方差分析、相关与回归分析、时间序列分析、统计指数分析、统计综合评价。为方便教学，书中穿插有大量的实例和丰富的二维码资源链接，读者通过手机等终端扫描二维码后可阅读相关内容。

本书由刘飞担任主编，吴培乐担任副主编，赵青、刘文萍、贺信、王红亮、尹磊参与编写。具体编写分工如下：刘飞编写单元 1，贺信编写单元 2 和单元 4，吴培乐编写单元 3 和单元 8，赵青编写单元 5 和单元 6，王红亮编写单元 7 和单元 10，刘文萍编写单元 9，尹磊编写单元 11。全书由刘飞负责统稿。

在编写本书的过程中，编者参阅了国内外同行的文献，并得到了许多同行和专家的大力支持和帮助，在此一并表示感谢。

由于编者水平有限，加之时间仓促，书中难免存在疏漏和不足之处，恳请广大读者批评指正。

编 者

2017 年 12 月

# 目 录

|                     |    |
|---------------------|----|
| 单元 1 绪论             | 1  |
| 1.1 学习应用统计学的三种境界    | 2  |
| 1.2 什么是统计学          | 3  |
| 1.2.1 统计学的定义        | 3  |
| 1.2.2 统计学的发展简史      | 3  |
| 1.2.3 统计学的应用领域      | 4  |
| 1.3 统计学的基本概念        | 5  |
| 1.3.1 总体和样本         | 5  |
| 1.3.2 统计指标和指标体系     | 5  |
| 1.3.3 变量和数据         | 6  |
| 1.4 统计学研究方法         | 7  |
| 1.4.1 大量观察          | 7  |
| 1.4.2 统计描述          | 8  |
| 1.4.3 统计推断          | 8  |
| 1.5 大数据时代统计学的发展     | 9  |
| 1.6 统计分析软件          | 10 |
| 思考与练习               | 12 |
| 单元 2 统计数据搜集         | 13 |
| 2.1 数据的来源           | 14 |
| 2.1.1 数据的直接来源       | 14 |
| 2.1.2 统计数据的间接来源     | 17 |
| 2.1.3 国内外主要统计数据查询网站 | 17 |
| 2.2 获取数据的方法         | 18 |
| 2.2.1 询问法           | 19 |
| 2.2.2 报告法           | 20 |
| 2.2.3 观察法           | 20 |
| 2.3 问卷设计            | 20 |
| 2.3.1 问卷的特点         | 21 |
| 2.3.2 问卷的内容         | 21 |
| 2.3.3 问卷调查的一般程序     | 21 |
| 2.3.4 问卷的设计         | 22 |
| 2.4 大数据时代统计调查       | 25 |
| 2.4.1 大数据对数据收集的影响   | 25 |

|                     |    |
|---------------------|----|
| 2.4.2 大数据的获取方法      | 26 |
| 2.4.3 大数据的分析与整合     | 26 |
| 思考与练习               | 27 |
| 单元3 统计数据整理          | 28 |
| 3.1 数据的预处理          | 29 |
| 3.1.1 数据审核          | 29 |
| 3.1.2 数据筛选          | 29 |
| 3.1.3 数据排序          | 30 |
| 3.1.4 数据填补          | 30 |
| 3.2 品质数据的整理         | 31 |
| 3.2.1 定类数据的整理与图示    | 31 |
| 3.2.2 定序数据的整理与图示    | 33 |
| 3.3 数值型数据的整理        | 34 |
| 3.3.1 单变量值分组        | 35 |
| 3.3.2 组距式分组         | 36 |
| 3.4 大数据的整理          | 41 |
| 3.5 频数分布            | 42 |
| 3.5.1 频数分布的概念与种类    | 42 |
| 3.5.2 频数分布的表示与特征    | 42 |
| 3.5.3 频数分布的实际应用     | 44 |
| 3.6 数据整理的计算机实现      | 47 |
| 3.6.1 Excel 品质数据整理  | 47 |
| 3.6.2 Excel 数值型数据整理 | 49 |
| 思考与练习               | 52 |
| 单元4 数据特征描述          | 55 |
| 4.1 数据的总量特征         | 56 |
| 4.1.1 总量指标的概念       | 56 |
| 4.1.2 总量指标的种类       | 56 |
| 4.1.3 总量指标的单位       | 57 |
| 4.2 数据的相对特征         | 58 |
| 4.2.1 相对指标的概念       | 58 |
| 4.2.2 计划完成程度相对指标    | 58 |
| 4.2.3 结构相对指标        | 61 |
| 4.2.4 比例相对指标        | 61 |
| 4.2.5 比较相对指标        | 61 |
| 4.2.6 动态相对指标        | 62 |
| 4.2.7 强度相对指标        | 62 |

