

Statistics Theories and Way

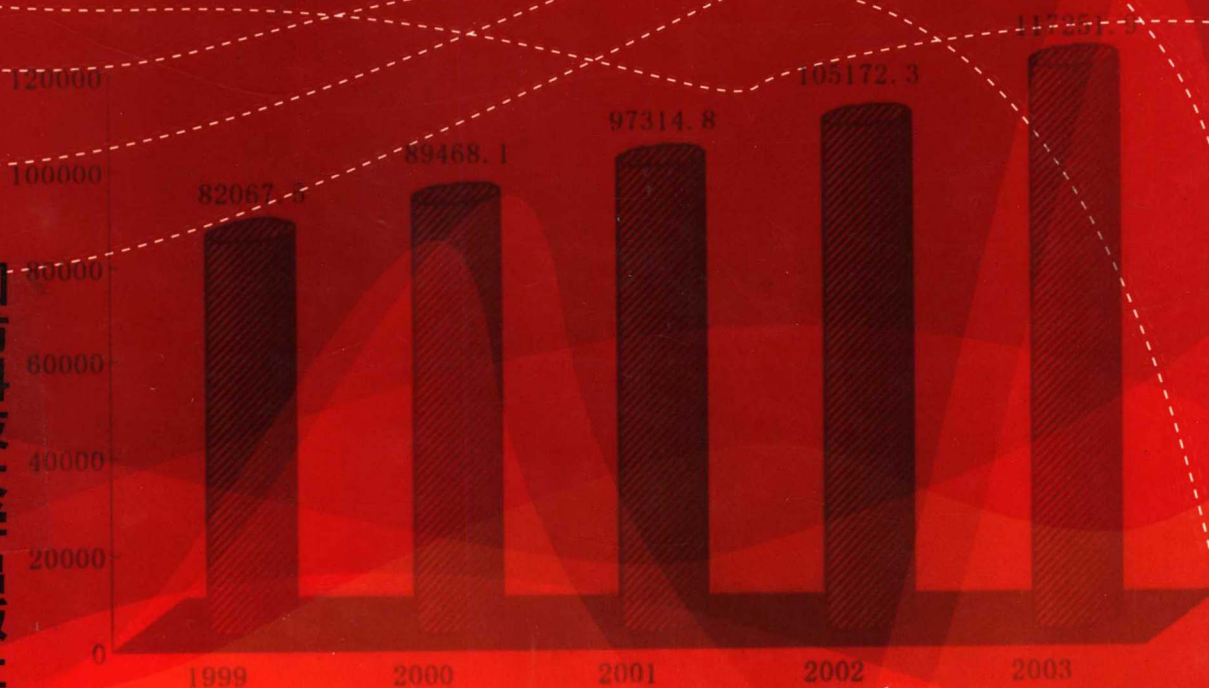
统计学

理论与方法

肖彦花 编著

日本	3956	-3018	938	657	-1079	915	-257	1125	-33	-634	461
德国	6150	-4928	1222	1060	-1505	1033	-1093	466	-2	-771	-20
英国	2799	-3500	-702	1304	-1077	1897	-1561	-267	16	89	-6
法国	3056	-2966	90	867	-690	808	-679	257	-2	-334	-40
意大利	2537	-2371	165	603	-635	433	-579	-67	7	112	32
加拿大	2641	-2272	368	372	-425	202	-377	149	31	-117	-2
澳大利亚	651	-705	-54	179	-181	82	-199	-174	4	172	1
捷克	385	-407	-22	71	-64	22	-60	-45	0	112	66
波兰	467	-540	-72	100	-92	20	-38	-50	0	70	6
匈牙利	348	-369	-21	78	-72	12	-28	-26	2		
罗马尼亚	139	-165	-26	23	-23	4	-9	-15	1		
印度	527	-652	-124	249	-187	23	-62	47	35	81	
印度尼西亚	592	-357	235	67	-171	13	-84	78	0	-12	
菲律宾	344	-340	4	31	-43	79	-34	42	0	-27	
泰国	668	-570	98	153	-167	34	-47	77		-27	
新加坡	1284	-1098	185	297	-273	144	-155	187		-157	
巴基斯坦	98	-104	-6	24	-22	1	-34	39		-48	
中国	1232	-148	102	181	-356	569	-63	61			

