



工业和信息化部普通高等教育“十三五”规划教材立项项目

21世纪高等学校规划教材

统计学

Statistics

徐晓岭 王磊 主编

- 吸收国内外优秀教材之长处
- 厚基础重应用，强调实用性
- 商科案例丰富，注重前沿性



中国工信出版集团

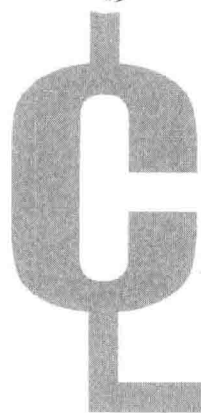


人民邮电出版社
POSTS & TELECOM PRESS



工业和信息化部普通高等教育“十三五”规划教材立项项目

21世纪高等学校规划教材



统计学

Statistics

■ 徐晓岭 王磊 主编



人民邮电出版社

北京

图书在版编目(CIP)数据

统计学 / 徐晓岭, 王磊主编. — 北京: 人民邮电出版社, 2015. 9
21世纪高等学校规划教材
ISBN 978-7-115-39347-0

I. ①统… II. ①徐… ②王… 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(2015)第188879号

内 容 提 要

本书根据经济、金融、管理、文学、法学等专业的特点, 系统介绍了统计学的基本统计思想和基本统计分析方法, 既有较宽的理论基础, 又突出实用性。主要内容包括: 绪论、统计数据的搜集、统计数据的整理、描述性统计、数理统计基础、参数估计、假设检验、非参数假设检验、方差分析、线性回归与相关分析、抽样调查、时间序列分析、指数分析、统计预测与决策、常用统计方法简介、若干统计分析专题简介等。全书内容深入浅出, 难易适度。通过本书的学习, 学生可具备基本的统计思想, 掌握基本的统计方法, 提高运用统计分析方法解决实际问题的能力。

本书不仅可以作为经济、金融、管理、文学、法学等专业的统计学课程的教材, 也可以作为应用统计专业硕士的考试参考用书。

◆ 主 编 徐晓岭 王 磊

责任编辑 张孟玮

执行编辑 税梦玲

责任印制 沈 蓉 彭志环

◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路11号

邮编 100164 电子邮件 315@ptpress.com.cn

网址 <http://www.ptpress.com.cn>

三河市潮河印业有限公司印刷

◆ 开本: 787×1092 1/16

印张: 26.25

字数: 608千字

2015年9月第1版

2015年9月河北第1次印刷

定价: 59.80元

读者服务热线: (010) 81055256 印装质量热线: (010) 81055316

反盗版热线: (010) 81055315

前言

本书是为高等院校非统计学专业的学生编写的一本公共必修课程教材，其适用于经济、金融、管理、文学以及法学等专业学生。由于统计学是与现实生活联系较为紧密且实用性较强的一门课程，所以学生对其掌握程度的好坏，对人才培养的质量有着重要的影响。

学生在初次接触统计学课程时普遍会感到“三难”——基本概念难懂、方法灵活难以掌握、习题难做，以致在学生中流传着著名的呐喊——“我就是不懂！”为了帮助学生能更好地学好这一门课程，深入理解统计学的基本思想，掌握基本的统计分析方法，提高发现问题、分析问题与解决问题的能力，我们系统地梳理了在长期担任统计学课程教学过程的经验体会，注重实际应用背景与实用性，在博采国内外优秀教材之长，突出近二十年研究成果的基础上，经过精心加工整理形成了全书以商科案例为教学目标，以标准化学术写作为教学成果的特色统计学教程。

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 3 个部分。第 1 章至第 3 章为描述统计部分，第 4 章至第 8 章为推断统计部分，第 9 章至第 15 章为常用统计方法部分。本书按一学期 54 学时教授，教师可根据实际教学效果，适当调整教学内容。其中打“*”的第 10、13、14、15 章不建议做课堂讲解，这些内容可以作为学生的阅读提高材料，可为学生毕业论文的撰写提供帮助。

本书由上海对外经贸大学的徐晓岭、王磊任主编，参与编写的老师还有顾蓓青、邵杨、周平、雷平。邵杨编写了第 1 章，周平编写了第 2、11 章以及第 15 章中的国民经济核算部分，顾蓓青编写了第 5、6、10 章，雷平编写了第 3、4 章，王磊编写了第 7、9、12 章以及第 15 章中的服务贸易统计简介部分，徐晓岭编写了第 8、13、14 章以及第 15 章中的金融统计部分，全书由徐晓岭、王磊负责统稿。本书的撰写得到了上海对外经贸大学统计系周晓东主任的关心与帮助，同时还得到了赵飞、沙丹、陈洁、徐冰纨四位老师的关心与帮助。此外，上海对外经贸大学数量经济专业 2013 级硕士研究生张罗赛、李慧杰，2014 级硕士研究生顾蕤，他们也对本书的撰写提供了协助，在此一并深表感谢！

本书的出版得到了上海对外经贸大学——2015 年统计学重点专业建设的资助（编号：R085132023），得到了《商务统计》市级示范性全英语课程建设的资助（编号：QYYKC2013-01）。