|                        |           |
|------------------------|-----------|
| 4.3 数据的集中趋势 .....      | 63        |
| 4.3.1 众数 .....         | 63        |
| 4.3.2 中位数 .....        | 64        |
| 4.3.3 分位数 .....        | 66        |
| 4.3.4 平均数 .....        | 68        |
| 4.4 数据的离散程度 .....      | 72        |
| 4.4.1 极差 .....         | 72        |
| 4.4.2 四分位差 .....       | 73        |
| 4.4.3 方差和标准差 .....     | 73        |
| 4.4.4 标准化值 .....       | 75        |
| 4.4.5 离散系数 .....       | 76        |
| 4.5 数据的分布形态 .....      | 76        |
| 4.5.1 偏度及其测度 .....     | 76        |
| 4.5.2 峰度及其测度 .....     | 78        |
| 4.6 统计表 .....          | 78        |
| 4.6.1 统计表的构成 .....     | 78        |
| 4.6.2 统计表的分类 .....     | 79        |
| 4.6.3 统计表的设计要点 .....   | 80        |
| 4.7 统计图 .....          | 80        |
| 4.7.1 线形图 .....        | 80        |
| 4.7.2 雷达图 .....        | 81        |
| 4.7.3 散点图 .....        | 82        |
| 4.7.4 其他图形 .....       | 83        |
| 思考与练习 .....            | 83        |
| <b>单元 5 参数估计</b> ..... | <b>85</b> |
| 5.1 统计推断基础 .....       | 86        |
| 5.1.1 抽样的种类 .....      | 86        |
| 5.1.2 统计量抽样分布 .....    | 88        |
| 5.1.3 样本均值的抽样分布 .....  | 90        |
| 5.1.4 样本比例的抽样分布 .....  | 91        |
| 5.1.5 样本方差的抽样分布 .....  | 92        |
| 5.1.6 抽样平均误差 .....     | 92        |
| 5.2 点估计和区间估计 .....     | 95        |
| 5.2.1 点估计 .....        | 95        |
| 5.2.2 区间估计 .....       | 98        |
| 5.3 单个总体参数区间估计 .....   | 99        |
| 5.3.1 总体均值区间估计 .....   | 99        |
| 5.3.2 总体比例区间估计 .....   | 103       |

|                                 |     |
|---------------------------------|-----|
| 5.3.3 总体方差区间估计                  | 104 |
| 5.4 两总体参数区间估计                   | 106 |
| 5.4.1 两总体均值差区间估计                | 106 |
| 5.4.2 两总体比例差区间估计                | 108 |
| 5.4.3 两总体方差比区间估计                | 109 |
| 5.5 样本容量的确定                     | 110 |
| 5.5.1 总体均值估计的必要样本容量             | 111 |
| 5.5.2 总体比例估计的必要样本容量             | 112 |
| 5.5.3 影响必要样本容量的因素               | 112 |
| 5.6 参数估计 Excel 的实现              | 112 |
| 5.6.1 $\sigma^2$ 未知时, 总体均值的区间估计 | 113 |
| 5.6.2 $\sigma^2$ 已知时, 总体均值的区间估计 | 113 |
| 思考与练习                           | 113 |
| 单元 6 假设检验                       | 115 |
| 6.1 假设检验的基本概念                   | 116 |
| 6.1.1 假设的提出                     | 116 |
| 6.1.2 两类错误和假设检验的规则              | 117 |
| 6.1.3 假设检验的步骤                   | 118 |
| 6.2 单个总体参数的检验                   | 119 |
| 6.2.1 总体均值 $\mu$ 的检验            | 119 |
| 6.2.2 总体比例 $\pi$ 的检验            | 121 |
| 6.2.3 总体方差 $\sigma^2$ 的检验       | 122 |
| 6.3 两个总体参数的检验                   | 124 |
| 6.3.1 两个总体均值之差的检验               | 124 |
| 6.3.2 两个总体比例之差的检验               | 125 |
| 6.3.3 两总体方差相等的检验: $F$ 检验法       | 126 |
| 6.4 单侧检验假设                      | 127 |
| 6.5 Excel 中假设检验的处理              | 130 |
| 6.5.1 单总体均值检验                   | 130 |
| 6.5.2 双总体等均值假设检验                | 131 |
| 思考与练习                           | 132 |
| 单元 7 方差分析                       | 134 |
| 7.1 方差分析概述                      | 135 |
| 7.1.1 方差分析的相关概念                 | 135 |
| 7.1.2 方差分析的基本思想                 | 135 |
| 7.1.3 方差分析的基本假定                 | 136 |
| 7.2 单因素方差分析                     | 136 |



