

普通高等学校“十三五”规划教材

统 计 学

Statistics

叶仁道 刘干 薛洁 主编



西安电子科技大学出版社
<http://www.xduph.com>

普通高等学校“十三五”规划教材

统 计 学

叶仁道 刘 干 薛 洁 主编

西安电子科技大学出版社

内 容 简 介

本书系统性地介绍了统计学的基本理论、方法和应用。全书共分 12 章, 主要内容包括导论、数据收集、统计数据整理与显示、统计数据的分布特征描述、概率基础、参数估计、假设检验、列联分析、方差分析、相关与回归分析、时间数列分析、统计指数等。本书力求使用较少的数学知识, 强调统计学概念与方法的阐释, 并注意举例的多样性。

本书可作为高等学校经济、管理、人文、社科、工科、农业和医学等相关专业的统计学课程的教材, 亦可作为实际统计工作者的参考书。

图书在版编目(CIP)数据

统计学/叶仁道, 刘干, 薛洁主编. —西安: 西安电子科技大学出版社, 2016. 8

普通高等学校“十三五”规划教材

ISBN 978-7-5606-4239-0

I. ① 统… II. ① 叶… ② 刘… ③ 薛… III. ① 统计学—高等学校—教材 IV. ① C8

中国版本图书馆 CIP 数据核字(2016)第 189195 号

策 划 马乐惠

责任编辑 马乐惠 王 静

出版发行 西安电子科技大学出版社(西安市太白南路 2 号)

电 话 (029)88242885 88201467 邮 编 710071

网 址 www.xduph.com 电子邮箱 xdupfxb001@163.com

经 销 新华书店

印刷单位 陕西天意印务有限责任公司

版 次 2016 年 8 月第 1 版 2016 年 8 月第 1 次印刷

开 本 787 毫米×1092 毫米 1/16 印 张 16.5

字 数 389 千字

印 数 1~3000 册

定 价 29.00 元

ISBN 978-7-5606-4239-0/C

XDUP 4531001-1

*** 如有印装问题可调换 ***

前 言

随着物联网、云计算、大数据时代的兴起，政府部门、大中型企业、金融机构和咨询机构面临着大量的数据处理工作，包括对海量数据进行采集、存储、分析、展示、解释、推断、预测等，这些都需要统计学概念、理论、方法等专业知识作为支撑。

统计学作为一门关于数据收集、整理和分析的方法性科学，其目的是探索数据内在的数量规律性，以达到对客观事物的科学认识，进而为决策服务提供数据支撑。本书共有 12 章，可分为两部分。第一部分为描述统计，包括导论、数据收集、统计数据整理与显示、统计数据的分布特征描述等；第二部分为推断统计，包括概率基础、参数估计、假设检验、列联分析、方差分析、相关与回归分析、时间数列分析、统计指数等。

本书的完稿凝聚了许多一线教师的心血。其中，第 1、3、4、10 章由薛洁编写，第 2、9、11、12 章由刘干编写，第 5、7、8 章由叶仁道编写，第 6 章由罗堃编写，全书由叶仁道统稿。本书的编写工作得到了杭州电子科技大学校级教材建设项目的资助，并参考了一些国内外优秀的统计学教材，在此表示诚挚的谢意。

由于编者水平所限，书中不当之处在所难免，恳请广大同行及读者不吝赐教。

编 者

2016 年 5 月

目 录

第 1 章 导论	1	3.1.3 统计资料的审核	40
1.1 什么是统计	1	3.2 统计分组	41
1.1.1 统计的含义	1	3.2.1 统计分组的概念与原则	41
1.1.2 统计研究的基本过程与方法	3	3.2.2 统计分组的种类	41
1.2 统计学的产生与发展	6	3.2.3 统计分组的方法	42
1.2.1 统计学的发展历程	6	3.3 频数分布	45
1.2.2 统计学的发展动向与趋势	8	3.3.1 频数分布的概念	45
1.2.3 统计学与相关学科的关系	10	3.3.2 分布数列的编制	45
1.2.4 为什么学习统计学	12	3.3.3 累计频数和累计频率	47
1.3 统计数据的类型	13	3.4 统计数据的显示	48
1.3.1 数据的计量尺度	13	3.4.1 统计表	48
1.3.2 数据的类型	14	3.4.2 统计图	51
1.4 统计学的基本概念	15	本章小结	57
1.4.1 总体与样本	15	思考与练习	57
1.4.2 标志与指标	16	拓展阅读	59
1.4.3 统计指标体系	18		
本章小结	22	第 4 章 统计数据的分布特征描述	62
思考与练习	22	4.1 数据对比分析	62
拓展阅读	23	4.1.1 数据对比分析的概念及其表现形式	62
第 2 章 数据收集	28	4.1.2 相对指标的种类	62
2.1 统计调查的种类与方法	28	4.2 数据集中趋势的测量	63
2.1.1 统计调查的种类	28	4.2.1 集中趋势测量的含义和作用	63
2.1.2 统计调查的常用方法	28	4.2.2 数值平均数	64
2.1.3 收集数据的方法	30	4.2.3 位置平均数	70
2.2 统计调查方案设计	32	4.3 数据离散趋势的测量	73
2.3 调查问卷设计	34	4.3.1 离散趋势测量的意义	73
2.3.1 问卷结构	34	4.3.2 极差、四分位差与平均差	74
2.3.2 问卷设计步骤	35	4.3.3 方差与标准差	76
2.3.3 设计问卷应注意的问题	36	4.3.4 变异系数	81
本章小结	37	4.4 分布的偏度和峰度	81
思考与练习	38	4.4.1 矩	81
		4.4.2 偏度系数	82
第 3 章 统计数据整理与显示	39	4.4.3 峰度系数	83
3.1 统计数据整理	39	本章小结	84
3.1.1 统计数据整理的概念	39	思考与练习	84
3.1.2 统计数据整理的内容与程序	39		

