

应用统计

贾怀勤 主编

对外经济贸易大学出版社

应 用 统 计

修 订 本

贾怀勤 主编

对外经济贸易大学出版社

(京)新登字 182 号

图书在版编目 (CIP) 数据

应用统计/贾怀勤主编. 修订本. —北京: 对外经济贸易大学出版社, 1998.9

ISBN 7-81000 865 X

I. 应… II. 贾… III. 应用统计学 IV. C8

中国版本图书馆 CIP 数据核字 (98) 第 17917 号

©1998 年 对外经济贸易大学出版社出版发行

版权所有 翻印必究

应用统计

贾怀勤 主编

责任编辑: 张孟秋

对外经济贸易大学出版社

北京惠新东街 12 号 邮政编码: 100029

北京飞达印刷厂印刷 新华书店北京发行所发行

开本: 850×1168 1/32 印张: 19 490 千字

1998 年 9 月北京第 1 版 2001 年 1 月北京第 2 次印刷

ISBN 7-81000-865-X/F · 340

印数: 5001—8000 册 定价: 29.00 元

修 订 说 明

本书1994年6月出版以来,历经一次重印。作为一部配合“国际经济贸易专业”和“企业管理专业”,而非统计专业教学需要而编写的教材,它受到了特定读者群的欢迎和有关人士的关注。今年,本书已荣幸地被列为国家教委大学推荐教材的候选对象。

此次修订工作包括:①对小部分内容进行了改写;②对一些例题及相关统计图表进行了更替;③订正了原书中属于排版的错误。担任修订任务的人员及其负责范围是:贾怀勤修订绪论、第一章、第五章、第十三章和第十四章;张杰修订第八章至第十二章;杜学孔修订第四章;王玉蓉修订第七章;杨宝峰修订第二章;赵宇辉修订第三章和第六章。

四年来,广大读者对本书提出了许多富有参考价值的意见。我们在这次修订时对这些宝贵意见尽可能采纳。敬请读者继续垂顾,批评指正。

编 者

1998年2月

前 言

80年代初,对外经济贸易大学在创办国际企业教育时,即将“应用统计”(Applied Statistics)定为一门主干课程。十余年来,本着以国际企业需求为导向、中西融通、文理渗透的方针,经老中青三代统计学教师不断探索,已建立起具有经贸特色的统计学教学体系,包括一整套教学大纲、教科书、教学参考资料、计算机应用软件和相应的教学方法,并有幸获得北京市优秀教学成果一等奖(1993年)。本课程所用教材,原为李志伟先生等编著的《统计分析概论》(对外贸易教育出版社,1984年初版,1989年修订再版)。

建设社会主义市场经济必然对经贸高等教育,当然也包括对“应用统计”的教学提出新的要求,特别是要求经贸工作者提高在国内外市场的复杂环境中运用统计信息对经营管理进行分析,并作出科学决策的能力。为此,我们决定编写这部《应用统计》,作为各经贸专业本科“应用统计”课程的教科书,并力求使之更好地适应新形势下培养经贸人才的需要。这部教材对“应用统计”的知识框架重新作了安排,在内容的广度和深度上都较原用教材有所发展。

本教材由贾怀勤主编。其中,绪论、第一章、第五章、第十三章和第十四章由贾怀勤执笔;第八章、第九章、第十章、第十一章和第十二章由张杰执笔;第二章、第三章和第六章由朱雅华执笔;第四章由杜学孔执笔;第七章由王玉蓉执笔。

恳请读者对本书不当和疏漏之处提出宝贵意见。

编 者

1993年5月

目 录

绪论	(1)
第一章 统计设计	(16)
第一节 统计设计概述	(16)
第二节 统计指标的设计	(25)
第三节 实验设计	(31)
第二章 统计资料的搜集和整理	(37)
第一节 统计资料的类别和来源	(37)
第二节 统计调查的种类	(38)
第三节 统计资料的整理和表述	(47)
第三章 数据的静态对比分析	(73)
第一节 相对数的意义和种类	(73)
第二节 结构相对数	(75)
第三节 其它静态相对数	(84)
第四章 单变量截面数据的描述性分析	(92)
第一节 频数分布数列的图形表述	(93)
第二节 集中趋势指标	(102)
第三节 离散趋势指标	(117)
第四节 偏态与峰度	(124)
第五节 位次指标	(127)
第五章 双变量截面数据的描述性分析	(145)

