

STATISTICS

统计学

S T A T I S T I C S

肖彦花 马知遥 编著

国防科技大学出版社

统计学

STATISTICS

陈永贵 主编

统 计 学

肖彦花 马知遥 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/肖彦花,马知遥编著. —长沙:国防科技大学出版社,2004.2

ISBN 7-81099-025-X

I. 统… II. ①肖…②马… III. 统计学—高校教材 IV. O221

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:28.5 字数:640千
2004年2月第1版第1次印刷 印数:1-2500册

ISBN 7-81099-025-X/O·3

定价:39.80元

目 录

	第1章 导 论
1	1.1 统计学的起源和发展
1	1.1.1 统计的起源
2	1.1.2 统计学的三个发展阶段
4	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
5	1.2.1 统计的涵义
7	1.2.2 未来的统计
7	1.2.3 统计学的研究对象及其特点
9	1.2.3 统计学的性质
10	1.3 统计学的研究方法和研究程序
10	1.3.1 大数定律的方法论意义
10	1.3.2 统计研究的基本方法
12	1.3.3 统计研究程序
14	1.4 统计学的分科
14	1.4.1 描述统计学和推断统计学
14	1.4.2 理论统计学和应用统计学
15	1.5 统计的职能
17	1.6 统计学的基本范畴
17	1.6.1 统计总体和总体单位
18	1.6.2 标志和标志表现
19	1.6.3 统计指标和指标体系
24	1.6.4 变异、变量与变量值
25	小结
26	思考与练习
	第2章 统计设计和统计调查
27	2.1 统计设计
27	2.1.1 统计设计的意义和目标

28	2.1.2 统计指标和指标体系的设计
32	2.2 数据的计量与分类
32	2.2.1 数据的计量尺度
34	2.2.2 数据的分类
37	2.3 统计调查
37	2.3.1 统计调查的意义、形式和方法
43	2.3.2 统计调查方案
48	2.3.3 统计调查的组织方式
58	小结
58	思考与练习
	第3章 统计整理与数据显示
61	3.1 统计整理
61	3.1.1 统计整理的意义及步骤
63	3.1.2 统计分组
72	3.1.3 次数分布
82	3.1.4 统计汇总
85	3.2 数据显示
85	3.2.1 统计表
91	3.2.2 统计图
100	小结
101	思考与练习
	第4章 综合指标
105	4.1 总量指标
105	4.1.1 总量指标的意义
106	4.1.2 总量指标的种类
107	4.1.3 总量指标的计量单位
109	4.1.4 总量指标的计算
114	4.2 相对指标
114	4.2.1 相对指标的意义
114	4.2.2 相对指标的表现形式
116	4.2.3 相对指标的计算
123	4.2.4 相对指标的应用条件

124	4.2.5 正确运用相对指标应注意的问题
125	4.3 平均指标
125	4.3.1 平均指标的意义和种类
128	4.3.2 算术平均数
135	4.3.3 调和平均数
138	4.3.4 几何平均数
140	4.3.5 众数与中位数
144	4.3.6 平均指标的应用
147	4.4 变异指标
147	4.4.1 变异指标的意义
149	4.4.2 异从比率、极差、分位差和平均差
160	4.4.3 偏度与峰度
163	小结
164	思考与练习

第5章 概率与概率分布

167	5.1 试验、事件与概率
167	5.1.1 试验与事件
168	5.1.2 概率
169	5.2 随机变量
170	5.2.1 离散型随机变量的概率分布
170	5.2.2 连续型随机变量的概率分布
172	5.2.3 概率分布的特征
174	5.3 几种重要的概率分布
175	5.3.1 二项分布
175	5.3.2 正态分布
178	5.3.3 χ^2 分布
179	5.3.4 t 分布
180	5.3.5 F 分布
181	小结
182	思考与练习