|                              |     |
|------------------------------|-----|
| 7.2.1 各水平实验次数相同时的方差分析        | 137 |
| 7.2.2 各水平实验次数不同时的方差分析        | 140 |
| 7.2.3 方差分析中的多重比较问题           | 144 |
| 7.3 双因素方差分析                  | 145 |
| 7.3.1 无交互作用的方差分析             | 145 |
| 7.3.2 有交互作用的方差分析             | 150 |
| 7.4 方差分析的计算机实现               | 156 |
| 7.4.1 用 Excel 进行单因素方差分析      | 156 |
| 7.4.2 用 Excel 进行无交互作用双因素方差分析 | 158 |
| 7.4.3 用 Excel 进行有交互作用双因素方差分析 | 159 |
| 思考与练习                        | 160 |
| 单元 8 相关与回归分析                 | 163 |
| 8.1 相关关系概述                   | 164 |
| 8.1.1 相关关系的概念                | 164 |
| 8.1.2 相关关系的种类                | 164 |
| 8.1.3 相关关系的研究方法              | 165 |
| 8.1.4 相关关系分析的步骤              | 166 |
| 8.2 相关分析                     | 166 |
| 8.2.1 相关表与相关图                | 166 |
| 8.2.2 相关系数及其检验               | 166 |
| 8.2.3 计算相关系数应注意的问题           | 170 |
| 8.3 一元线性回归分析与预测              | 171 |
| 8.3.1 一元线性回归函数与模型            | 171 |
| 8.3.2 一元线性回归模型的参数估计          | 173 |
| 8.3.3 一元线性回归模型的检验            | 175 |
| 8.3.4 一元线性回归模型的应用            | 177 |
| 8.4 多元线性回归分析与预测              | 179 |
| 8.4.1 多元线性总体回归函数与模型          | 180 |
| 8.4.2 多元线性样本回归函数与模型          | 180 |
| 8.4.3 多元线性模型的参数估计            | 182 |
| 8.4.4 多元线性回归模型的检验            | 182 |
| 8.5 非线性回归分析与预测               | 183 |
| 8.5.1 非线性回归分析的意义             | 183 |
| 8.5.2 非线性回归函数形式的确定           | 184 |
| 8.5.3 非线性回归模型的估计             | 185 |
| 8.6 相关与回归分析的计算机实现            | 186 |
| 8.6.1 用 Excel 绘制散点图          | 186 |
| 8.6.2 用 Excel 计算相关系数         | 187 |