第一节	列联表·····	(145)
第二节	相关和回归分析·····	(153)
第六章	时间数列分析·····	(178)
第一节	时间数列的动态比较和平均分析·····	(178)
第二节	时间数列的解析分析·····	(194)
第七章	指数·····	(235)
第一节	指数的概念·····	(235)
第二节	指数的编制方法·····	(237)
第三节	指数的应用·····	(253)
第四节	因素分析·····	(264)
第八章	概率·····	(276)
第一节	概率的概念·····	(276)
第二节	概率计算规则·····	(282)
第九章	概率分布·····	(309)
第一节	随机变量与概率分布·····	(309)
第二节	常用的离散型概率分布·····	(321)
第三节	正态分布·····	(342)
第十章	抽样和抽样分布·····	(358)
第一节	抽样的基本概念·····	(358)
第二节	抽样分布的基本原理·····	(369)
第十一章	参数估计·····	(403)
第一节	点估计与区间估计·····	(403)
第二节	区间估计的应用·····	(415)

第三节	样本容量的确定	(425)
第十二章	假设检验	(434)
第一节	假设检验的有关概念	(434)
第二节	单一样本的假设检验	(451)
第三节	双样假设检验	(466)
第十三章	关于简单回归和相关的统计推断	(480)
第一节	关于因变量的估计和推算	(480)
第二节	关于回归和相关参数的推断	(491)
第三节	可决系数	(495)
第十四章	多元回归和多重相关分析	(507)
第一节	多元线性回归关系的描述和推断	(508)
第二节	多重相关分析	(518)
第三节	多元回归分析中的 F 检验	(528)
第四节	关于多元回归模型假定条件的讨论	(532)
第五节	多元回归和多重相关分析的综合运用	(538)
统计数表		(556)
附表 1	二项分布累积概率表	(556)
附表 2	普哇松分布累积概率表	(559)
附表 3	标准正态分布	(564)
附表 4	随机数字 4000 个	(565)
附表 5	t 分布	(568)
附表 6	F 分布	(569)
附表 7	杜宾-沃岑检验上下界	(579)
词汇表 (中英文对照)		(581)

绪 论

人类的统计活动，尤其是调查和整理社会经济总合现象数据的活动，有着悠久的历史。据一些统计史专家考证，关于人口、土地数字的搜集和整理活动，在中国可以上溯到殷、周时代，在外国可以追及古埃及、古希腊和古罗马。然而，由统计实务上升到理论和系统的方法的研究，最早也只能从17世纪算起。至今，统计学的发展大体上经历了三个阶段。

从17世纪中叶到19世纪中叶，为统计学的初创阶段。具体地讲，可以把英国的威廉·配第（William Petty）于1671—1676年写成《政治算术》一书作为统计实务上升到理论和方法的标志。或者说，初创阶段的统计学，始于“政治算术”这门有统计学之实、却无统计学之名的学问。马克思说威廉·配第“在某种程度上也可以说是统计学的创始人”^①。在此阶段，统计学主要用来描述和比较国家的综合国力，及人口研究，从研究对象上说是实质性而不是方法论的学问，从方法上看是经验研究多于理论研究，尚未进入系统的理论研究。

从19世纪中叶到本世纪40年代，是统计学发展的第二阶段——近代统计学阶段。1869年凯特勒（L. A. J. Quetelet）发表了《社会物理学》，揭开了近代统计学发展的序幕。凯特勒把作为数学分支的概率论引入了统计研究，解决了客观总合现象在数量变化上的规律性问题。而后，统计理论和方法又取得了一系列进步，如高尔登（F. Galton）的回归理论，戈塞特（R. A. Gosset）的小

^① 《马克思恩格斯全集》第23卷，第302页，人民出版社，1972年。

样本理论，费煦 (R. A. Fisher) 的 F 分布理论和方差分析方法，皮尔逊 (E. S. Pearson) 的区间估计方法等等，都极大地推动了统计理论和方法的发展和应用，特别是在自然科学和工程技术中的应用。至此，统计学已经建立起系统的理论和方法，并且实现了由实质性科学向方法论科学的转变。

从本世纪 50 年代起，统计学进入了它的第三阶段——现代统计学阶段。这个阶段带有三个明显的特点：其一是统计理论和方法的应用有了广泛的发展，不仅自然科学研究方面大量应用统计方法，就是社会和人文科学也越来越广泛地应用统计方法，特别是在经济和工商管理领域尤为如此；其二，进一步开发出一系列新的统计方法，如统计预测的新方法、多元统计方法和探索性数据分析等等；其三是统计研究与电子计算机应用密切结合。这三个特点也是相互联系的。统计学的应用扩展到自然科学、工程技术、经济和企业管理、社会、人口、语言等各个学科领域，极大地推动了这些学科的发展。反过来，统计学在各个实质性学科的应用又促进了统计理论和方法的发展。这既指一般统计理论和方法的进步，又指适用于专门领域的统计方法的开发。而计算机的应用使得现代统计方法的应用有了计算手段的保障，同时又为新的统计方法的开发提供了方便。