国防科技大学出版社



统计学：理论与方法

肖彦花 编著

国防科技大学出版社
湖南 长沙

图书在版编目(CIP)数据

统计学:理论与方法/肖彦花编著. —长沙:国防科技大学出版社,2005.8

ISBN 7-81099-221-X

I. 统… II. 肖… III. 统计学—高等学校—教材 IV. C8

中国版本图书馆 CIP 数据核字(2005)第 084244 号

国防科技大学出版社出版发行

电话:(0731)4572640 邮政编码:410073

E-mail: gfkdcbs@ public. cs. hn. cn

责任编辑:文 慧 责任校对:唐卫葳

新华书店总店北京发行所经销

国防科技大学印刷厂印装

*

开本:787×1092 1/16 印张:31 字数:659 千

2005 年 8 月第 1 版第 1 次印刷 印数:1-3000 册

ISBN 7-81099-221-X/C·4

定价:42.50 元



前言

QIANYAN QIANYAN

欢迎使用这本《统计学:理论与方法》。本书是对统计学理论与方法如何在现代社会经济生活中发挥作用这一课题进行多年理解的结果,可作为大学经济及管理专业本科生、MBA 和相关专业研究生的专业教材,也可作政府部门、企事业单位从事统计和经济分析工作人员的参考用书。

统计学誉称渊远流长,在现代社会经济生活中的应用也十分广泛。在中国这样的地域大国,统计学一直是政府部门执行公共职能的应用学科,其次才是各种组织、个人的认识工具,因而在某种程度上它是一门官学或显学。这样的学科,其体系、构架甚至行文风格都是成熟的和讲究章法的,一般读者大多只能适应它而无法触动它。考虑到这些,我的工作主要定位在系统地、恰当地反映既存的统计学科理论和方法。本书展示了现代统计学专业领域的全貌,并追踪学术界最新观点与业界前沿动态,在基本理论和方法上作了系统的辨析。仅在某些极为有限的领域,我才表达了属于自己的观点。

其次考虑到本书将作为教科书来使用,我的大学同事们提醒我,它不仅需要有较成熟的学科生命基础,而且要有符合认识规律、教学规律的逻辑安排和表述。我想这些也是我必须遵从的。因此我要做的是尽量在由浅入深、循序渐进、信息充分等方面有较好的

把握。本书注重统计学在经济和管理工作中的应用,其内容取舍均服从此目标。由于现代统计不可避免地涉及概率论、统计推断、置信区间、方差分析、回归分析等较高深晦涩的内容,本书在叙述方法上作了一些探讨,把定义、定理、推导和验证有机地结合起来,配以丰富的案例习题和图表,辅以全新动态的版式。我想本书虽然内容深厚,但只要详略有致,处置得当,用于教和学仍然是游刃有余的。

另一种考虑是节约导向。“一切节约都是时间的节约”(马克思)。本书的要旨是为读者节约时间。与一些教科书“兼顾不同程度读者的需要”不同,本书强调它的针对性和专业性,服务群体是大学经济及管理专业本科生、MBA 和相关专业研究生,目的是帮助他们顺利地完课程学习。因此本书力求简明易懂、内容系统和实用,着重阐明统计学的基本理论和方法,并在每章后面附有小结、思考练习题及参考答案。由于我们所处的计算机及其网络环境,本书还结合了 Excel 软件的应用教学。目前许多软件如 SAS、SPSS、Statistica、Minitab、Eviews 等均有很强的统计功能,但 Excel 更普及,为大多数人所熟悉,其统计功能已能满足计算和分析的需要,而且 Excel 能够与 Office 中的其他软件很好地结合,比如与 PowerPoint 结合制作教学幻灯片等。相信这种安排会给读者带来便利。