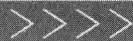
|                                |     |
|--------------------------------|-----|
| 8.6.3 用 Excel 建立回归方程 .....     | 188 |
| 思考与练习 .....                    | 191 |
| <b>单元 9 时间序列分析</b> .....       | 194 |
| 9.1 时间序列分析的一般性问题 .....         | 195 |
| 9.1.1 时间序列的概念 .....            | 195 |
| 9.1.2 时间序列的种类 .....            | 196 |
| 9.1.3 时间序列分析的常用方法 .....        | 198 |
| 9.1.4 时间序列的编制原则 .....          | 198 |
| 9.2 时间序列的动态指标分析 .....          | 199 |
| 9.2.1 时间序列的水平分析 .....          | 199 |
| 9.2.2 时间序列的速度分析 .....          | 203 |
| 9.2.3 应用速度指标应注意的问题 .....       | 205 |
| 9.3 确定性时间序列构成因素分析 .....        | 206 |
| 9.3.1 确定性时间序列分解 .....          | 206 |
| 9.3.2 长期趋势的分解 .....            | 208 |
| 9.3.3 季节变动的分解 .....            | 213 |
| 9.4 随机性时间序列分析的方法 .....         | 219 |
| 9.4.1 随机性时间序列的含义 .....         | 219 |
| 9.4.2 AR 模型 .....              | 219 |
| 9.4.3 MA 模型 .....              | 220 |
| 9.4.4 ARMA 模型 .....            | 221 |
| 9.5 Excel 在时间序列分析中的应用 .....    | 221 |
| 9.5.1 测定平均发展水平 .....           | 221 |
| 9.5.2 测定增长量和平均增长量 .....        | 223 |
| 9.5.3 测定发展速度、增长速度和平均发展速度 ..... | 224 |
| 9.5.4 计算长期趋势 .....             | 225 |
| 9.5.5 计算季节变动 .....             | 225 |
| 思考与练习 .....                    | 227 |
| <b>单元 10 统计指数分析</b> .....      | 229 |
| 10.1 指数概述 .....                | 230 |
| 10.1.1 指数的概念及性质 .....          | 230 |
| 10.1.2 指数的分类 .....             | 230 |
| 10.1.3 指数的作用 .....             | 231 |
| 10.2 总指数的编制 .....              | 232 |
| 10.2.1 综合指数的编制 .....           | 232 |
| 10.2.2 平均指数的编制 .....           | 236 |
| 10.3 指数体系 .....                | 240 |

|                           |            |
|---------------------------|------------|
| 10.3.1 指数体系的概念 .....      | 240        |
| 10.3.2 指数体系的分类 .....      | 241        |
| 10.4 指数因素分析 .....         | 242        |
| 10.4.1 总量指数因素分析 .....     | 242        |
| 10.4.2 平均指标指数因素分析 .....   | 244        |
| 10.5 经济类指数 .....          | 246        |
| 10.5.1 消费者价格指数 .....      | 246        |
| 10.5.2 生产者价格指数 .....      | 249        |
| 10.5.3 股票价格指数 .....       | 250        |
| 10.5.4 采购经理人指数 .....      | 252        |
| 10.5.5 消费者信心指数 .....      | 253        |
| 10.6 大数据时代的新指数 .....      | 254        |
| 10.6.1 百度指数 .....         | 254        |
| 10.6.2 阿里指数 .....         | 256        |
| 10.6.3 微信指数 .....         | 257        |
| 思考与练习 .....               | 257        |
| <b>单元 11 统计综合评价</b> ..... | <b>261</b> |
| 11.1 综合评价概述 .....         | 262        |
| 11.1.1 综合评价的概念 .....      | 262        |
| 11.1.2 综合评价的一般步骤 .....    | 262        |
| 11.2 综合评价指标的选择 .....      | 263        |
| 11.2.1 评价指标体系的构建原则 .....  | 263        |
| 11.2.2 评价指标体系的构建方法 .....  | 264        |
| 11.3 评价指标同度量化处理 .....     | 266        |
| 11.3.1 相对化处理法 .....       | 266        |
| 11.3.2 功效系数法 .....        | 266        |
| 11.3.3 标准化法 .....         | 267        |
| 11.4 综合评价指标权重的设定 .....    | 267        |
| 11.4.1 权重的分类 .....        | 268        |
| 11.4.2 定性指标权重的确定方法 .....  | 268        |
| 11.4.3 定量指标权重的确定方法 .....  | 269        |
| 11.5 综合评价指标加总方法 .....     | 272        |
| 11.5.1 加权算术平均综合评价模型 ..... | 272        |
| 11.5.2 加权几何平均综合评价模型 ..... | 272        |
| 思考与练习 .....               | 273        |
| <b>参考文献</b> .....         | <b>274</b> |