第 5 章 概率基础	87	7.1.3 p -值检验	120
5.1 基本概念	87	7.1.4 双侧检验与单侧检验	120
5.1.1 随机试验与随机事件	87	7.1.5 两类错误	121
5.1.2 事件的概率	88	7.1.6 检验功效	122
5.2 概率性质与运算法则	90	7.2 总体均值的检验	123
5.2.1 概率性质	90	7.2.1 方差已知时	123
5.2.2 运算法则	90	7.2.2 方差未知时	124
5.3 离散型随机变量及其分布	94	7.3 总体成数的检验	125
5.3.1 随机变量的定义	94	7.4 两个总体参数的检验	127
5.3.2 离散型随机变量	94	7.4.1 两个总体均值之差的检验	127
5.3.3 几种常见离散型分布	95	7.4.2 两个总体成数之差的检验	128
5.4 连续型随机变量及其分布	96	7.4.3 两个总体方差之比的检验	129
5.4.1 密度函数与分布函数	96	本章小结	130
5.4.2 几种常见的连续型分布	97	思考与练习	131
5.5 大数定律与中心极限定理	101	第 8 章 列联分析	133
5.5.1 大数定律	101	8.1 列联表概述	133
5.5.2 中心极限定理	102	8.1.1 分类数据	133
本章小结	102	8.1.2 列联表	133
思考与练习	102	8.2 列联表的检验	136
第 6 章 参数估计	104	8.2.1 χ^2 检验概述	136
6.1 抽样分布	104	8.2.2 独立性检验	137
6.1.1 抽样的基本概念	104	8.2.3 拟合优度检验	138
6.1.2 抽样分布	106	8.3 列联表中相关程度测量	139
6.2 点估计及评价标准	108	8.3.1 φ 相关系数	139
6.2.1 点估计	108	8.3.2 列联相关系数	140
6.2.2 估计量的评价标准	109	8.3.3 V 相关系数	141
6.3 区间估计	110	本章小结	141
6.3.1 总体均值的区间估计	110	思考与练习	141
6.3.2 总体成数的区间估计	112	第 9 章 方差分析	145
6.4 样本容量的确定	113	9.1 方差分析概述	145
6.4.1 估计总体均值时样本容量的 确定	113	9.1.1 方差分析的概念及常用术语	145
6.4.2 估计总体成数时样本容量的 确定	114	9.1.2 方差分析的基本原理	148
本章小结	115	9.1.3 方差分析的基本假定	148
思考与练习	115	9.2 单因素方差分析	149
第 7 章 假设检验	118	9.2.1 单因素方差分析的程序和方法	149
7.1 假设检验的基本问题	118	9.2.2 方差分析中的多重比较	151
7.1.1 假设检验的概念	118	9.3 双因素方差分析	152
7.1.2 假设检验的步骤	119	9.3.1 双因素方差分析的种类	152
		9.3.2 无交互作用的双因素 方差分析	152

9.3.3 有交互作用的双因素 方差分析	155	11.4.1 时间数列的构成要素与模型	203
本章小结	159	11.4.2 长期趋势分析	204
思考与练习	159	11.4.3 季节变动分析	213
第 10 章 相关与回归分析	161	本章小结	215
10.1 相关分析	161	思考与练习	216
10.1.1 函数关系和相关关系	161	第 12 章 统计指数	219
10.1.2 相关关系的种类	162	12.1 指数的概念和种类	219
10.1.3 相关关系的判断与测度	162	12.1.1 指数的概念和作用	219
10.2 一元线性回归分析	167	12.1.2 指数的种类	220
10.2.1 回归分析的概念和特点	167	12.1.3 指数的特性	221
10.2.2 标准的一元线性回归模型	168	12.2 综合指数	222
10.2.3 一元线性回归模型的估计	170	12.2.1 综合指数的概念和编制方法	222
10.2.4 一元线性回归模型的检验	172	12.2.2 同度量因素的确定及综合指数的 基本公式	223
10.2.5 一元线性回归模型的预测	176	12.2.3 拉氏指数	225
10.3 多元线性回归分析	178	12.2.4 帕氏指数	226
10.3.1 标准的多元线性回归模型	178	12.2.5 拉氏指数和帕氏指数的比较	227
10.3.2 多元线性回归模型的估计	179	12.3 平均指数	227
10.3.3 多元线性回归模型的检验	180	12.3.1 平均指数的概念和编制方法	227
10.3.4 多元线性回归模型的预测	182	12.3.2 算术平均指数	228
10.4 非线性回归分析	185	12.3.3 调和平均指数	229
10.4.1 非线性函数形式的确定	185	12.3.4 固定权数的平均指数	230
10.4.2 非线性回归模型的估计	186	12.4 指数体系与因素分析	230
本章小结	188	12.4.1 指数体系的概念与作用	230
思考与练习	188	12.4.2 指数因素分析法的 意义和种类	232
第 11 章 时间数列分析	191	12.4.3 指数因素分析法的应用	232
11.1 时间数列的概念和编制	191	12.5 几种常见的经济指数简介	238
11.1.1 时间数列的概念	191	12.5.1 物价指数	238
11.1.2 时间数列的种类	191	12.5.2 工业生产指数	239
11.1.3 时间数列的编制	193	12.5.3 股票价格指数	241
11.2 时间数列的水平指标	193	12.5.4 进出口贸易指数	242
11.2.1 发展水平	193	本章小结	243
11.2.2 平均发展水平	194	思考与练习	243
11.2.3 增长量	199	附录 常用统计表	245
11.2.4 平均增长量	200	附表 1 标准正态分布表	245
11.3 时间数列的速度指标	200	附表 2 t 分布表	246
11.3.1 发展速度	200	附表 3 F 分布表	247
11.3.2 增长速度	201	附表 4 χ^2 分布表	253
11.3.3 平均发展速度	202	参考文献	255
11.3.4 平均增长速度	203		
11.4 时间数列分解分析	203		