第6章 抽样估计

185	6.1 抽样推断中的基本概念
-----	----------------

185	6.1.1 总体和样本
186	6.1.2 总体参数和样本统计量
186	6.1.3 重置抽样和不重置抽样
187	6.1.4 样本容量和样本个数
187	6.1.5 抽样单位与抽样框
188	6.1.6 抽样误差
188	6.2 抽样分布
189	6.2.1 抽样分布的特征值
189	6.2.2 抽样分布定理
192	6.3 抽样估计的基本方法
192	6.3.1 点估计
194	6.3.2 区间估计
202	6.3.3 样本容量
203	小结
204	思考与练习
	第7章 假设检验
205	7.1 假设检验概述
205	7.1.1 假设检验的概念
206	7.1.2 假设检验的检验法则
207	7.1.3 假设检验中的两类错误
209	7.1.4 假设检验的基本思想
210	7.1.5 假设检验的步骤
212	7.2 总体均值的假设检验
212	7.2.1 单一总体均值的假设检验
216	7.2.2 两总体均值之差的假设检验
221	7.3 总体成数的假设检验
221	7.3.1 单一总体成数的假设检验
222	7.3.2 两总体成数之差的假设检验
224	7.4 总体方差的假设检验
225	7.4.1 单一正态总体方差的假设检验
227	7.4.2 两不同正态总体方差的假设检验
229	小结
230	思考与练习

第 8 章 相关与回归分析

233	8.1 相关分析概述
233	8.1.1 函数关系与相关关系
234	8.1.2 相关关系的种类
235	8.1.3 相关分析的概念及内容
236	8.2 相关分析方法
236	8.2.1 相关表与相关图
239	8.2.2 相关系数
244	8.3 回归分析概述
244	8.3.1 回归分析的概念及内容
245	8.3.2 回归分析与相关分析的关系
246	8.3.3 相关分析与回归分析应注意的事项
247	8.4 线性回归分析
247	8.4.1 简单线性回归分析
254	8.4.2 多元线性回归分析
261	8.5 非线性回归分析
261	8.5.1 非线性回归分析概述
262	8.5.2 非线性回归方程的测定
269	小结
270	思考与练习

第 9 章 时间数列

275	9.1 时间数列的编制
275	9.1.1 时间数列的意义和种类
280	9.1.2 编制时间数列的基本原则
280	9.2 时间数列的分析指标
282	9.2.1 时间数列的水平分析指标
288	9.2.2 时间数列的速度分析指标
293	9.2.3 应用时间数列分析指标时应注意的问题
294	9.3 长期趋势的测定
294	9.3.1 时间数列的构成要素与模型
295	9.3.2 长期趋势的测定
303	9.3.3 趋势线的选择

304	9.4 季节变动的测算
304	9.4.1 季节变动及其测算目的
304	9.4.2 季节变动分析的原理与方法
310	9.4.3 季节变动的调整
312	9.5 循环变动的测算
312	9.5.1 循环变动及其测算目的
313	9.5.2 循环变动的测算方法
317	9.6 不规则变动的测定
317	9.7 时间数列的预测方法
317	9.7.1 移动平均法
319	9.7.2 指数平滑法
321	9.7.3 趋势外推法
323	9.7.4 季节指数法
324	小结
326	思考与练习
	第 10 章 统计指数分析
329	10.1 统计指数概述
329	10.1.1 统计指数的概念、性质和作用
332	10.1.2 统计指数的种类
335	10.2 总指数的编制
335	10.2.1 总指数编制的基本问题
337	10.2.2 总指数的编制方法
343	10.3 指数体系与因素分析
343	10.3.1 指数体系
344	10.3.2 因素分析法
352	10.4 指数数列
353	10.4.1 定基指数与环比指数
353	10.4.2 不变权数指数与可变权数指数
355	10.5 几种常用的经济指数
355	10.5.1 居民消费、商品零售价格指数
361	10.5.2 工业生产指数
362	10.5.3 农副产品收购价格指数
362	10.5.4 股票价格指数

370	10.5.5 产品成本指数
370	10.5.6 空间价格指数
371	10.6 综合评价指数
371	10.6.1 综合评价的基本思想
372	10.6.2 综合评价指数的构建
374	10.6.3 几种常用的综合评价指数
381	小结
382	思考与练习
386	附录1 部分习题参考答案
389	附录2 Excel 在统计中的应用
414	附录3 Excel 统计函数一览表
419	附录4 常用统计表
444	后 记
445	参考文献

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

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认识和了解,对我国统计学的发展和统计工作的现状有一个基本轮廓。

1.1.1 统计的起源

从历史上看,统计实践活动远远早于统计学的诞生。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。早在原始社会末期奴隶社会的形成过程中,我国就出现了社会经济统计的萌芽。从公元前 21 世纪夏禹立国开始,中国的统计不仅详细地记录了重大历史活动成果,如人口和土地,而且明显地被新兴的奴隶制国家用作治国的手段。到了西周时期,人口调查开始确立了一些制度。户籍制度在春秋时期有所发展。