# 1

## 单元

## 绪 论



### ◎ 导入案例

2017年热播的反腐剧《人民的名义》引发众多网友称赞。剧中京州市市委书记李达康因做派耿直、性情直爽、敢想敢干而赢得无数赞扬。围绕他，人们议论的话题可谓千奇百怪，其中有人认为“达康书记的眼里只有GDP”。那么，什么是GDP (gross domestic product, 国内生产总值)呢？它是如何计算的？它有什么优缺点？大家未必能回答得全面。学习完本单元，大家就会明白GDP和其他统计指标的相关知识。由此可见，统计学知识在我们身边随处可见，细心的你能否再举一些例子呢。

### ◎ 学习内容

1. 统计学的研究对象。(重点)
2. 统计学的基本概念。(重点)
3. 统计学的研究方法。(难点)
4. 统计学的应用领域。

### ◎ 学习目标

1. 掌握统计学研究对象的特征。
2. 掌握统计学的基本概念。
3. 理解统计学的研究方法。
4. 了解统计学的应用领域。

## 1.1 学习应用统计学的三种境界

### 王国维的三境界论

我国晚清学者王国维在《人间词话》中对做学问的“三境界”的描述如下：

第一境界：昨夜西风凋碧树，独上高楼，望尽天涯路。

第二境界：衣带渐宽终不悔，为伊消得人憔悴。

第三境界：蓦然回首，那人却在灯火阑珊处。

其中，第一境界可以看作寻找目标、发现问题的阶段；第二境界是不断探索、试探解决问题的阶段；第三境界是顿悟灵感的阶段。

凡事都有境界。我们学习应用统计学也应该有三种境界。

第一种境界，统计是一门学科。在第一种境界，我们要上课、写作业、参加考试，以掌握统计学知识为主。在掌握了统计学知识后，再在实践中去应用，就进入第二种境界。在这个境界，统计就成了一种工具。

在第二种境界，我们会用所学的统计学的方法和工具来解决实际问题。例如，马克思的《资本论》将近 180 万字，其中将近 100 万字与统计资料有直接的关系。马克思搜集了大量的数据，除了《大英联合王国统计资料汇编》《爱尔兰农业统计》，以及英格兰和威尔士 1851 年的普查资料等专门出版物外，也有英国议会众多委员会的材料及各类调研报告，还有一些工厂的实际资料。另外，马克思对通过艰辛劳动所获取的统计资料并不是简单地照搬照抄，而是以审慎的态度经过一番去粗取精、去伪存真的筛选才引用到《资本论》中。在《资本论》中，随处可以见到算术平均数、调和平均数、相对数、统计指数等基本统计指标的影子。在一定程度上可以说，正是有了这些统计数据，以及马克思对这些数据的分析，才成就了《资本论》。正如马克思所说，研究必须充分地占有资料，并分析它的各种发展形式，以探寻这些形式的内在联系。又如，我国著名经济学家林毅夫教授，从美国学成回国后，一直在国务院农村发展研究中心工作。在 20 世纪 80 年代后期，我国农业发展陷入徘徊状态，一些人开始怀疑家庭联产承包责任制改革是否合适，一时之间不同观点相持不下。林毅夫教授通过严格的统计分析，得出结论：1978~1984 年，我国不变价农产品产值增长 42.23%，其中 46.89% 归功于家庭联产承包责任制，高于化肥施用量引起的 32.2%，因此家庭联产承包责任制是这一时期农业大发展的主要源泉。

在利用统计学方法解决应用问题之后，我们继续研究，当统计对我们来说成为一种思想时，我们就会如统计学家一样思考问题。我们的思想境界、思维方式等都会达到一种新的高度，这就到了统计的第三种境界。我国著名统计学家邱东教授写过一本散文集——《统计使人豁达》，其中提到良好的统计训练使人的宏观意识增强，看待和处理问题时不那么偏激，还会培养人的分寸感。我们再来看另外一位统计学家的例子。第二次世界大战时期，美国空军遇到一个关键问题：在欧洲和太平洋地区的盟军飞机正在以惊人的速度被击落。美英联军迫切需要对所有的飞机性能进行加固，军事专家认为给飞机底层加一层钢板，可以减少飞机被击落的概率，但这样会使飞机质量过大从而带来负面影响。美国国防部委