由于编者水平有限，书中难免有不妥或谬误之处，衷心欢迎广大读者和专家批评指正。

编 者

2015 年 4 月于上海

目 录

第1章 绪论 1

1.1 统计的含义 1

1.1.1 统计的含义 1

1.1.2 统计的研究对象 2

1.2 统计的应用 3

1.2.1 统计的产生与发展 3

1.2.2 统计的应用领域 5

1.2.3 统计的分科 6

1.3 统计学的基本概念 7

1.3.1 总体与样本 7

1.3.2 统计指标和统计标志 8

1.3.3 参数、统计量与变量 11

1.4 统计数据的收集 12

1.5 统计数据的类型 14

1.5.1 分类数据、次序数据与数值型数据 14

1.5.2 截面数据、时间数据与面板数据 14

1.5.3 统计数据的误差 15

1.6 统计软件简介 16

1.6.1 SPSS 简介 16

1.6.2 SAS 简介 16

1.6.3 R 语言简介 16

1.6.4 Excel 简介 17

习题 1 18

第2章 统计数据的整理 19

2.1 统计数据的预处理 20

2.1.1 统计数据的审核、筛选和排序 20

2.1.2 统计数据的插补 21

2.2 统计分组与变量数列 22

2.2.1 统计分组 22

2.2.2 分布数列 22

2.2.3 变量数列的编制 23

2.3 统计表 25

2.4 统计图 25

2.4.1 饼图 26

2.4.2 条形图 26

2.4.3 茎叶图 27

2.4.4 直方图 27

2.4.5 折线图 28

2.4.6 曲线图 28

习题 2 29

第3章 数据的描述性分析 32

3.1 相对指标 32

3.1.1 相对指标概述 32

3.1.2 几种常见的相对指标 33

3.1.3 应用相对指标应注意的问题 35

3.2 集中趋势指标 36

3.2.1 算术平均数 37

3.2.2 调和平均数 39

3.2.3 几何平均数 41

3.2.4 中位数 42

3.2.5 分位数 44

3.2.6 众数 45

3.3 离散趋势指标 47

3.3.1 极差 48

3.3.2 平均差 48

3.3.3 标准差和方差 49

3.3.4 变异系数 51

3.3.5 数据的标准化 52

3.3.6 是非标志的标准差 52

3.4 偏度和峰度 54

3.4.1 偏态及其测量 54

3.4.2 峰度及其测量 56

习题 3 57

第4章 数理统计基础 60

4.1 数理统计的基本概念 60

4.1.1 简单随机样本 60

4.1.2	统计量与枢轴量	62
4.1.3	次序统计量、经验分布及常用 统计量	62
4.2	常用统计分布	66
4.3	抽样分布	71
4.3.1	正态总体下样本均值和样本方差 的分布	71
4.3.2	样本比例的分布	74
4.3.3	非正态总体统计量的极限 分布	75
习题 4		76

第 5 章 参数估计

5.1	参数的点估计	77
5.1.1	矩法估计	77
5.1.2	极大似然估计	79
5.2	估计量的评选标准	81
5.2.1	无偏性	82
5.2.2	相合估计量 (一致性)	83
5.2.3	有效性	84
5.3	参数的区间估计	85
5.3.1	区间估计的定义	85
5.3.2	正态总体参数的区间估计	87
5.3.3	大样本场合下的区间估计	94
5.3.4	两点分布总体参数的区间 估计	95
5.3.5	泊松分布总体参数的区间 估计	98
习题 5		99

第 6 章 假设检验

6.1	基本概念	101
6.1.1	假设检验的引入	101
6.1.2	判断“假设”的根据	102
6.1.3	如何确定原假设 H_0 和对立 假设 H_1	104
6.1.4	两类错误, 检验的水平与 功效	108
6.1.5	假设检验的程序	110
6.2	正态总体参数的假设检验	110
6.2.1	方差已知情况下正态总体均值 的检验	110

6.2.2	方差未知情况下正态总体均值 的检验	111
6.2.3	均值未知正态总体方差的 检验	112
6.2.4	两正态总体均值的检验	113
6.2.5	两正态总体方差的检验	115
6.3	非正态总体参数的假设检验	117
6.3.1	大样本场合总体均值的 检验	117
6.3.2	大样本场合两个总体均值的 检验	118
6.3.3	大样本场合两点分布总体参数 的检验	118
6.3.4	大样本场合两个两点分布总体 参数的检验	119
6.3.5	大样本场合泊松总体参数的 检验	120
6.3.6	检验的 p 值	120
习题 6		121

第 7 章 非参数统计

7.1	分布的拟合检验	123
7.1.1	拟合优度 χ^2 检验	123
7.1.2	柯尔莫哥洛夫拟合检验	126
7.1.3	柯尔莫哥洛夫—斯米尔诺夫 两样本检验	128
7.1.4	两样本秩和检验	130
7.2	不相关与独立性检验	132
7.2.1	不相关检验	132
7.2.2	列联表和独立性检验	133
7.3	成对比较检验	135
7.3.1	符号检验	135
7.3.2	威尔科克森带符号的等级检验	137
7.4	曼-惠特尼 U 检验	140
7.4.1	曼-惠特尼 U 检验的基本原理	140
7.4.2	小样本 U 检验	140
7.4.3	大样本 U 检验	141
7.5	游程检验	142
7.6	等级相关检验	144