第1章 导 论

日常生活中,我们经常会接触到各种统计数据,比如国内各大航空公司航班延误情况、CPI指数、70个大中城市住宅销售价格指数、PMI指数等。当然,还有一些结论,如身材高的父亲,其子女的身材也较高等。这些结论是否正确,你相信这些结论吗?要想正确理解数据,并从中得到正确的结论,需要具备一定的统计学知识。本章是全书的导入篇,主要介绍统计学的发展和基本概念,为学习以后各章奠定基础。

1.1 什么是统计

1.1.1 统计的含义

伴随着物联网、云计算、大数据时代的接踵而至,政府部门、大中型企业、金融机构和咨询机构面临大量的数据处理工作,包括对海量、类型多样的数据进行采集、存储、分析、展示、解释、推断、预测等,这些工作正是统计学科的基本范畴,统计理论与方法已成为国家宏观管理与决策、企业内部管理与决策、科学研究等的重要理论工具和实用方法。那么,究竟什么是统计?统计的研究对象是什么?统计是如何开展研究的?统计中常用的研究方法有哪些?这是本节要讲述的主要内容。

1. 日常生活中统计的含义

日常生活中,我们经常会听到关于“统计”二字不同的用法。例如,足球世界杯要“统计”各个参赛球队的积分;生产车间要“统计”每天的产品产量;银行要“统计”存款额、贷款额等。人们也常常从互联网、报纸杂志、电视新闻中获得我国居民消费价格指数、房地产价格指数、经济增长速度等统计数据。再如,我们问“某某同学您是学什么的?”学生回答说:“我是学统计的”,等等。可以看出,在不同的场合,统计的含义是不同的。通常情况下,统计有三种含义:一是统计实践活动,二是统计数据,三是统计学。

统计实践活动,也称为统计工作,它是对社会现象客观存在的数量方面进行搜集、整理和分析的活动过程。如上述的统计比赛积分、统计产量和统计存款额等都是在从事一项统计实践活动。

统计数据,也称为统计资料,它是指在统计实践活动过程中取得的各项数据资料和其他相关资料的总称。如上述的比赛积分、产量、存款额、居民消费价格指数、房地产价格指数、经济增长速度等都是统计数据。

统计学是一门方法论科学,它是对大量的各类社会现象数量方面的调查研究。它是研究如何测定、搜集、整理和分析数据,最终为决策服务的方法论科学。我们正在学习的统计,就是统计学。例如,如何搜集所需的统计资料?对搜集到的统计资料应如何整理以反映其分布特征?用什么样的方法对统计资料进行深入分析以反映现象变化的规律性?此类

问题都需要统计学知识来解决。

从上面的讨论中可以明显看出,“统计”在日常生活中的三种含义之间存在着密切的联系。如图 1.1 所示,统计实践活动与统计数据是活动过程与成果的关系,而统计学与统计实践活动又是理论与实践的关系,理论源于实践,且高于实践,反过来又指导实践。



图 1.1 统计的三种含义及其关系

2. 统计的研究对象及其特点

一般来说,统计研究的是客观世界总体的数量特征和数量关系,通过这些数量方面反映客观世界规律性的表现。一方面,统计研究的是大量客观世界中总体的数量方面,反映该现象发展变化的规律性在具体时间、地点和条件下的数量表现,揭示事物的本质、联系、发展规律和发展趋势;另一方面,统计研究的是现象的综合数量,也就是说,统计要集合大量单位的调查资料,汇总概括,得出反映客观现象总体的数量特征,说明其变化规律性。

综上所述,统计研究对象的特点可以归纳为如下三点:

1) 数量性

数量性是统计研究对象的基本特点,数据是统计的语言、核心,也是统计的原料。任何客观事物都是由质和量两个方面构成的,一定的质规定着一定的量,一定的量又表现为一定的质,二者密切联系、共同规定着事物的性质。但是从认识的角度来看,质和量是可以区分的,可以在一定质的情况下,单独研究事物的数量方面,通过认识事物的量认识事物的质。因此,事物的数量是我们认识客观现象的重要方面,通过分析统计数据,得到其统计规律性,就可以达到我们统计研究的目的。例如,要研究固定资产投资,就要对其总额、构成及其数量变化趋势等进行分析,这样才能准确地掌握固定资产投资的规律性;要分析居民消费价格指数,就要对城镇、农村居民的消费价格指数、构成及其增长速度等的变化趋势进行了解,从而准确地分析我国城镇、农村居民消费价格指数的规律性。