托统计学家亚伯拉罕·沃尔德为飞行在敌人领土上空的飞机设计和开发新的装甲防护。空军为他提供来自前线的数据,包括返回基地的所有飞机上弹孔的数量和位置。由于返回的飞机的伤痕集中在尾部和机翼,所以空军高层建议加强对这些伤痕累累部位的装甲。沃尔德激烈反对,他认为返回的飞机上的标记显示这是飞机最强大的部位,这些部位能承受伤害并依然能够返回基地。因此,沃尔德认为这些部位并不是需要加强的关键。有成千上万的飞机没有飞回来,因为它们被击中其他地方:发动机、尾部和驾驶舱,这些部位更应该被重点保护。美国军方采纳了沃尔德的建议,从此飞机被击落的概率大大减小。后来沃尔德写了一篇文章,将这种现象称为“生存偏差”,这是一种困扰所有领域数据分析师的认知偏差。一个统计学家的思想改变了美国飞机的命运,统计思想的威力可见一斑。

## 1.2 什么是统计学

### 1.2.1 统计学的定义

统计学,顾名思义,是一门学科,既然是一门学科,就与理论有关。这就和我们通常所理解的统计工作和统计资料区别开来。人们对于统计学有不同的定义。《大英百科全书》认为统计学是一门搜集数据、分析数据,并根据数据进行推断的艺术和科学,最初与政府搜集数据有关,现在又包括了范围广泛的方法和理论。《美国百科全书》把统计学定义为一门在不确定性方面为了做出正确的推断而进行搜集、分析定量数据的科学和艺术。而《中国大百科全书》将统计学定义为一门研究如何有效地搜集、整理和分析带有随机性的数据,以对所考察的问题做出推断或预测,直至为采取一定的决策和行动提供依据和建议的学科。由此可见,统计学与数据息息相关。

统计学是统计工作的理论升华,从工作流程来说,统计工作一般包括统计设计、数据搜集、统计整理、统计分析等主要环节。统计设计主要是对统计指标、标准、统计方法、调查单位库、报表制度、工作流程等方面进行设计,包括制定统计标准和工作流程、设计指标体系、设计统计报表制度、设计统计调查方案等内容,属于统计工作的顶层设计环节。数据搜集则是根据统计研究的目的和要求,有组织、有计划地向客观现象搜集统计资料的工作过程,既包括对原始数据的统计调查,又包括对间接数据的搜集。它是统计认识事物的起点,是统计定量研究的开始。其中,统计调查在整个统计工作过程中担负着提供基础资料的任务,它的工作质量在很大程度上决定着统计工作全过程的质量。统计整理就是将搜集到的原始资料加以科学分组、归纳、综合,使原始资料条理化、系统化,从而获得比较简明的再生信息。它是从现象的描述过渡到现象规律性的认识,进而做出判断、评价的起点。这一阶段是统计工作的一个中间环节。统计分析就是对整理好的统计资料进行多方面的分析研究,以认识事物的本质和规律为目的的工作过程,其目的是寻找被研究现象的特征和规律,将现象由定量认识深化到定性认识,并提出见解。这一阶段是统计工作过程的决定性阶段,是较能体现统计价值的阶段。

### 1.2.2 统计学的发展简史

统计学作为一门学科已有 300 多年的历史。按统计方法及历史的演变顺序,通常可以



将统计学的发展史分为三个阶段，分别是古典统计学时期、近代统计学时期和现代统计学时期。

古典统计学的萌芽最早可以追溯到 17 世纪中叶，此时的欧洲正处于封建社会解体和资本主义兴起的阶段，工业、手工业快速增长，社会经历着重大变革。政治改革家们急需辅助国家经营和管理的数据证据以适应经济发展的需要，此时一系列统计学的奠基工作在欧洲各国相继展开。在这一时期，以威廉·配第和约翰·格朗特为代表的政治算术学派与海尔曼·康令创立的国势学派相互渗透和借鉴，服务与指导国家管理并改善了社会福利。

18 世纪末至 19 世纪末为近代统计学时期。这 100 年间欧洲各国先后完成了工业革命，科学技术开始进入全面繁荣时期，天文、气象、社会人口等领域的数据资料也有了一定规模的积累，对统计的需求已从国家层面扩展至社会科学各个领域。对事物现象静态性的描述已不能满足社会需求，后来数理统计学派创始人凯特勒率先将概率论引入古典统计学，提出了大数定律思想，使统计学逐步成为揭示事物内在规律、可用于任何科学的一般性研究方法。一些重要的统计概念也在这一时期提出，误差测定、正态分布曲线、最小二乘法、大数定律等理论方法的大量运用为社会、经济、人口、法律等领域的研究提供了指导。