7.6.1 等级相关检验的基本原理	144	预测	187
7.6.2 等级相关系数的应用	145	9.2.3 关于异方差性问题	193
习题 7	146	9.2.4 关于自相关性问题	198
第 8 章 方差分析	150	9.3 相关分析	205
8.1 方差分析的简要介绍	150	9.3.1 相关分析的概念	205
8.1.1 方差分析及其有关术语	150	9.3.2 相关关系类型	206
8.1.2 方差分析的基本思想和原理	151	9.3.3 相关关系的描述与测定	207
8.1.3 方差分析中的基本假定	153	9.3.4 相关系数的显著性检验	210
8.1.4 问题的一般提法	153	习题 9	211
8.2 单因素方差分析	154	第 10 章 抽样调查 *	215
8.2.1 模型	154	10.1 抽样调查概述	215
8.2.2 各水平试验次数相同时的方差分析与 F 检验	155	10.1.1 抽样调查的概念与特点	215
8.2.3 各水平试验次数不一定相同时的方差分析与 F 检验	158	10.1.2 抽样调查的应用领域	216
8.3 双因素方差分析	160	10.2 抽样估计	216
8.3.1 模型	160	10.2.1 全及总体与抽样总体	216
8.3.2 方差分析与 F 检验	160	10.2.2 全及指标与抽样指标	217
8.3.3 重复试验的情况 (无交互作用)	162	10.2.3 重复抽样与不重复抽样	217
8.3.4 交互作用问题	164	10.3 抽样误差	218
习题 8	166	10.3.1 抽样误差的概念	219
第 9 章 线性回归与相关分析	170	10.3.2 抽样平均误差	219
9.1 一元线性回归	170	10.3.3 抽样误差范围与估计的可靠性	220
9.1.1 一元线性回归分析模型	170	10.4 抽样组织形式	222
9.1.2 回归系数 α , β 及 σ^2 的估计	171	10.4.1 类型抽样	222
9.1.3 最小二乘估计的性质	172	10.4.2 机械抽样	224
9.1.4 假设检验与置信区间	174	10.4.3 整群抽样	226
9.1.5 可线性化的一元非线性回归	179	10.4.4 多阶段抽样	227
9.2 多元线性回归	180	习题 10	228
9.2.1 多元线性回归模型及参数的最小二乘估计	181	第 11 章 时间序列分析与预测	230
9.2.2 多元线性回归的假设检验与		11.1 时间序列概述	230
		11.1.1 时间序列的概念和要素	230
		11.1.2 时间序列的种类	230
		11.1.3 时间序列的编制	232
		11.2 时间序列描述性分析	232
		11.2.1 发展水平与平均发展水平	233
		11.2.2 增长量与平均增长量	236

11.2.3	发展速度与增长速度	237
11.2.4	平均发展速度与平均增长 速度	238
11.2.5	发展速度分析应注意的问题	239
11.3	时间序列的构成要素及模型 ...	239
11.3.1	时间序列的构成要素	239
11.3.2	时间序列的分解模型	240
11.4	时间序列的分解分析	240
11.4.1	长期趋势分析	240
11.4.2	季节波动分析	244
11.4.3	循环变动和不规则变动分析	247
习题 11		247
第 12 章	统计指数	250
12.1	统计指数的基本概念	250
12.1.1	统计指数的定义	250
12.1.2	统计指数的分类	251
12.1.3	统计指数的作用	251
12.2	综合指数的编制	252
12.2.1	简单综合法	252
12.2.2	加权综合法	253
12.2.3	拉氏指数和帕氏指数	254
12.2.4	其他加权综合指数	256
12.3	加权平均指数的编制	257
12.4	指数因素分解	258
12.4.1	指数因素分解的含义	258
12.4.2	指数分解的方法	259
12.5	平均指标指数因素分析	262
12.6	统计指数的应用	264
12.6.1	消费者价格指数	264
12.6.2	生产者价格指数	267
12.6.3	股票价格指数	268
12.6.4	空间价格指数	270
12.6.5	消费者满意度指数	271
习题 12		272

第 13 章 * 统计预测与决策 ... 275

13.1	统计预测的一般问题	275
------	-----------------	-----

13.1.1	统计预测的概念和作用 ...	275
13.1.2	统计预测方法的分类及其 选择	276
13.1.3	统计预测的原则和步骤 ...	278
13.2	统计预测的一般方法	279
13.2.1	主观概率法	279
13.2.2	三点预测法	279
13.2.3	指数平滑法	283
13.2.4	季节调整预测法	284
13.2.5	灰色预测法	286
13.3	预测误差	287
13.3.1	预测误差的意义及产生的 原因	287
13.3.2	预测误差的计算	287
13.4	统计决策概述	288
13.4.1	统计决策的概念和种类 ...	288
13.4.2	统计决策的作用和步骤 ...	289
13.4.3	统计决策的公理和原则 ...	290
13.5	风险型决策方法	290
13.5.1	风险型决策概述	290
13.5.2	不同标准的决策方法	291
13.5.3	决策树	291
13.6	贝叶斯决策方法	293
13.6.1	贝叶斯决策概述	293
13.6.2	贝叶斯决策方法的类型和 应用	294