2) 总体性

统计以客观世界总体的数量方面为研究对象。虽然统计研究从调查个体入手,对个体的具体事实进行观察研究,但是最终的目的是认识总体的数量特征。例如,对职工工资进行统计分析,我们并不是要分析和研究单个职工的工资,而是要分析和研究一个地区、一个部门、一个企事业单位的总体工资情况及其规律性。统计研究要做的就是从个体的具体数量归纳出客观现象的总规模、总水平,由此得出客观现象的比例关系和总趋势。当然,这并不代表统计忽略对个体事物的深入研究,相反,对具有代表性的典型单位进行具体分析,了解现象的内在联系,将有助于更好地认识总体的规律性。

3) 变异性

统计研究同类现象总体的数量特征,它的前提是总体各单位数量特征或属性特征表现存在差异,而且这种差异不是由某种固定的原因事先给定的,而是由随机因素引起的,统计上称之为变异。例如,一个地区的人口有多有少,居民的文化程度有高有低、收入有

高有低,消费水平也存在着高低差异,基于此,才需要研究该地区的人口总数、居民的文化结构、平均收入、平均消费水平等统计指标。如果各个单位间不存在任何差异,那么就不需要做统计;如果各单位间的差异是按已知条件可以事先推出的,那么也不需要用统计方法。例如一年内各日昼夜时间长短随着季节的变化而有规律地变化着,这属于大自然现象,与统计无关;而股票价格高低随时间不同,它是由多种复杂因素引起的,则是统计的研究对象。

1.1.2 统计研究的基本过程与方法

1. 统计研究的基本过程

一般来说,利用统计认识客观世界总体数量规律的全过程包括统计设计、统计调查、统计整理、统计分析和统计资料的管理与开发应用,这一过程如图 1.2 所示。

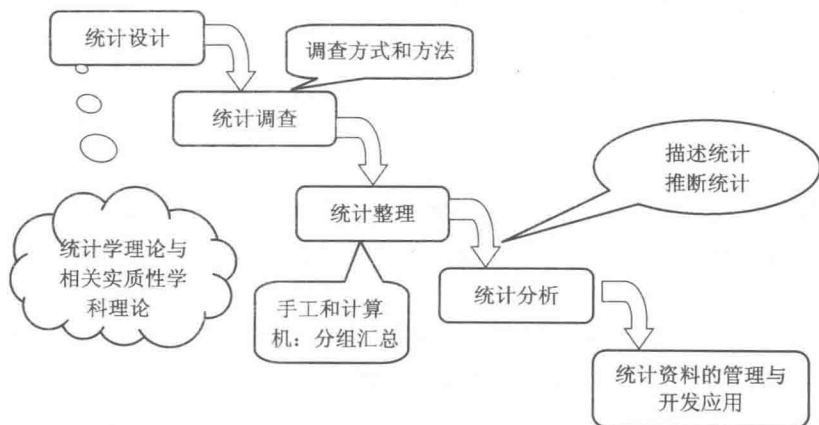


图 1.2 统计研究的基本过程

1) 统计设计

统计设计是统计研究工作的起始,它是对统计活动进行全面安排的关键环节。一般地,接到一项统计研究课题后,应该先对该研究课题进行通盘考虑,合理安排统计活动中的各个方面和各个环节,并形成一套完整的统计设计方案。它类似于项目计划书,其内容包括统计工作的目标、研究对象、研究内容(通常以指标体系来反映)、数据的调查方案和整理方案,以及数据分析、提供与保管,还有工作期限和进度安排等方面,最终的统计研究工作将以此为依据来展开。例如,要了解民众对社区居家养老问题的看法,需要明确调查背景与目的,明确调查对象与范围,根据研究内容设计调查表,制定抽样方案,指出数据的搜集方法,整理方案 and 数据分析方法,并在人员和时间上有统一安排。当然,在进行统计设计的时候,仅具有一般的统计方法知识是远远不够的,还需要具备相关研究领域的学科理论知识。

2) 统计调查

统计调查即统计数据的搜集环节。它是根据统计设计方案的要求,采用各种统计调查组织形式和调查方法,有计划、有组织地向客观实际搜集准确而充分的统计数据的过程。统计调查是统计认识活动由定性过渡到定量认识的阶段,统计研究过程中各种数据皆由此

环节取得,这个阶段所搜集的资料是否客观、全面、系统、准确、及时,将直接影响到统计整理的好坏、统计分析结论的正误。所以,该环节是统计研究工作的基础,决定了统计研究工作完成的质量。常用的统计调查方式有统计报表制度、普查、抽样调查、重点调查和典型调查 5 种,调查方法有观察实验法、访问调查法两种,本书将在第 2 章对每一种调查方式、方法做详细介绍。

3) 统计整理

统计整理是对统计调查得到的数据进行分组汇总的过程。通过对资料的整理使资料系统化、条理化,把反映各个单位个体特征的资料转化为反映总体特征的资料,方便后续的资料统计分析。统计整理介于统计调查和统计分析之间,属于统计研究工作的第三阶段,在统计工作中具有承上启下的作用,既是统计调查阶段的延续,又是统计分析的基础和前提,是人们对社会经济现象从感性认识上升到理性认识的过渡阶段。统计整理的步骤主要包括:① 制定统计整理方案;② 对统计调查资料进行审核;③ 根据研究目的,选择分组标志,建立统计分组;④ 对分组后的资料进行汇总和计算,使得反映总体单位特征的资料转化为反映总体数量特征的资料;⑤ 通过各类统计表、图呈现汇总的结果。