很幸运能够与我的读者们在现代统计理论及方法的迷人旅途上,作一次愉快的旅行。

作 者

2005 年 8 月

第1章 导 论

1.1 统计学的起源和发展	(2)
1.1.1 统计的起源	(2)
1.1.2 统计学的3个发展阶段	(3)
1.1.3 中国统计学的现状	(5)
1.2 统计学的研究对象和性质	(6)
1.2.1 统计的涵义	(6)
1.2.2 对未来统计学的展望	(8)
1.2.3 统计学的研究对象及其特点	(9)
1.2.4 统计学的性质	(10)
1.3 统计学的研究方法和研究程序	(11)
1.3.1 统计研究的基本方法	(11)
1.3.2 统计研究程序	(13)
1.4 统计学的分科	(14)
1.4.1 描述统计学和推断统计学	(15)
1.4.2 理论统计学和应用统计学	(15)
1.5 统计的职能	(16)
1.6 统计学的基本范畴	(17)
1.6.1 统计总体和总体单位	(17)
1.6.2 标志和标志表现	(19)
1.6.3 统计指标和指标体系	(20)
1.6.4 变异、变量与变量值	(24)
小结	(25)
思考与练习	(26)

第2章 统计设计

2.1 统计设计的意义和种类	(30)
2.1.1 统计设计的意义	(30)
2.1.2 统计设计的种类	(30)
2.1.3 中国统计设计的目标	(31)
2.2 统计设计的原则和内容	(31)
2.2.1 统计设计的原则	(31)

2.2.2 统计设计的内容	(32)
2.3 统计指标和指标体系的设计	(34)
2.3.1 统计指标和统计指标体系设计的原则	(34)
2.3.2 统计指标和统计指标体系设计的内容	(35)
2.3.3 统计指标的评价标准	(37)
2.3.4 统计指标体系的特点及判别标准	(39)
2.4 实验的统计设计	(41)
2.4.1 实验设计的原则	(41)
2.4.2 实验设计的要素	(47)
2.4.3 正交试验设计	(48)
小结	(51)
思考与练习	(51)

第3章 统计数据的搜集

3.1 数据的计量与分类	(54)
3.1.1 数据的计量尺度	(54)
3.1.2 数据的分类	(56)
3.2 统计调查的意义和种类	(59)
3.2.1 统计调查的意义	(59)
3.2.2 统计调查的种类	(60)
3.3 数据搜集的方法	(62)
3.3.1 询问调查法	(62)
3.3.2 直接观察法	(64)
3.3.3 报告法	(64)
3.3.4 实验调查法	(64)
3.3.5 卫星遥感法	(65)
3.4 调查方案设计	(65)
3.4.1 调查目的	(65)
3.4.2 调查对象和调查单位	(65)
3.4.3 调查项目和调查表	(67)
3.4.4 调查方法	(69)
3.4.5 调查时间和调查期限	(69)
3.4.6 调查工作的组织实施计划	(69)
3.5 统计调查的组织方式	(70)
3.5.1 统计报表	(70)

3.5.2 专门调查	(74)
3.6 统计数据的质量	(80)
3.6.1 统计数据误差	(80)
3.6.2 统计数据的质量要求	(81)
小结	(81)
思考与练习	(81)

第4章 统计数据的整理与显示

4.1 数据整理的意义及程序	(84)
4.1.1 数据整理的意义	(84)
4.1.2 数据整理的程序	(84)
4.2 统计分组	(86)
4.2.1 统计分组的意义	(86)
4.2.2 统计分组的种类	(88)
4.2.3 分组标志的选择	(89)
4.2.4 统计分组的方法	(90)
4.2.5 统计分组体系	(93)
4.3 频数分布	(94)
4.3.1 频数分布的概念	(94)
4.3.2 变量数列	(96)
4.3.3 累计频数与累计频率	(100)
4.3.4 频数分布的类型	(103)
4.4 数据整理的组织和技术	(105)
4.4.1 数据整理的组织形式	(105)
4.4.2 数据整理的技术	(106)
4.5 数据显示	(108)
4.5.1 统计表	(109)
4.5.2 统计图	(114)
小结	(123)
思考与练习	(124)