中国早在四千多年前夏王朝,《尚书·禹贡》把当时的中国分为九州,分别叙述了各地的物产、交通、植物特征等情况,又依照土质不同,按照复合分组的方式把田地及贡赋分为九等,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,不仅制定了定期的报表制度,在统计方法上还应用了专门调查、统计图示及账册,

当时的中国人已知道并运用了统计分组、平均数、相对数等近代统计方法。进入封建时代,由于朝廷统治的需要,统计工作进一步完善,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《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。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,没有了从横向进行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,埃及为了征稽金字塔的建造费,对全国人口和财产进行了普查。17 世纪至 18 世纪是资本主义的上升时期,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,迄今已有二千三百多年的历史。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也不断有所发展。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—1637)《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于其分析的结论,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想说的内容,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·配第的《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·配第写作《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·配第的心目中,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”(Comparativstatistik),《爱尔兰的政治解剖》是“个别统计”(Einzelstatistik)。

构成“政治算术”方法的核心是“数字”。配第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。配第在书中使用了很多这类数据,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。配第把这种运用数字和符号进行的推理称为“代数的算法”。从配第使用数据的方法看,“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据,约翰·格朗特利用这些资料发表了《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(William Sealy Gosset, 1876—1937)发表了《关于小样本t分布》的论文,这是一篇在统计学发展史上具有划时代意义的文章,它创立了小样本代替大样本的方法,开创了统计学的新纪元。其后由费雪尔(Ronald Aylmer Fisher, 1890—1962)加以充实,并由波兰统计学家尼曼(J. Neyman, 1894—?)以及英国数理统计学家皮尔逊(Karl Pearson, 1857—1936)等人进一步发展,建立了统计假设理论。后来美国统计学家瓦尔德(A. Wald, 1902—1950)又将统计学中的估计和假设理论加以归纳,创立了“决策理论”;美国的威尔克斯(S. S. Wilks, 1906—1964)、英国的威沙(J. Wishart, 1898—1956)等统计学家对样本分布理论又加以充实和发展;美国的科克伦(W. G. Cochran, 1909—1980)等在1957年又提出了实验设计的理论和方法,进一步拓宽了统计学的研究范围。现代统计学的代表人物首推比利时统计学家奎特莱(Adolphe Quélet),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,因为他深信统计学是可以用于研究任何科学的一般研究方法。

现代统计学的理论基础概率论始于研究赌博的机遇问题,大约开始于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,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方法则。于是,现代统计方法便有了比较坚实的理论基础。

1.1.3 中国统计学的现状

旧中国的统计比较落后,为数不多的统计学者主要受英美数理统计学派的影响。新中国成立后,指导我国统计工作的理论全盘照搬了前苏联的社会经济统计学,由于受到前苏联1954年统计科学会议的影响,即把统计学定义为一门具有阶级性、党性的社会科学。这种统计学理论体系被中国统计界的一些学者称为“马克思主义的无产阶级统计学”。虽然前苏联的统计模式给中国统计界提供了良好的示范作用,但把统计学确立为前苏联模式后,由于在科学概念上把社会经济统计学与数理统计学割裂开来,就使得社会经济统计学的发展始终限定在以简单描述为工具的狭小圈子里。因而1980年以前中国的社会经济统计学的发展缺乏生机,进步迟缓。

改革开放加快了中国统计学的发展。中国统计面临着世界统计的严峻挑战,

挑战和机遇要求我们必须坚持实事求是的科学态度。经过一个较长时期的讨论和探讨,我国统计界终于发生了质的变化和飞跃。1992年11月,国家技术监督局正式批准统计学为一级学科,1998年教育部进行的专业调整也将统计学归为一级学科。统计学一级学科地位的确立表明中国统计在与国际接轨的进程中迈出了重要的一步。

在国际惯例上根本就不存在社会经济统计学和数理统计学之分。统计学是一门综合性很强的边缘科学,它既不是数学,也不是经济学。统计学是研究如何对事物随机现象总体数量(包括人文与社会、自然等广泛的领域)进行搜集、加工、整理以及推断、分析、预测和决策的科学。因此,统计学是以定性为依据,定量地认识事物随机现象、数量、规律性的方法论。可以说,建立涵盖社会经济统计学和数理统计学的“大统计学”观点已经有了现实的基础。

“大统计”概念是1993年12月在中国统计学会会长会议上提出来的,基本观点是认为只有一门统计学,并且认为统计学是收集和分析大量数据的方法论科学。收集和分析数据既需要定量分析,也需要定性分析;既需要理论指导,也需要技术支撑;既需要描述方法,也需要推断方法。因此,社会经济统计学和数理统计学不应当相互对立,而应当相互融合,在融合过程中发展现代统计理论,开创新的统计方法。