4) 统计分析

统计分析即统计资料的分析,它是对加工汇总的数据资料展开分析研究的过程。这里主要是通过各种统计分析方法,包括综合指标法、回归模型和假设检验等方法,从静态和动态两方面对数据进行分析,以揭示现象发展的规律性和发展趋势。这是统计工作最主要的环节,也是统计学研究的主要内容。

在统计整理与统计分析过程中,所运用的方法包括两大类,即描述统计和推断统计,而描述统计学和推断统计学也常被称为统计学的两大分支。

(1) 描述统计。描述统计是指搜集由试验或调查所获得的资料,对搜集的数据进行登记、审核、整理、归类,在此基础上采用统计分组、计算综合指标的方法得到现象总体的数量特征,并以图表的形式表示经过归纳分析而得到的各种有用的统计信息,以达到认识总体内在数量规律性的目的。描述统计是统计研究过程的基础,为统计推断、统计预测与决策提供事实依据。

(2) 推断统计。推断统计是在对样本数据进行描述的基础上,利用一定的方法根据样本数据去估计或检验总体的数量特征。在实际统计研究过程中,往往是总体规模庞大,又受经费与人力的限制,我们总是无法得到总体数据,相比之下,部分单位的数据更易获取,但是我们关心的是总体的数量特征及其规律性,如企业产品的合格率、城镇居民的消费水平等问题,所以需要运用统计推断的方法来解决。

如果引入描述统计学和推断统计学两分支,统计学探索客观现象总体数量规律性的过程亦可以用图 1.3 来表示。统计研究始于对数据的搜集,终于利用整理过的数据反映客观现象总体的数量特征。在统计研究过程中,如果能直接搜集到反映客观现象的总体数据,利用描述统计学的方法即可达到认识总体内在数量规律性的目的;如果得到的只是样本数据,那么此时还需要引入概率论的理论知识,如大数定律、中心极限定理等,运用推断统计学的方法,利用样本信息推断总体特征,最终达到认识总体内在数量规律的目的。现实中,我们所获取的大部分数据均为样本数据,所以,推断统计方法日益运用广泛。

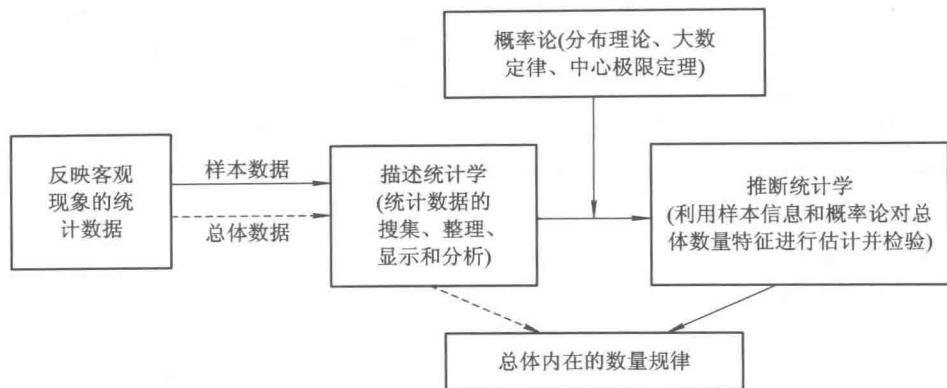


图 1.3 统计学探索客观现象数量规律性的过程

5) 统计资料的管理与开发应用

在经过统计整理和分析环节后，可以得到有条理的统计资料，但是统计资料的得到并不代表统计研究的终结。统计的目的在于认识客观世界总体的数量规律，仅有一次统计的资料，还远远不能很好地发现客观世界存在的数量规律。因此，需要将搜集整理的数据进行积累，建立各种统计数据库，并通过统计信息网络，以各种各样的灵活方式向领导部门以及社会提供资料，以实现统计信息化。如何更好地将统计资料和统计方法应用于各个研究领域，则是应用统计学研究的重要方面之一。

2. 统计的研究方法

统计学是一门方法论学科，它拥有自己的一套方法体系。在统计研究过程中，常用的统计方法有大量观察法、统计分组法、综合指标法、统计模型法和归纳推断法。

1) 大量观察法

所谓大量观察法，就是对所要研究的客观事物的全部或足够多的单位进行观察，以反映总体数量特征和数量规律的方法。大量观察法的数理依据是大数定律，该定律是关于随机事件和随机变量分布规律的描述，其基本含义本书会在第5章具体阐述，这里用一个简单的例子来说明，即掷硬币试验。在该试验中，每掷一次只有两种结果：正面朝上或反面朝上。试验次数越多，两面分别朝上的频率就越接近于50%。通过大量观察，不仅可以掌握客观事物所必需的总体的各种总量，还可以通过个体离差的相互抵消，在一定范围内排除某些偶然因素的影响，从数量上反映总体的本质特征。

在我国统计实践活动中，广泛运用大量观察法组织各种统计调查，例如统计报表制度、普查、重点调查和抽样调查等。

2) 统计分组法

根据统计研究目的和所研究现象总体的特点，选取一定的标志，把所研究的现象总体划分为两个或两个以上组成部分(或组)的统计研究方法称为统计分组法。统计分组法贯穿于统计研究全过程。统计调查离不开分组，在统计整理的过程中，统计分组更是核心技术，而作为统计分析的基本工具——统计指标和指标体系的应用同样需要统计分组作支撑。