第 14 章 * 常用统计方法简介 ... 297

14.1	定性数据的建模分析	297
14.1.1	对数线性模型基本理论和 方法	297
14.1.2	泊松对数线性模型	299
14.1.3	Logistic 回归基本理论和方法	301
14.2	聚类分析	308
14.2.1	聚类分析基本概念与方法 原理	309
14.2.2	分类统计量与聚类方法 ...	312
14.2.3	类和类的特征	316

14.2.4 系统聚类法(谱系聚类法)	318
14.3 判别分析	325
14.3.1 判别分析的基本思想	325
14.3.2 距离判别法	327
14.4 主成分分析	332
14.4.1 主成分分析的基本思想	332
14.4.2 主成分的推导及性质	338
14.5 因子分析	342
14.5.1 基本概念与方法原理	342
14.5.2 因子分析的数学模型	344
14.5.3 因子载荷阵的估计方法	347
14.5.4 因子旋转	348
14.5.5 因子得分	350
14.5.6 计算步骤及实例	351

第15章* 若干统计分析专题简介

15.1 国民经济核算	356
15.1.1 国民经济核算概述	356
15.1.2 国民经济生产总量核算	358
15.1.3 投入产出核算	360
15.1.4 资金流量核算	362
15.1.5 国际收支核算	364
15.2 金融统计	366
15.2.1 金融统计学基础	366
15.2.2 中央银行统计	366
15.2.3 商业银行统计	368
15.2.4 政策性银行统计	374
15.2.5 证券期货市场统计	375
15.2.6 保险统计	376

15.2.7 对外金融统计	377
15.2.8 外汇市场统计	378
15.2.9 金融监管统计	378
15.3 服务贸易统计简介	379
15.3.1 从服务贸易概念的提出到服务贸易统计制度的建立	379
15.3.2 服务贸易统计商品分类和国别分组	382
15.3.3 服务贸易统计制度	384

附表一 泊松分布函数表	386
-------------	-----

附表二 标准正态分布函数表	388
---------------	-----

附表三 χ^2 分布上侧分位数表	390
-----------------------	-----

附表四 t 分布上侧分位数表	393
------------------	-----

附表五 F 分布上侧分位数表	395
------------------	-----

附表六 柯尔莫哥洛夫检验 临界值表	405
----------------------	-----

附表七 两样本秩和检验 临界值表	407
---------------------	-----

附表八 杜宾—沃森检验上下界表	408
-----------------	-----

第 1 章

绪论

1.1 统计的含义

1.1.1 统计的含义

统计一词起源于国情调查，最早被译为国情学。原始的统计工作即人们收集数据的原始形态已经有几千年的历史，而作为一门科学是从 17 世纪开始的。英语中统计学家和统计员是同一个单词，但统计学并不是直接产生于统计工作的经验总结。每一门科学都有其建立、发展和客观条件，统计科学则是统计工作经验、社会经济理论、计量经济方法融合、提炼、发展而来的一种边缘性学科。

一般来说，统计包括三个含义：统计工作、统计资料和统计科学。三者之间的关系是：统计工作的成果是统计资料，统计资料和统计科学的基础是统计工作，统计科学既是统计工作经验的理论概括，又是指导统计工作的原理、原则和方法。

(1) 统计工作。指利用科学的方法搜集、整理、分析和提供关于社会经济现象数量资料的工作的总称，是统计的基础。它也称统计实践，或统计活动，是在一定统计理论指导下，采用科学的方法，搜集、整理、分析统计资料的一系列活动过程。它是随着人类社会的发展、治国和管理的需要而产生和发展起来的，至今已有四五千年的历史。现实生活中，统计工作作为一种认识社会经济现象总体和自然现象总体的实践过程，一般包括统计设计、统计调查、统计整理和统计分析四个环节。

(2) 统计资料。指通过统计工作取得的、用来反映社会经济现象的数据资料的总称。统计工作所取得的各项数字资料及有关文字资料，一般反映在统计表、统计图、统计手册、统计年鉴、统计资料汇编和统计分析报告中。它也称统计信息，是反映一定社会经济现象总体或自然现象总体的特征或规律的数字资料、文字资料、图表资料及其他相关资料的总称。包括刚刚调查取得的原始资料和经过一定程度整理、加工的次级资料，其形式有：统计表、统计图、统计年鉴、统计公报、统计报告和其他有关统计信息的载体。

(3) 统计科学。也称统计学，是统计工作经验的总结和理论概括，是系统化的知识体系。指研究如何搜集、整理和分析统计资料的理论与方法。统计学也可以看作是应用数学的一个分支，主要通过利用概率论建立数学模型，收集所观察系统的数据，进行量化的分析、总结，进而进行推断和预测，为相关决策提供依据和参考。它被广泛地应用在各门学科之上，从物理和社会科学到人文科学，甚至工商业及政府的情报决策之上。