现代统计学发展到今天，已经成为与数学和一系列实质性学科互有交叉的综合性的横断学科。十多年前，当中国的统计学研究刚刚从十年浩劫中复苏过来时，曾经有过“一门”还是“两门”之争。现在比较普遍的认识是，统计学是一个体系庞大、分支众多的大家族。在统计学这个学科体系中，尽管有许多分支学科，但是它们的基本理论和基本方法却都是共同的。阐述这些共同的统计基本理论和基本方法的统计分支学科，通常被国内统计学界称为“统计学原理” (Principles of Statistice)，国外也称为“基础统计学” (Elementary Statistics)。本书作为国际工商管理本科的统计学教材，在阐述共同的统计原理和方法的过程中，以其

在企业管理和经贸调研中的应用为导向，故定名为《应用统计》(Applied Statistics)。

一、统计学的定义、性质和研究对象

统计学是在质与量的辩证统一中，通过大量观察和平均方法，对客观总合现象 (aggregate phenomena) 的数量特征进行研究的方法论学科。

统计学离不开数。统计学涉及的数，不是随意的或单纯的数，而是能表明客观现象某一方面特征的数，如气温数、水压数、钢产量数、入境人数、不合格产品数，等等。有的客观现象特征不适合用数字描述。另外一些客观现象特征既可以用文字描述，又可以用数字描述。对于后者，使用数字描述往往更确切。比如说“黑龙江省北部地区冬天非常冷”，就不如说那里“一月份平均气温在 -30°C 以下”更确切。关于客观现象特征的记载，有文字型资料 (alphabetic material) 和数据型资料 (numerical data) 两类。只有后者才是统计研究的素材。

统计资料 (statistics) 是反映一定空间、时间条件下客观总合现象数量特征的数据型资料。统计资料又称为统计数据或统计数字 (statistical data)。它有数量性、客观性、具体性和总合性四个特点。统计资料存在的价值，就在于说明现象的数量特征。统计资料是客观的，即真实的，是通过科学的搜集方法得到的，反映过去或现时状况的资料。虚构的数字，不是统计资料。咨询机构作出的预测数字，计划部门制定的计划数字，尽管都与统计资料有关，但是其本身并不属于统计资料的范畴。统计资料是具体的，任何统计资料都不是以孤立的数值形式存在的，它总是与特定时间、空间和计量单位联系着。统计资料具有总合性，其用途就在于描述客观总合现象，而不是个别现象。对个体观察取得的数据，总是有待于汇总。

统计学所观察和分析的客观现象包括自然现象和社会现象。按照辩证唯物主义的基本原理，客观世界存在着从无生命运动到

有生命运动，再到人类的社会运动，这样的不同运动形式。社会现象虽然是最高级的运动形式，但它与自然现象也不是截然对立的。应该看到，它们在总合数量特征的研究方法上有许多共同的东西。例如，通过定期观察气温求得平均气温的分析方法，与通过对居民户进行选点调查求得居民家庭平均收入的分析方法，从本质上是相同的，又如，设立试验田块，观察、研究施肥量对农作物产量的影响，与抽取一定数目的企业，研究“研究与开发”经费支出与企业经济效益的关系，从分析方法上讲并没有什么不同。

统计学是借助于观察和分析具体的总合现象来研究观察和分析总合现象数量特征的一般方法。统计学不研究各种现象自身的发展规律。各种自然现象和社会现象都有其自身发展规律，对这些规律的研究只能由各门实质性科学承担。统计学提供的一系列研究方法，使得人们有可能透过对客观总合现象数量特征的观察与分析觉察这种现象自身发展规律的存在，或者加深对这种规律的理解。但是，对这样的实质性规律的论证和解释要由研究特定现象的实质性科学作出。我国第四次全国人口普查资料表明：零岁人口性比例为 1.11，由此往上，一直到 64 岁人口的性比例的 1.04，都是男性多于女性。自 65 岁人口起，性比例开始逆转，为 0.96，随着年岁增大，性比例急剧下降，100 到 85 岁以上人口的 0.40。其它国家的人口统计数字也有类似分布。这些统计资料向人们揭示出一条人口规律，初生儿男性比例略大于女性，在较低年龄段和中年龄段，大体仍保持这一结构，然而由于女性寿命平均高于男性，到了高年龄段则呈现女性比例越来越超过男性的趋势。至于如何解释这条规律，则不属于统计学研究的任务，恐怕要由关于人的生理规律及心理规律的学科来回答。此处所用的对比分析方法，同样可以用来研究我国出口商品结构的变化。据海关统计，1980 年出口额中制成品与初级产品的比例是 50 : 50，1984 年此比例上升到 54 : 46，1988 年又上升到 70 : 30，到了 1991 年则是 78 : 22。这一系列统计数字揭示出我国出口商品中制