第5章 综合指标

5.1 总量指标	(128)
5.1.1 总量指标的意义	(128)
5.1.2 总量指标的种类	(128)

5.1.3 总量指标的计量单位	(129)
5.1.4 总量指标的计算	(132)
5.2 相对指标	(136)
5.2.1 相对指标的意义	(136)
5.2.2 相对指标的表现形式	(137)
5.2.3 相对指标的计算	(139)
5.2.4 相对指标的应用条件	(146)
5.2.5 正确运用相对指标应注意的问题	(147)
5.3 平均指标	(148)
5.3.1 平均指标的意义和种类	(148)
5.3.2 算术平均数	(150)
5.3.3 调和平均数	(156)
5.3.4 几何平均数	(158)
5.3.5 众数与中位数	(160)
5.3.6 平均指标的应用	(160)
5.4 变异指标	(168)
5.4.1 变异指标的意义	(168)
5.4.2 异从比率、极差、分位差和平均差	(169)
5.4.3 偏度与峰度	(180)
小结	(183)
思考与练习	(184)

第6章 抽样推断

6.1 抽样推断概述	(188)
6.1.1 抽样推断的意义及特点	(188)
6.1.2 抽样推断的基本概念	(190)
6.1.3 抽样推断的组织形式	(195)
6.1.4 抽样推断的理论依据	(197)
6.2 抽样误差	(199)
6.2.1 实际抽样误差	(199)
6.2.2 抽样平均误差	(200)
6.2.3 抽样极限误差	(202)
6.3 参数估计的基本方法	(203)
6.3.1 点估计	(203)
6.3.2 区间估计	(206)

6.4 单一总体均值和总体比率的区间估计	(207)
6.4.1 单一总体均值的区间估计	(207)
6.4.2 单一总体比率的区间估计	(210)
6.5 样本容量的确定	(212)
6.5.1 确定样本容量的必要性	(212)
6.5.2 影响样本容量的因素	(213)
6.5.3 确定样本容量的方法	(214)
6.5.4 计算时应注意的问题	(216)
6.6 两个总体均值和两个总体比率之差的区间估计	(216)
6.6.1 两个总体均值之差的区间估计	(216)
6.6.2 两个总体比率之差的区间估计	(220)
6.7 单一正态总体方差及两正态总体方差比的区间估计	(221)
6.7.1 单一正态总体方差的区间估计	(221)
6.7.2 两正态总体方差比的区间估计	(222)
小结	(224)
思考与练习	(224)

第7章 假设检验

7.1 假设检验概述	(228)
7.1.1 假设检验的概念	(228)
7.1.2 假设检验的检验法则	(228)
7.1.3 假设检验中的两类错误	(229)
7.1.4 假设检验的基本思想	(232)
7.1.5 假设检验的步骤	(232)
7.2 总体均值的假设检验	(235)
7.2.1 单一总体均值的假设检验	(235)
7.2.2 两总体均值之差的假设检验	(239)
7.3 总体比率的假设检验	(244)
7.3.1 单一总体比率的假设检验	(244)
7.3.2 两总体比率之差的假设检验	(245)
7.4 总体方差的假设检验	(248)
7.4.1 单一正态总体方差的假设检验	(248)
7.4.2 两不同正态总体方差的假设检验	(251)
7.5 假设检验中的其他问题	(253)
7.5.1 区间估计与假设检验的关系	(253)

7.5.2 假设检验中的 P 值	(255)
小结	(256)
思考与练习	(256)