1.2 统计学的研究对象和性质

1.2.1 统计的涵义

统计最早出现于中世纪拉丁语的 Status,意思是指各种现象的状态和状况。由这一语根组成意大利语 Stato 表示“国家”的概念,也含有国家结构和国情知识的意思。根据这一语根,最早作为学名使用的“统计”是在18世纪德国政治学教授阿亨瓦尔(G. Achenwall)在1749年所著《近代欧洲各国国家学纲要》一书的绪言中,把国家学名定为“Statistika”(统计)这个词。原意是指“国家显著事项的比较和记述”或“国势学”,认为统计是关于国家应注意事项的学问。此后,各国相继沿用“统计”这个词,并把这个词译成各国的文字,法国译为 Statistique,意大利译为 Statistica,英国译为 Statistics。日本最初译为“政表”、“政算”、“国势”、“形势”等,直到1880年在太政官中设立了统计院,才确定以“统计”二字正名。1903年(清光绪廿九年)由钮永建、林卓南等翻译了4本横山雅南所著的《统计讲义录》一书,把“统计”这个词从日本传到我国。1907年(清光绪卅三年)彭祖植编写的《统计学》在日本出版,同时在国内发行,这是我国最早的一本“统计学”书籍。“统计”一词就成了记述国家和社会状况的数量关系的总称。

“统计”一词,英语为 Statistics,用作单数名词时指的是统计学;用作复数名词时

表示“统计数据”或“统计资料”,这也表明统计数据不是指单个的数字,而是由多个数据构成的数据集。单个的数字显然用不着统计方法进行分析,仅凭一个数据不可能得出事物的规律,只有经过对同一事物进行多次观察或计量后得到的大量数据,才能利用统计的方法探索出内在的必然规律性。

统计是随着社会生产的发展、适应国家与社会管理的需要而产生和发展的,其涉及的范围也逐步扩展到社会的各个领域。统计具有不同的涵义和范围。“统计”一词在中文中一般解释成这样三种涵义,即统计工作、统计数据和统计学。

1. 统计工作

统计工作即统计实践,是指为了研究客观事物的数量关系和数量特征而进行调查研究的认识活动。一项统计活动过程通常由以下几个环节构成:统计目的,统计设计,即制定调查研究和实(试)验研究的计划;统计调查,即统计数据的搜集;统计整理,即对数据资料进行清理、纠错并加以存储;统计分析,即计算相应指标以反映研究对象的内在特征和规律;统计服务,即数据资料的提供、管理和开发。一个环节衔接另一个环节,由参加工作的各个地区、部门和单位组成统计网络,密切协作,互相配合,共同完成。从事统计工作的人员,一般简称“统计人员”。例如说“某某是搞统计的”,即说明某某是从事统计工作的。

2. 统计数据

统计数据即统计信息,是指通过统计工作过程所取得的各项数据资料以及与之相关的其他资料的总称。在一般情况下,统计数据资料都集中收编和反映在统计公报、统计年鉴、数据手册、统计资料汇编、统计图表和统计分析报告中。为了保证统计工作的质量,人们对统计数据提出了基本要求,即准确、及时和全面。准确,就是数据要真实可靠,如实反映情况,决不允许虚报、瞒报或任意编造统计数字;及时,就是要讲求数据的时效性,不拖延滞后;全面,就是数据资料不缺不漏,周详完整。通常人们说统计怎么样,其意义就是用统计数据来说明问题。

3. 统计学

统计学即统计理论,是“收集和分析数据的科学和艺术”。统计学由一套收集和处理统计数据的方法所组成,这些方法来源于对统计数据的研究,其目的是探索数据内在的数量规律,以达到对客观事物的科学认识。统计数据如果不用统计方法去分析,就仅仅是一堆杂乱无章的数据而已,无法得出任何有益的结论。

统计的三种涵义之间存在着密切的联系(见图1-1)。统计数据既是统计工作的基本要素,又是统计工作的成果体现。统计学一般说来是统计工作实践经验的理论概括,同时,它又反过来从理论和方法上指导统计工作,推动统计工作不断发展。因此,统计工作和统计数据是统计活动和统计成果的关系,统计学和统计工作是理论和实践的关系。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、发展的客观条件,统计科学则是统计工作经验、社会经济理论、