现实中，我们所要研究的客观现象是复杂多样的。质与量要做到统一，探索客观事物的数量特征需要分析事物的质，那么，对研究事物总体分成性质不同的组是进行统计加工

整理和深入分析的前提。例如,要研究服务业行业结构及其对国民经济的影响,就必须首先把服务业分为交通运输、旅游、教育、餐饮、批发零售、住宿、科学研究、计算机服务、文教卫生娱乐、公共组织等若干部门,然后分别调查和分析各个部门的增加值、销售收入、固定资产、资金占用、利润、从业人员数及职工工资总额等方面的情况;要研究我国金融产品及其衍生品的发展情况,就应选择金融产品类别作为分组标志进行分组,基于此做详细的统计分析。

3) 综合指标法

所谓综合指标法,就是指利用统计指标和指标体系对现象总体的数量特征和数量关系进行描述和分析的方法。统计研究对象的基本特点之一是数量性,而面对大量客观现象,总体数量特征的研究离不开统计指标和指标体系。所以,综合指标法毫无疑问地成为统计研究的基本方法之一。在统计实践中,存在着数量指标、相对指标、平均指标等综合指标,分别从静态和动态上综合反映现象总体的规模、水平、结构、比例以及依存关系等数量特征和数量规律。

还有一点要说明的是,综合指标法和统计分组法是密切联系且相互依存的。如果没有相应的统计指标来反映现象的规模水平,统计分组就不能揭示现象总体的数量特征;而如果没有科学的统计分组,综合指标就无法清晰划分事物变化的数量界限。所以在研究现象的数量关系时,必须科学分组,合理设置统计指标,二者相辅相成且结合使用才可在统计研究过程中充分发挥它们的重要作用。

4) 统计模型法

统计模型法是根据一定的学科理论和假定条件,采用数学方程模拟现实各类现象数量关系的一种研究方法。利用该方法能够对客观现象和过程中存在的数量关系进行描述,并利用模型对现象的变化进行数量上的评估和预测。统计模型法把客观存在的总体规模、内部结构、各因素的相互关系,通过一定的数学形式有机地结合起来,大大提高了统计的认识能力。

5) 归纳推断法

在统计研究过程中,常常从总体中各单位的特征入手,通过推断得出总体的某种信息。这种从个别到一般,从演绎到归纳,从事实到概括的推理方法,称为归纳推断法。

归纳推断法既可用于对总体数量特征的估计,也可以用于对总体数量特征某些假设的检验。现实中,统计研究面临的多是样本信息,因而归纳推断法被广泛地应用于统计研究的许多领域,例如已知部分员工的平均工资额估计全部员工的平均工资额,建立回归模型需要对模型参数进行估计和检验,利用时间序列进行预测需要对原序列进行估计和检验。因此,归纳推断法是现代统计学最基本且相当重要的方法之一。

1.2 统计学的产生与发展

1.2.1 统计学的发展历程

统计作为一种社会实践活动已具有非常悠久的历史。原始社会,从结绳记事开始,就有了统计的萌芽,在奴隶社会夏朝,有了人口和土地数字的记载,这是我国最早的统计资

料,在西周朝代建立了最早的统计报告制度;在国外的古希腊、罗马时代,开始了对人口和财产的调查,预示着统计实践活动在国外的萌芽。从“统计”一词来说,统计的英文为 Statistics,来自拉丁文,其词根与国家(State)相同,意味着有了国家,就有了统计实践活动的存在。而统计学的产生则要晚许多,直到17世纪中叶,各国政府成立统计部门,统计变成了一项专业工作后,人口、社会、经济统计才得以快速发展,总之,统计学的出现距今仅有300多年的历史。在统计学发展的历史过程中,产生了以下几个主要学派:

1. 政治算术学派

最早的统计学产生于17世纪中叶的英国,代表人物是威廉·配第(William Petty, 1623—1687年)和约翰·格朗特(John Graunt, 1620—1674年)。其中,威廉·配第的代表作《政治算术》(1676年)被誉为是经济学和统计学史上的重要著作,其中,“政治”是指政治经济学,“算术”就是统计方法。书中,威廉·配第运用数字、重量和尺度等工具对英国、法国和荷兰3国的国情国力作了系统的数量对比分析,为统计学的形成和发展奠定了方法论基础。马克思曾经将威廉·配第这一政治经济学之父认为是统计学的创始人。而另一个代表人物约翰·格朗特则以1604年伦敦教会每周一次发表的“死亡公报”为研究资料,于1662年发表了《关于死亡公报的自然和政治观察》的著作。该书通过大量观察的方法发现了许多与人口相关的数量规律,如人口各年龄组的死亡率、性别比例等重要指标的数量规律,较科学地估算了人口总数,并第一次编制了“生命表”,对死亡率与人口寿命作了较为细致的分析。因此,他被认为是人口统计学的创始人。