20 世纪科学技术的发展速度远超之前的时代，以描述性方法为核心的近代统计已无法满足需求，统计学的重心开始转为推断性统计，统计学进入现代统计学时期。20 世纪初细胞学开始发展，农业育种工作全面展开。1923 年，英国著名统计学家费希尔为满足作物育种的研究需求，提出了基于概率论和数理统计的随机实验设计技术及方差分析等一系列推断统计的理论和方法。推断性统计方法的应用对工农业生产和科学研究起到了很大的促进作用。自 20 世纪 30 年代，随着社会经济的发展和医学先进理念的吸收融合，人们对于医疗保险和健康管理的需求日益增长，统计思想渗透到医学领域，形成了现代医学统计方法。例如，在生存质量研究领域，通过分析横向、纵向资料，逐步形成了重复测量资料的方差分析、质量调整生存年法等统计方法。这一阶段，统计在分子生物学、临床试验等生物医学领域获得了大量的应用，这些领域的发展又带动统计方法不断创新，主成分估计、非参数估计等方法应运而生。随着现代生物医学的发展，计算机技术的进步，人类对健康的管理和疾病的治疗已进入基因领域，从而对基因数据分析产生了大量需求。高维海量的基因数据具有全新的数据特征，由此一系列面向高维数据的统计分析方法相继产生。

那么，到了 21 世纪，统计学又有什么新发展了呢？我们将在 1.5 节介绍。

### 1.2.3 统计学的应用领域

目前统计方法已被应用到自然科学和社会科学的众多领域，统计学也已经发展成为由若干分支学科组成的学科体系。可以说，大部分涉及数据分析的研究领域都要用到统计方法。

1) 企业发展战略：发展战略是一个企业长远发展的方向。制定发展战略，一方面，需要及时了解和把握整个宏观经济的状况及发展趋势，了解市场的变化；另一方面，需要对企业进行合理的市场定位，把握企业自身的优势和劣势。这些离不开统计学，需要运用统计学提供可靠的数据，并利用统计学方法对数据进行科学的分析和预测等。

2) 产品质量管理：质量是企业的生命，是企业持续发展的基础。在一些知名的跨国公司，6 $\sigma$ 准则已成为一种重要的管理理念。质量控制已成为统计学在生产领域中的一项重要

应用,各种统计质量控制图被广泛应用于监测生产过程。

3) 市场研究:企业要在激烈的市场竞争中取得优势,首先必须了解市场,要了解市场,就需要广泛的市场调查,取得所需的信息,并对这些信息进行科学的分析,以便作为生产和营销的依据,这些需要统计学的支持。

4) 财务分析:上市公司的财务数据是投资决策的重要参考依据。一些投资咨询公司根据上市公司的财务和统计数据进行分析,为股民提供咨询参考。企业自身的投资,也离不开对财务数据的分析,其中要用到大量的统计学方法。

5) 经济预测:企业要对未来的市场进行预测,经济学家也常常对宏观经济或某一方面进行预测。他们在进行预测时要用到各种统计信息和统计学方法。例如,企业要对自己产品的市场潜力做出预测,以便及时调整生产计划,这就需要利用市场调查取得数据,并对数据进行统计分析。经济学家在预测通货膨胀时,要利用有关生产价格指数、失业率、生产能力等统计数据,以及统计模型进行预测。

6) 人力资源管理:利用统计方法对企业员工的年龄、性别、受教育程度、工资等进行分析,并以此作为企业制订工资计划、奖惩制度的依据。

## 1.3 统计学的基本概念

### 1.3.1 总体和样本

统计总体,就是我们所要研究事物的全体。例如,我们要研究中国的人口状况,则所有拥有中国国籍的人就是研究的总体。构成总体的每个个体单元称为总体单位,也称为个体。统计总体和总体单位是相对的。例如,我们在研究全国或各地区通信企业的情况时,每个企业是总体单位,当研究一个企业内部问题时,该企业则为总体。