第 8 章 相关与回归分析

8.1 相关分析概述	(260)
8.1.1 函数关系与相关关系	(260)
8.1.2 相关关系的种类	(261)
8.1.3 相关分析的概念及内容	(262)
8.2 相关分析方法	(263)
8.2.1 相关表与相关图	(263)
8.2.2 相关系数	(265)
8.3 回归分析概述	(270)
8.3.1 回归分析的概念及内容	(270)
8.3.2 回归分析与相关分析的关系	(271)
8.3.3 相关分析与回归分析应注意的事项	(272)
8.4 线性回归分析	(273)
8.4.1 简单线性回归分析	(273)
8.4.2 多元线性回归分析	(281)
8.5 非线性回归分析	(288)
8.5.1 非线性回归分析概述	(288)
8.5.2 非线性回归方程的测定	(289)
小结	(296)
思考与练习	(297)

第 9 章 时间数列分析与预测

9.1 时间数列的编制	(302)
9.1.1 时间数列的意义和种类	(302)
9.1.2 编制时间数列的基本原则	(307)
9.2 时间数列的分析指标	(308)
9.2.1 时间数列的水平分析指标	(308)
9.2.2 时间数列的速度分析指标	(314)
9.2.3 应用时间数列分析指标时应注意的问题	(320)
9.3 长期趋势的测定	(320)
9.3.1 时间数列的构成要素与模型	(320)

9.3.2 长期趋势的测定	(321)
9.3.3 趋势线的选择	(330)
9.4 季节变动的测算	(331)
9.4.1 季节变动及其测算目的	(331)
9.4.2 季节变动分析的原理与方法	(331)
9.4.3 季节变动的调整	(337)
9.5 循环变动的测算	(340)
9.5.1 循环变动及其测算目的	(340)
9.5.2 循环变动的测算方法	(340)
9.6 不规则变动的测定	(344)
9.7 时间数列的预测方法	(345)
9.7.1 移动平均法	(345)
9.7.2 指数平滑法	(347)
9.7.3 趋势外推法	(348)
9.7.4 季节指数法	(350)
小结	(352)
思考与练习	(353)

第 10 章 统计指数分析

10.1 统计指数概述	(358)
10.1.1 统计指数的概念、性质和作用	(358)
10.1.2 统计指数的种类	(361)
10.2 总指数的编制	(363)
10.2.1 总指数编制的基本问题	(363)
10.2.2 总指数的编制方法	(366)
10.3 指数体系与因素分析	(372)
10.3.1 指数体系	(372)
10.3.2 因素分析法	(373)
10.4 指数数列	(381)
10.4.1 定基指数与环比指数	(381)
10.4.2 不变权数指数与可变权数指数	(382)
10.5 几种常用的经济指数	(383)
10.5.1 居民消费、商品零售价格指数	(383)
10.5.2 工业生产指数	(389)
10.5.3 农副产品收购价格指数	(390)

10.5.4 股票价格指数	(391)
10.5.5 产品成本指数	(398)
10.5.6 空间价格指数	(399)
10.6 综合评价指数	(399)
10.6.1 综合评价的基本思想	(399)
10.6.2 综合评价指数的构建	(401)
10.6.3 几种常用的综合评价指数	(403)
小结	(410)
思考与练习	(411)
附录1 部分习题参考答案	(414)
附录2 Excel 在统计中的应用	(418)
附录3 Excel 统计函数一览表	(444)
附录4 常用统计表	(449)
参考文献	(480)
后 记	(481)

第1章

导 论

读书是在别人思想的帮助下建立自己的思想。

鲁巴金

统计是静止的历史,历史是流动的统计。

奥古斯特·路德维格·施廖采尔

学习目标

- ◆ 了解统计学的历史。
- ◆ 掌握统计的含义。
- ◆ 了解统计学的研究对象。
- ◆ 掌握统计研究的基本方法。
- ◆ 了解统计研究的程序。
- ◆ 了解统计所具有的职能。
- ◆ 重点掌握统计学的基本范畴。

1.1 统计学的起源和发展

统计学的发展源远流长,我们只能概要地向读者介绍迄今为止统计学的发展脉络,让读者尤其是非统计学专业的读者对统计学有一个初步的认识和了解,对我国统计学的发展和统计工作的现状在头脑中有一个基本轮廓。