2. 国势学派

国势学派又称记述学派,产生于17世纪的德国,其代表人物是海尔曼·康令(Hermann Conring, 1606—1681年)和阿亨华尔(Gottfried Achenwall, 1719—1772年)。海尔曼·康令于1660年把国势学从法学、史学和地理学等学科中独立出来;阿亨华尔在哥廷根大学开设“国家学”课程,其主要著作是《近代欧洲各国国势学纲要》,作者认为统计学是关于一国或多国的显著事项的学问,它主要运用对比分析的方法研究了国家组织、领土、人口、资源财富和国情国力,比较了各国实力的强弱,为德国的君主政体服务。因为在外文中,“国势”与“统计”词义相通,所以后来国势学被正式命名为“统计学”。而又因为国势学派只是对国情的文字记述,偏重事物性质的定性分析,并不注重数量的分析。所以,人们也把它叫做记述学派(旧学派或德国学派),并认为国势学派有统计学之名而无统计学之实。

3. 数理统计学派

18世纪,概率理论日益成熟。19世纪中叶,概率论被引进统计学从而形成数理学派,其创始人是比利时的阿道夫·凯特勒(Lambert Adolphe Jacques Quetelet, 1796—1874年),他在其著作《社会物理学》中将古典概率论引入了统计学,使统计学进入一个新的发展阶段。凯特勒认为概率论是适于政治及道德科学中以观察与计数为基础的方法,并运用该方法观察自然现象和社会现象的规律性,认为要促进科学的发展,就必须更多地应用数学。

数理统计学派在20世纪得到快速发展,推断统计学理论成为发展的主导。例如,1907年,英国人戈塞特(1876—1937年)提出了小样本 t 统计量理论,也称为“学生” t 分布,丰富了抽样分布理论,为统计推断奠定了基础;英国的费希尔(R. A. Fisher, 1890—1962年)提

出了极大似然估计量的概念,成为了估计参数的重要方法,他还提出样本相关系数的分布、试验设计和方差分析等方法;英国科学家弗朗西斯·高尔顿(Francis Galton, 1822—1911年)提出了相关与回归思想,并给出计算相关系数的明确公式;英国统计学者 K. 皮尔逊提出拟合优度检验,还对卡方统计量及其极限分布展开了研究;波兰学者奈曼(J. Neyman, 1894—1981年)创立了区间估计理论,并和 E. 皮尔逊发展了假设理论;美国学者瓦尔德提出决策理论和序贯抽样方法;美国化学家威尔科克森(Frank Wilcoxon)发展了一系列非参数统计方法,开辟了统计学的新领域;等等。

4. 社会统计学派

19 世纪后半叶社会统计学派产生,创始人是德国经济学家、统计学家克尼斯(K. G. A. Knies, 1821—1898 年),主要代表人物有厄恩斯特·恩格尔(Christian Lonrenz Ernst Engel, 1841—1896 年)、乔治·冯·梅尔(Georg von Mayr, 1841—1925 年)等人。基于政治算术学派与国势学派的观点,沿着凯特勒的“基本统计理论”向前发展,但在学科性质上认为统计学是一门社会科学,是与数理统计思想完全不同的用来研究社会现象变动原因和规律性的实质性科学。社会统计学派认为统计学是利用大量观察来研究总体的数量特征与规律。这是社会统计学派的“实质性科学”的显著特点,从这点看出,社会统计学派的观点与现代统计学的研究对象方法已经非常类似了。

1.2.2 统计学的发展动向与趋势

现如今,随着社会经济发展和科学技术的进步,统计学得到快速的发展。

一方面,数理统计学是统计学的一个重要分支,而且自 20 世纪以来,数理统计学发展迅速。目前,对于数理统计的研究主要集中在贝叶斯研究、生存分析和高维统计。其中:

(1) 不论是美国统计学会 JASA 期刊还是英国皇家学会 JRSS 期刊等,几乎每期都有关于“贝叶斯统计”^①的论文;

(2) 生存分析则是将事件的结果和出现这一结果所经历的时间结合起来分析的一种统计分析方法,该方法广泛应用于生物医药的生存时间或疗效、工程故障时间、金融保险公司存在时间、寿命等领域,是统计学的一个重要课题;

(3) 高维统计在金融、基因、光谱学分析中应用较多,特别是针对大规模数据集的采集、存储和处理,仅仅通过降维是远远不够的,而模型预测准确性、改进模型估计方法、特殊类型的高维数据检验等问题是当前研究的热点问题。

另一方面,随着科技的不断进步,在物联网、云计算、移动互联网等技术的推动下,数据发生了“大爆炸”,“大数据”油然而生并影响着传统统计学的理论与方法。牛津大学教授维克托·迈尔·舍恩伯格被誉为“大数据时代的预言家”,他和肯尼思·库克耶共同编写了《大数据时代》一书,书中明确指出在大数据时代,传统的数据分析思想应做三大转变:一是转变抽样思想,在大数据时代,样本就是总体,要分析与某事物相关的所有数据,而不是依靠少量数据样本;二是转变数据测量的思想,要乐于接受数据的纷繁芜杂,不再追求精确的数据;三是不再探求难以捉摸的因果关系,转而关注事物的相关关系。这里提到的