统计学主要又分为描述统计学和推断统计学。给定一组数据，统计学可以摘要并且描述这组数据，这种用法称作描述统计学。另外，观察者以数据的形态建立起一个用以解释其随机性和不确定性的数学模型，以推论研究中的步骤及总体，这种用法被称作推论统计学。这两种用法都可以被称作应用统计学。另外，也有一个叫作数理统计学的学科专门用来讨论这门科学背后的理论基础。

总体来说，“统计”一词三方面的含义是紧密联系的，统计资料是统计工作的成果，统计工作与统计科学之间是实践与理论的关系。

1.1.2 统计的研究对象

从哲学的意义出发，任何事物都存在质和量两个方面，是质和量的统一。研究一种事物可以从量的方面进行，也可以从质的方面进行，对事物量的方面的研究是在对事物质的方面有所把握的基础上进行的。统计是从量的方面对社会经济现象进行观察研究的，即统计的认识对象是社会经济现象的数量方面。虽然统计是研究社会经济现象数量方面的，但这并不是孤立进行的，而是在质与量的相互联系中研究量的。如果离开了事物质的方面，仅为研究量而研究量，那就不是统计学了。

统计研究事物数量方面的目的，在于通过对事物量的方面的观察和量变规律的研究，逐步把握事物的质和对事物质的方面的认识。因此，统计对社会经济现象数量方面的认识包括量的规模、现象之间的数量联系、现象数量的变化规律、现象质与量互变的数量界限等。而对事物量的这些方面的研究，都不是仅对个别事物观察所能得到的，必须通过对现象的大规模研究才能有效。因此，统计的研究对象具有如下特点。

1. 总体性

统计认识社会经济现象的数量方面必须是对总体现象的认识，而非对个体现象的认识。因为，只有通过总体数量方面的观察，才能发现现象存在的共性和规律性。例如，我们可以通过对一个国家或地区的众多工业企业的研究，了解工业企业的生产能力、生产规模、产品结构和工业品满足社会需要的程度等方面的情况。但如果只对该国或该地区的个别工业企业进行观察，则无论我们的工作做得多么细致，也不可能得到整个工业产品的结构及其满足需要程度的信息，因为它不具备代表性。

统计对社会经济现象的研究要求具有总体性，是基于满足统计研究的目的来考虑的。但强调总体性的要求，并不排斥统计对社会经济个体现象观察的重视。事实上，统计对总体事物的研究是从对个体的观察开始的。例如在人口统计中，如果没有对一个自然人各方面情况的仔细观察和记录，就得不到对人口总体的总人数、性别比例、地区分布、出生率、平均寿命等方面的数量认识。因此，统计对个体现象进行观察的目的，是认识总体的数量特征。

2. 社会性

统计对象的社会性可以从三个方面进行考察：一是统计的认识对象是社会经济现象的数量方面，因而统计本身也就有了社会性。二是统计认识的主体是社会的人，人的阶级性（社会性）决定了认识立场和认识结论上的社会性。三是一切社会经济活动都和人

的利益有关,不同的人群有着不同的利益和利益关系,因此人们相互间的利益分割和利益冲突必将在统计上显示出来。统计为一定的阶级和一定的社会集团服务,古今中外,概莫能外。因此我们说,统计具有社会性。

如前所述,统计学是统计实践活动的理论概括和总结,并反过来指导统计实践活动。因此,统计学的研究对象可以表述为:社会经济总体现象的数量特征及其规律性、统计认识活动过程本身和认识方法。

1.2 统计的应用

1.2.1 统计的产生与发展

统计学的英文 Statistics 最早是源于现代拉丁文 Statisticum Collegium (国会) 以及意大利文 statista (国民或政治家)。德文 Statistik 最早是由 Gottfried Achenwall (1749) 所使用,代表对国家的资料进行分析的学问,也就是“研究国家的科学”。19 世纪统计学在广泛的数据以及资料中探究其意义,并且由 John Sinclair 引进到英语世界。

统计学是一门很古老的科学,一般认为其学理研究始于古希腊的亚里斯多德时代,迄今已有两千三百多年的历史。它起源于研究社会经济问题,在两千多年的发展过程中,统计学至少经历了“城邦政情”“政治算数”和“统计分析科学”三个发展阶段。所谓“数理统计”并非独立于统计学的新学科,确切地说它是统计学在第三个发展阶段所形成的所有收集和分析数据的新方法的一个综合性名词。概率论是数理统计方法的理论基础,但是它不属于统计学的范畴,而属于数学的范畴。

统计学的发展可划分为以下 3 个阶段。

(1) “城邦政情”阶段。“城邦政情”阶段始于古希腊的亚里斯多德撰写“城邦政情”或“城邦纪要”。“城邦政情”式的统计研究延续了一两千年,直至 17 世纪中叶才逐渐被“政治算数”这个名词所替代,并且很快演化为“统计学”(Statistics)。统计学依然保留了城邦(state)这个词根。