1.1.1 统计的起源

统计的起源很早,在原始社会里,人类最初的一般计数活动即蕴藏着统计的萌芽。远在四五千年前,居住在“两河流域”一带的人们就在石头上和泥板上刻出一个个像楔子形的符号,如“丫”和“<”,分别表示 1 和 10。但那时的计数活动还主要表现在人们对仅有的剩余劳动成果或其视线所及的劳动对象加以清点与度量。人类社会经济生活越是向纵深发展,这种计数活动就越是频繁,并且越是复杂,从无意识地单纯处理集体劳动成果的计数开始,变成一种有意识地比较自己与他人劳动成果的计数。随着人类社会组织机构的不断发展与健全,逐步产生了一种专为全社会所需要的、共同一致的计数活动,即出现了一种具有总体性特点的计数实践。这主要是为了获得人类社会组织内有关人口、经济等方面内容的总的认识。特别是建立了国家之后,统治者为了实行其对内对外的职能,需要进行征兵、徭役和课税,就开始了人口、财产和军事统计。在外语中,“统计”一词与“国家”一词来自同一词源。因此,一般认为,有了国家就有了统计实践活动。

我国是世界上最早开始统计活动的国家之一。在原始社会,就产生了一般的计数活动,形成了统计工作的萌芽。结绳记事不仅产生了简单的总量指标,并且应用了简单的统计分组。随着社会生产力的发展,书契记数代结绳记事而起,发展为一种比较进步的计量方法——以数字为主体的经济记录。在夏商奴隶社会时期,《尚书·禹贡》是一篇反映夏代国势调查的重要文献,记述九州的基本情况,并按土质的优劣把九州的田、赋进行了复合分组,从而形成了统计的雏形。以后的周王朝在统计方面更为完善,不仅制定了乡的定期报表制度,在统计方法上还应用了专门调查、统计图示及帐册,当时的中国人就知道并运用了统计分组、平均数、相对数等近代统计方法。进入了封建时代,随着朝廷统治的需要,统计工作进一步完善,并且一些杰出的思想家提出过不少科学的统计思想和先进的统计方法。如汉代司马迁在他编的《史记》中多次用到统计表,晋代在分组上所采用的两端开口组,宋代中位数的应用,等等。尽管如此,当时中国统计工作却始终没有发展成系统的、理论形态的知识体系,这是由于中国始终没有进入资本主义社会,商品经济始终没有形成一种社会经济形态,所以统计学始终没有形成一个系统的学科。另外,自秦汉以来,中国就形成了封建的大一统国家,结束了春秋战国时期群雄并立的局面,几千年中,再没有从横向作国情国力比较的压力。

在国外统计实践也有悠久的历史。公元前 3050 年,埃及为了征集金字塔的建造费,对全国人口和财产进行了普查。到了 17 世纪至 18 世纪资本主义上升时期,随着经济文化的发展和社会分工的日益发达,引起了对统计的新需要,统计的范围扩大到经济活动的各个领域,统计有了很大的发展。19 世纪中叶以后,统计得到了日益广泛的应用。

1.1.2 统计学的三个发展阶段

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 2300 多年的历史。我国历史上产生的统计思想,在当时世界上也居于光辉地位。例如,早在春秋战国时代的管仲就曾指出:“不明于计数而欲举大事,犹无舟楫而欲经于水险也”。商鞅指出:“欲强国,不知国十三数,地虽利,民虽众,国愈弱至削。”他们已把统计活动的重要性提到哲学和关系国家兴亡的高度。以后历代在统计思想上不断有所发展。所有这些,给我们留下了珍贵的遗产。统计学起源于研究社会经济问题,今天仍然是我们研究社会经济问题的基本方法。在 2000 多年的发展过程中,统计学至少经历了“城邦政情”、“政治算术”和“统计分析科学”三个发展阶段。