^① 贝叶斯统计是指利用先验信息和后验信息来确定概率,基于此进行统计推断的理论和方法。

抽样、样本、总体、因果关系和相关关系等术语,均是统计学的基本概念和核心内容。

第三,在现代统计学理论与方法研究快速发展的背景下,不得不提到与统计学前沿最为相关的统计大会。1853年,第一届世界统计大会在比利时布鲁塞尔顺利举行,此后每两年举办一届,1995年第50届世界统计大会在北京举行,2015年7月26~31日,第60届世界统计大会在巴西里约热内卢召开,此次大会的主题是“数据变革中的可持续发展”。中国统计学年会也是每两年一届,每届一主题,参会人员主要来自统计实务部门、高校和研究中心的统计学专家学者。2014年10月在厦门大学召开的第五届中国统计学年会讨论的主题是“大数据时代下的统计学”,包括大数据统计分析与应用研究,经济统计理论、方法与应用研究,数理统计理论、方法与应用研究等议题。

此外,作为统计学基础课程的学习者,有必要知道与统计学相关的权威级奖项。众所周知,“诺贝尔”奖设立有许多领域的奖项,如文学奖、经济学奖、和平奖等,而在众多的诺贝尔奖项中并没有设立“诺贝尔”统计学奖,由于与统计学最相关的就是计量经济学,因此,在诺贝尔经济学奖项目中,2/3以上的研究成果均与统计和计量经济分析有关。那么,在统计学界是否存在类似的奖项呢?回答是肯定的,它也有自己的“诺贝尔”奖,即考普斯总统奖(COPSS President's Award),该奖项设立于1976年,由美国统计学会、数理统计学会、美东及美西计量协会以及加拿大统计学会等五个统计学会会长组成的委员会提名颁奖,每年只颁给一位40岁以下的统计学最杰出的学者。截至2013年,全球荣获COPSS总统大奖的学者共有33位,其中华裔学者有9位,如表1.1所示,他们中多数人在本科阶段所学的专业要么是统计学,要么是数学或计算机,这也充分说明统计学与数学、计算机学科的紧密联系。

表 1.1 COPSS 总统奖获奖者一览表

时间	获奖人	获奖时工作单位	时间	获奖人	获奖时工作单位
1979年	Peter J. Bickel	加州大学伯克利分校	1998年	Pascal Massart	巴黎第十一大学
1982年	Stephen Fienberg	卡内基梅隆大学	1999年	Larry A. Wasserman	卡内基梅隆大学
1983年	黎子良	斯坦福大学	2000年	范剑青	北卡罗来纳大学教堂山分校
1984年	David V. Hinkley	加州大学圣芭芭拉分校	2001年	孟晓犁	芝加哥大学
1985年	James O. Berger	杜克大学	2002年	刘军	哈佛大学
1986年	Ross L. Prentice	Fred Hutchinson 癌症研究中心	2003年	Andrew Gelman	哥伦比亚大学
1987年	吴建福	威斯康星大学	2004年	Michael A. Newton	威斯康星大学
1988年	Raymond J. Carroll	德克萨斯 A&M 大学	2005年	Mark J. van der Laan	加州大学伯克利分校
1989年	Peter Hall	澳大利亚国立大学	2006年	林希虹	哈佛大学

续表

时间	获奖人	获奖时工作单位	时间	获奖人	获奖时工作单位
1990 年	Peter McCullagh	芝加哥大学	2007 年	Jeffrey S. Rosenthal	多伦多大学
1991 年	Bernard Silverman	布里斯托大学	2008 年	蔡天文	宾夕法尼亚大学
1992 年	Nancy Reid	多伦多大学	2009 年	Rafael Irizarry	约翰霍普金斯大学
1993 年	王永雄	哈佛大学	2010 年	David Dunson	杜克大学
1994 年	David L. Donoho	斯坦福大学	2011 年	Nilanjan Chatterjee	美国国家癌症研究所
1995 年	Iain M. Johnstone	加州大学伯克利分校	2012 年	寇星昌	哈佛大学
1996 年	Robert J. Tibshirani	多伦多大学	2013 年	Marc A. Suchard	加州大学洛杉矶分校
1997 年	Kathryn Roeder	卡内基梅隆大学	合计	33	

资料来源：[http://baike.baidu.com/view/7232416.htm? fr=aladdin](http://baike.baidu.com/view/7232416.htm?fr=aladdin)。

1.2.3 统计学与相关学科的关系

国际知名统计学家劳(C. R. Rao)提到过“统计学基本上是寄生的，靠研究其他领域内的工作而生存”。如图 1.4 所示，只有当统计学与其他学科紧密结合，统计学才能发挥其强大的生命力，产生一系列专业领域的统计学，形成统计学学科体系。

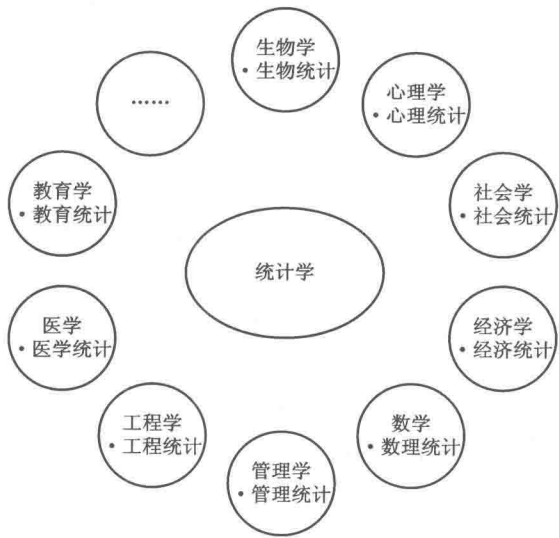


图 1.4 统计学学科体系

本节将重点说明统计学与其关系最为紧密的数学、计量经济学、会计学、计算机科学之间的联系和区别。