(2) “政治算数”阶段。与“城邦政情”阶段没有很明显的分界点,本质的差别很小。“政治算数”的特点是统计方法与数学计算和推理方法开始结合;分析社会经济问题更加注重运用定量分析方法。1690 年,英国威廉·配弟出版《政治算数》一书作为这个阶段的起始标志。威廉·配弟用数字、重量和尺度将社会经济现象数量化的方法是近代统计学的重要特征。因此,威廉·配弟的《政治算数》被后来的学者评价为近代统计学的来源,他本人也被评价为近代统计学之父。配弟在书中使用的数字有三类:第一类是对社会经济现象进行统计调查和经验观察得到的数字。第二类是运用某种数学方法推算出来的数字。第三类是为了进行理论性推理而采用的例示性的数字。

(3) “统计分析科学”阶段。在“政治算数”阶段出现的统计与数学的结合趋势逐渐发展形成了“统计分析科学”。19 世纪末,欧洲大学开设的“国情纪要”或“政治算数”等课程名称逐渐消失,代之而起的是“统计分析科学”课程,当时的“统计分析科学”课程的内容仍然是分析研究社会经济问题。“统计分析科学”课程的出现是现代统

计发展阶段的开端。1908年,“学生”氏(William Slessy Gosset 的笔名)发表了关于 t 分布的论文,创立了小样本代替大样本的方法,开创了统计学的新纪元。现代统计学的代表人物首推比利时统计学家奎特莱,他将统计分析科学广泛应用于社会科学、自然科学和工程技术科学领域。现代统计学的理论基础概率论始于研究赌博的机遇问题,大约为 1477 年。数学家为了解释支配机遇的一般法则进行了长期的研究,逐渐形成了概率论理论框架。在概率论进一步发展的基础上,到 19 世纪初,数学家们逐渐建立了观察误差理论、正态分布理论和最小平方法则。于是,现代统计方法便有了比较坚实的理论基础。

20 世纪初以来,科学技术迅猛发展,社会发生了巨大变化,统计学进入了快速发展时期。归纳起来,有以下几个方面。

(1) 由记述统计向推断统计发展。记述统计是对所搜集的大量数据资料进行加工整理、综合概括,通过图示、列表和数字,编制次数分布表、绘制直方图、计算各种特征数等,对资料进行分析和描述。而推断统计则是在搜集、整理观测的样本数据基础上,对有关总体作出推断。其特点是根据带随机性的观测样本数据以及问题的条件和假定(模型),而对未知事物作出的、以概率形式表述的推断。西方国家所指的科学统计方法,主要就是推断统计。

(2) 由社会、经济统计向多分支学科发展。在 20 世纪以前,统计学的领域主要是人口统计、生命统计、社会统计和经济统计。随着社会、经济和科学技术的发展,统计的范畴已覆盖了社会生活的领域,成为通用的方法论科学。它被广泛用于研究社会和自然界的各个方面,并发展成为有着许多分支学科的科学。

(3) 统计预测和决策科学的发展。传统的统计是对已经发生和正在发生的事物进行统计,提供统计资料和数据。20 世纪 30 年代以来,特别是第二次世界大战以来,由于经济、社会、军事等方面的客观需要,统计预测和统计决策科学有了进一步发展。

(4) 信息论、控制论、系统论与统计学的相互渗透和结合,使统计科学进一步得到发展和日趋完善。信息论、控制论、系统论在许多基本概念、基本思想、基本方法等方面有着共同之处,分别从不同角度、侧面提出了解决共同问题的方法和原则。三论的创立和发展,改变了世界的科学图景和科学家的思维方式,也使统计科学和统计工作出现了新的发展趋势。

(5) 计算技术和一系列新技术、新方法在统计领域不断得到开发和应用。伴随着计算机技术的不断发展,使统计数据的搜集、处理、分析、存贮、传递、印制等过程日益现代化,提高了统计工作的效能。计算机技术的发展,日益扩大了传统的和先进的统计技术的应用领域,促使统计科学和统计工作发生了革命性的变化。而今,计算机科学已经成为统计科学不可分割的组成部分。随着科学技术的发展,统计理论和实践深度与广度方面也在不断发展。

(6) 统计在现代化管理和社会生活中的地位日益重要。英国统计学家哈斯利特说:“统计方法在生活 and 习惯中的应用是这样普遍,应当重视统计的作用。”后来,甚至有的科学还称之为“统计时代”。

统计学产生于应用,在应用过程中发展壮大。随着经济社会的发展、各学科相互融合趋势的发展和计算机技术的迅速发展,统计学的应用领域、统计理论与分析方法也将

不断发展。

1.2.2 统计的应用领域

随着社会和科学技术的蓬勃发展，统计学与其他学科的相互渗透与影响更加广泛深入，应用于各个领域。在商业以及工业中，统计被用来了解与测量系统变异性、程序控制、对决策提供数据支持。在第一产业方面，可运用统计计算出各种农产品的需求情况及价格分布，从而指导生产；在生产行业中，统计学可以运用在产品开发、营销、财务管理等方面，从而提高企业的营运能力；在服务行业中，例如在金融行业中，运用统计技术对各种交易资料加以分类、整理，从而得到如客户贡献度、客户偏好、存款变动趋势、产品分析、行业发展等数据，进而为管理层提供决策依据等。