1. “城邦政情”阶段

“城邦政情”(matters of state)阶段始于古希腊的亚里斯多德撰写“城邦政情”或“城邦纪要”。他一共撰写了 150 余种纪要,其内容包括各城邦的历史、行政、科学、艺术、人口、资源和财富等社会和经济情况的比较、分析,具有社会科学特征。“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“政治算术”这个名词所替代,并且很快被演化为“统计学”(statistics)。统计学依然保留了城邦(state)这个词根。

2. “政治算术”阶段

“政治算术”(political arithmetic)阶段与“城邦政情”阶段并没有很明显的分界点,二者也没有很本质的差别。“政治算术”的特点是,统计方法与数学计算和推理方法开始结合,分析社会经济问题的方式更加注重运用定量分析方法。一般以 1690 年英国古典政治经济学的奠基人威廉·配第(W. Petty, 1623 - 1687)《政治算术》一书的出版作为这个阶段的起始标志。在这部著作中,他以数字资料为基础,用计算和对比的方法对英、法、荷的实力进行了比较,论证了英国称雄世界的条件和地位。威廉·配第这本书出名之处,不在其分析的结论,而在于其所使用的独创的方法。用他自己的话来说:“我进行这种工作所使用的方法,在目前还不是常见的。因为我不采用比较级或最高级的词语进行思辩式的议论,相反地采用了这样的方法(作为我很久以来就想建立的政治算术的一个范例),即用数字、重量和尺度来表达自己的想说的,只进行诉诸人们的感受的议论,借以考察在自然中有可见根据的原

因。”

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《政治算术》被后来的学者评价为近代统计学的来源。马克思称他为“政治经济学之父,在某种程度上说是统计学的创始人”。刺激近代统计学发展的主要原因是战争和工业革命。当时政治家需要有人专门致力于收集和分析社会、经济和政治方面的统计资料,帮助他们决策和管理国家。威廉·配弟写作《政治算术》大约在1671—1676年之间,正值爆发第三次英荷战争(1672—1674年)期间。与此同时,他还撰写了《爱尔兰的政治解剖》一书。在威廉·配弟的心目中,政治算术和政治解剖这两种方法是二位一体的。有些学者认为《政治算术》是“比较统计”(comparativ-statistik),《爱尔兰的政治解剖》是“个别统计”(einzel-statistik)。

构成“政治算术”方法的核心是“数字”。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。因为受历史条件的限制,书中通过严格的统计调查得到的数据少,根据经验得出的数字多。第二类是运用某种数学方法推算出来的数字。这类数据配弟在书中使用很多,其推算方法可分为三种:

- ① 以已知数或已知量为基础,循着某种具体关系进行推算的方法;
- ② 通过运用数字的理论性推理来进行推算的方法;
- ③ 以平均数为基础进行推算的方法。

第三类是为了进行理论性推理而采用的例示性的数字。配弟把这种运用数字和符号进行的推理称之为“代数的算法”。从配弟使用数据的方法看,“政治算术”阶段的统计学已经比较明显地体现了“收集和分析数据的科学和艺术”特点,统计实证方法和理论分析方法浑然一体,这种方法即使是现代统计学也依然继承。

政治算术学派的另一个代表人物是约翰·格朗特(J. Graunt, 1620—1674)。17世纪上半叶,英国多次发生严重的瘟疫,政府定期公布有关人口出生和死亡的数字,约翰·格朗特利用这些资料并发表了《关于死亡表的自然观察和政治观察》的论著,首次提出通过大量观察,可以发现新生儿性别比例具有稳定性和不同死因的比例等人口规律;并且第一次编制了初具规模的“生命表”,对年龄死亡率与人口寿命作了分析,从而引起了普遍的关注。他清楚地说明了统计学作为国家管理工具的重要作用。

3. “统计分析科学”阶段

在“政治算术”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”(science of statistical analysis)。

19世纪末,欧洲大学里开设的“国情纪要”或“政治算术”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。英国数学家戈塞特