当然,并不是任何事物的全体都可以构成总体。要成为一个总体,首先要有足够多的个体单位,即大量性。另外,不同的总体单位要同时具有共性和差异性。例如,要成为全国人口普查的对象,那么每个人至少在国籍上都是一样的,但每个人在其他方面存在着差异。统计学就是要从这些差异中寻找规律。

从统计总体中抽取一部分单位构成的集合就是样本。根据概率抽样得到的样本就是随机样本,不是根据概率抽样得到的样本称为非随机样本。

我们现在处在大数据时代,大数据具有以下 4V 特征,即海量(volume)、时效与速度(velocity)、价值(value)、多源异构(varity),然而很多所谓的大数据并不完全同时满足以上四个特征。此外,大数据这个概念是随着时间、空间和技术而变化的,是相对于当前的技术和资源而言的。在大数据时代,总体和样本的概念同样适用,甚至有一种说法认为在大数据时代样本等于总体。

### 1.3.2 统计指标和指标体系

统计指标是描述总体现象数量特征的概念,通常用于总体的范围。而描述一个总体中单个单位特征的概念,通常称为标志。目前的一些新教材中,模仿了国外教材的写法,不再介绍统计指标和统计标志的概念,但教材中经常用到这些概念,在现实中,我们也经常使用这些概念。



根据统计指标的表现形式不同,统计指标可分为总量指标、相对指标、平均指标、变异指标等。总量指标是说明一个总体的总规模、总水平,或者反映完成工作量多少的统计指标,是以绝对数形式表示的,因此也称为绝对数。相对指标是用两个有联系的指标之比



国际 ICT 发展指标  
专题研讨会

来反映现象之间联系程度的统计指标,是以相对数形式表示的,因此也称为相对数。平均指标是用一个综合数值来反映总体所有单位在某一个数量标志下所达到的一般水平的统计指标,是以平均数形式表示的,因此也称为平均数。变异指标通常用来表示一组数据与其均值的离散程度。另外,有时还会用到数量指标与质量指标的概念,数量指标是反映总体的规模和总水平的指标,用绝对数表示;质量指标是反映现象之间的各种对比关系和总体一般水平的指标,用相对数和平均数表示。

在现实中,我们要描述一个事物的特征,使用一个指标往往是不够的,这时通常要选择若干有联系的指标来描述该事物,我们把这些有联系的指标称为指标体系。许多问题指标体系的构建也是一个研究的热门话题。例如,国际电信联盟在分析各种信息化发展水平时,就利用了信息技术接入、应用和技能三个方面的不同指标来反映。再如“全面建成小康社会”是一个抽象的概念,但我们根据相关理论,可以把它概括为经济发展、环境优化、民生改善等方面。

另外,统计指标在不同的场合可能会有不同的叫法。例如,在抽样推断中,我们需要根据样本的数量特征推断总体的特征,这时我们把表示样本数量特征的随机变量称为统计量,把说明总体数量特征的变量称为参数。在这种情况下,由于样本是不确定的,所以统计量是已知的,但它不是唯一的,而总体参数往往是未知的,但它是唯一的。抽样推断就是要由已知的、不唯一的样本统计量来推断唯一的且未知的总体参数。

### 1.3.3 变量和数据

通常,我们把描述事物总体或总体单位特征的概念统称为变量,而把变量的取值称为变量值,也称为数据。例如,人的性别是一个变量,而一个人性别是男是女就是数据。统计数据是统计学的研究对象。有什么类型的变量,就有什么类型的统计数据,不同类型的统计数据的性质是不同的。从不同的角度,数据可以划分为不同的类型。

#### 1. 定性数据和定量数据

定性数据和定量数据是按照数据的计量尺度进行的分类。

1) 定性数据没有量的特征,只是描述事物的属性特征。定性数据可以分为分类数据和顺序数据。分类数据,也称为定类数据,是指只能归于某一类别的非数字型数据,如性别中的男女就是分类数据。顺序数据,也称为定序数据,是指只能归于某一有序类别的非数字型数据,如产品的等级。

2) 定量数据称为数值型数据,具有明确的数学含义,不仅可以分类,而且能够测量出具体的大小和差异。数值型数据是按数字尺度测量的观察值,它是自然或度量衡单位对事物进行测量的结果,如一个国家的总人口数、一个人的身高等。