(1) 企业发展战略。发展战略是一个企业长远的发展方向。控制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变化趋势，另一方面还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。所有这些都需要统计提供可靠的数据，利用统计方法进行科学的数据分析和预测。

(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。质量管理中离不开统计的应用。在一些知名的跨国公司，6 σ 准则已经成为一种重要的管理理念。质量控制已经成为统计学在生产领域的一项重要应用，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。要了解市场就需要进行广泛的市场统计调查，取得所需信息，并对这些信息进行统计分析，以便作为生产和营销的依据。

(4) 财务分析。上市公司的财务数据是股民投资的重要参考依据。一些投资咨询公司主要是根据上市公司提供的财务和统计数据进行分析，为股民提供参考。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方法。

(5) 经济预测。企业要对未来市场状况进行预测。例如，对产品的市场潜力进行预测，及时调整生产计划。这就需要利用统计方法进行收集、整理和分析数据。

(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，并作为企业制度工资计划、奖惩程度的依据。

统计学的应用非常广泛，除了管理科学之外，自然科学和社会科学等领域也需要统计学为其处理大量的数据问题。从某种意义上说，统计学是与数学一样的一种数据分析工具。表 1.1.1 列举了一些统计学的应用领域，由此可见统计学的应用之广泛。

表 1.1.1 统计学的应用领域

actuarial work (精算)	hydrology (水文学)
agriculture (农业)	industry (工业)
animal science (动物学)	linguistics (语言学)
anthropology (人类学)	literature (文学)
archaeology (考古学)	manpower planning (劳动力计划)
auditing (审计学)	management science (管理科学)
crystallography (晶体学)	marketing (市场营销学)

续表

demography (人口统计学)	medical diagnosis (医学诊断)
dentistry (牙医学)	meteorology (气象学)
ecology (生态学)	military science (军事科学)
econometrics (经济计量学)	nuclear material safeguards (核材料安全管理)
education (教育学)	ophthalmology (眼科学)
election forecasting and projection (选举预测和策划)	pharmaceutics (制药学)
engineering (工程)	physics (物理学)
epidemiology (流行病学)	political science (政治学)
finance (金融)	psychology (心理学)
fisheries research (水产渔业研究)	psychophysics (心理物理学)
gambling (赌博)	quality control (质量控制)
genetics (遗传学)	religious studies (宗教研究)
geography (地理学)	sociology (社会学)
geology (地质学)	survey sampling (调查抽样)
historical research (历史研究)	taxonomy (分类学)
human genetics (人类遗传学)	weather modification (气象改善)

1.2.3 统计的分科

任何一门科学，随着人们对它的研究逐步深入，总是在不断地发展与进步。由于研究人员观察的角度不同，研究的重点不同，必然会出现各个相互联系而又有所区别的分支。统计学也不例外，大致有以下两种分类。

1. 描述统计学和推论统计学

统计学分为描述统计学和推论统计学，一方面是反映统计发展的两个阶段，另一方面各有不同的侧重。描述统计学是研究如何对客观现象进行数学的计量、概括和表示的方法。有些客观现象的数字计量是比较简单的，如人的身高、体重等；而在某些领域如社会经济方面就比较复杂，例如要计量居民生活费的变动或不同国家人民生活水平的比较，就要涉及很多方面。因此需要确定一些反映现象数量特征的范畴，即统计指标。研究某一个问题，要选择恰当的统计指标。然而一个指标往往只能说明某一方面的问题，要想用来全面评价社会经济现象的发展状况常常是困难的。为了比较全面系统地认识社会经济现象，就需要用若干个相互联系的指标来反映所研究问题的各个侧面，这些相互联系的统计指标就构成了统计指标体系。有了统计指标体系后就必须搜集数据，搜集数据的方法也要根据不同的研究对象和研究目的来确定。在一些自然科学中通常是根据实验观察来获得数据，而在社会科学中往往通过调查和访问来取得。搜集来的数据开始总是许许多多杂乱无章的原始资料，难以直接看出什么问题，于是需要对数据进行简缩、加工，按各种项目分门别类加以整理，综合成一些图、表醒目地表达出来。这一系列的内容都是描述统计所要研究的。

推论统计学也称归纳统计学。在 20 世纪之前，统计学基本上处于描述阶段，戈赛特、费希尔、奈曼等人对统计学的发展，使推论统计学的理论不断丰富并成为统计学研

究的主流。它是研究如何根据部分数据去推论总体的情况，因为对客观现象搜集数据有时不可能对全部单位去作调查，即使有时可以调查工作量也太太。例如要研究我国人口的年龄结构、婚姻状况、文化水平以及死亡率等情况，用普查的方法工作量极大，如果没有足够数量且训练有素的工作人员，难免会出现误差，其结果也就不理想。故通常抽取部分样本进行研究，从而对总体作出推断。当然，样本只是总体的一部分，包含的信息不完备，而且抽样是随机进行的，必然会出现误差，对于推断的结论是否可靠要冒一定风险。但是推论统计根据概率论的原理可以使归纳推断所产生的不确定性得到度量。为了使样本对总体作出更可靠的推断，推论统计也要研究如何对抽样进行设计等内容。

2. 理论统计学和应用统计学

理论统计学是指统计学的数学原理，扎根于纯数学的一个领域——概率论。从广义来说，统计理论可以认为是包括概率论的，此外还包括一些并不属于传统概率论的内容，如随机化原则的理论、各种估计的原理、假设检验的原理以及一般决策的原理，可以看成是概率论公理的扩增。在统计实践中常常会遇到一些新问题，使原有的统计方法不适应，需要统计学家针对新问题去建立一个与实际情况相符合的统计模型，创造新的统计方法去分析。前面已经提到统计学是应用性很强的一门学科，因此将统计学的基本原理应用于各个领域就形成了各种各样的应用统计学。它包括一整套统计分析方法，有的是一般性的统计方法，适用于各个领域，如参数估计、假设检验、方差分析、相关与回归等。有的则是某一专业领域中特有的分析方法，例如经济统计学中的指数分析法等。近几十年来，由于统计研究的范围越来越广，一些科学实验也日趋复杂，统计方法也相应地复杂化和专门化，在应用统计方法中必须对因模型和实际情况不一致而引起的各种误差的性质和大小做出判断，或提出改进的措施。由于统计的工具更加专门化了，也就缺少通用性，一个统计学家要熟悉所有的专门工具已不可能。为适应这种发展的需要，要求既熟悉统计知识又熟悉某一领域业务的应用统计学家就应运而生，同时也产生了相应的应用统计学。这类统计学的特点不着重于统计数学原理的推导，而是侧重于阐明统计的思想，并将理论统计学的结论作为工具应用于各个具体领域。

1.3 统计学的基本概念

1.3.1 总体与样本

1. 总体

总体是包含研究的全部个体（数据）的集合，通常由所研究的一些个体组成，如中国全部人口构成的集合、由多个企业构成的集合、由某品牌移动电话的所有客户构成的集合等。组成总体的每一个元素称为个体，在由中国全部人口构成的总体中，每一个人就是一个个体；由多个企业构成的总体中，每一个企业就是一个个体；由某品牌移动电话的所有客户构成的总体中，每一个客户就是一个个体。

总体的确定有时比较容易,有时却比较困难。如,对一批产品的零件的使用寿命进行检验,这批零件构成的集合就是总体,每一个零件就是一个个体,总体的范围比较清楚。再如,对于新推出的一种新型电子产品,要想知道消费者是否能接受,首先必须弄清哪些人是消费对象,也就是确定构成该产品的消费者这一总体。但是事实上,我们很难确定哪些消费者想购买该产品,总体范围很难确定。当总体范围很难确定时,可以根据研究的目的来定义总体。

总体根据其所包含的单位数目是否可数分为有限总体和无限总体。有限总体是指总体的范围能够明确确定,而且元素的数目是有限可数的。例如,中国全部人口构成的总体就是有限总体,一批待检验的零部件也是有限总体。无限总体是指总体所包括的元素是无限的,不可数的。例如,在科学实验中,每一个实验数据可以看作一个总体的一个元素,而实验则可以无限地进行下去,因此由实验数据构成的总体就是一个无限总体。

总体分为有限总体和无限总体主要是为了判别在抽样中每次抽取是否独立。对于无限总体,每次抽取一个单位,并不影响下一次的抽样结果,因此每次抽取可以看作是独立的。对于有限总体,抽取一个单位后,总体元素就会减少一个,前一次的抽样结果往往会影响到第二次的抽样结果,因此每次抽取是不独立的。这些因素会影响到抽样推断的结果。

最后,再对总体的概念作进一步的说明。如要检验一批灯泡的寿命,这批灯泡构成的集合就是总体。在统计问题中,我们只是关心每一个灯泡的寿命,而不是灯泡本身,所以也可以把这批灯泡的寿命集合作为总体,这个总体是由一些实数构成的集合。一般而言,有限总体就是有限个实数的集合。如果不是针对一批特定的灯泡,而是全面地考察某企业生产的灯泡寿命,可能的寿命是多少呢?答案是 $[0, +\infty)$ 这样一个区间。或者这样看这个问题:随机从该企业生产的灯泡中拿出一个,问这个灯泡的可能寿命是多少。答案只能是“非负实数”,当然这个“非负实数”在实验检验前是未知的。这时称该企业生产的灯泡寿命总体是取值于 $[0, +\infty)$ 区间上的一个随机变量,这是一个无限总体。在统计推断中通常是针对无限总体的,因而通常把总体看作随机变量。通常情况下,统计上的总体是一组观测数据,而不是一群人或一些物品的集合。

2. 样本

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量。抽样的目的是根据样本提供的信息推断总体的特征。例如,从一批灯泡中随机抽取100个,这100个灯泡就构成了一个样本,然后根据这100个灯泡的平均使用寿命去推断这批灯泡的平均使用寿命。

1.3.2 统计指标和统计标志

1. 统计指标

(1) 统计指标的概念和构成要素

统计指标的含义有两种理解与使用方法。一种是:指标是反映总体现象数量特征的概念(或名称),例如国内生产总值、居民消费水平等。另一种是:指标是说明总体数