成品比重不断增大的趋势。要寻求我国出口商品结构优化的原因，就要从改革开放的基本国策，经济和外贸发展战略，技术进步和产业发展，外贸管理体制改革等多方面来论述。

统计学是一门通用的方法论科学，按照认识世界和研究世界的方法的适用面，各种学科可以划分成三个层次：首先是最一般的世界观和方法论的科学——哲学，其次是一般的方法论科学，再次是实质性科学。统计学属于第二层次。正如形式逻辑不研究思维的具体内容的特殊规律，而研究思维的逻辑形式及其规律一样，统计学不是阐明社会领域和自然领域万千事物本身的实质性规律，而是阐明适用于研究各类总合现象数量特征的一般理论和方法。当然，也正如形式逻辑在研究思维的逻辑形式及其规律时不能抛开思维的具体内容一样，统计学在阐述通用的总合现象数量特征的研究方法时，也不能脱离作为其研究素材的各种总合现象数量特征的统计资料。

既然统计学是通用的方法论科学，就只好截然断言它属于自然科学还是社会科学范畴。国外关于统计学的著述在谈到统计学的定义时，有的使用“理学”（science）的字眼，也有的使用“艺术”（art）的字眼，不得一致。还有人索性称为“一系列方法”（a body of methods），以避免文理之辩。

理论来源于实践，又反过来指导实践。统计学离不开统计活动（statistical activity）。凡是观察和分析客观总合现象数量特征的活动，或者说，凡是搜集、加工和分析关于客观总合现象数据类型资料的活动都属于统计活动的范畴。政府统计机构组织的调查和整理社会经济统计资料的系统活动，是统计活动；其它各种类型的企、事业单位乃至私人搜集、整理和分析统计资料的活动，也应该视为统计活动。这种统计活动可以包括企业为获得市场信息自己作的或委托咨询机构作的营销调研，舆论机构为反映和研究社会问题搞的民意测验，还可以包括为自然科学研究和技术开发而作的某一种客观总合现象的数据采集和加工，等等，国内习惯

把政府统计机构的活动称为统计工作，在国外则称为官方统计活动。值得指出的是，在社会主义市场经济体制下，无论是统计学还是统计活动，对其认识都应该放到“大统计”的基础上。只有这样才有利于统计方法的广泛应用，有利于统计活动在从宏观到微观的社会经济生活中，和科学研究中发挥其作用，有利于统计学自身的发展。

统计学、统计活动和统计资料三者既相互联系，又相互区别。日常人们习惯用“统计”二字作为上述三者的简称，这也未尝不可。只是不要将三者混淆，尤其不要把统计学的研究对象混同于统计活动的对象。

二、大量观察法和平均法是统计学的基本方法

统计活动是对客观总合现象的数量特征进行研究，因而须采用大量观察法 (method of mass observation)。所谓大量观察，就是对同质的客观现象所组成的集团中的大量个体逐一观察，由此取得反映所有个体现象某些特征的一组数据，作为加工、分析的原始资料 (raw data)。观察一词，来源于自然现象的研究，指注视被研究现象的状况及其变化，并及时和如实地加以测量和记录，当然这必须是现场作业。在社会经济现象的研究中，观察一词可作广义的理解，除了指研究者的现场测量和记录外，也包括研究者在事后向当事人提问有关数据，和其它取得有关数据的途径。无论怎样，总要有人亲历该事。

作为大量观察的实施对象的同质个体集团，即为总体 (population)。组成总体的集团成员，称为总体单位 (population element)。每一个体都具有一系列的属性 (characteristic)，如体现在一个人身上的属性有性别、民族、年龄、身高等等，体现在一个企业身上的属性有雇员人数、产值、主营行业等等。总体是在某一属性相同基础上许多个别事物的集合。例如一家保险公司要研究广州市居民参加家庭财产保险业务情况，这里的个体是居民户，居民户有许多属性，诸如常住地、家庭人口、是否拥有自己

住房、是否参加了家庭财产保险，等等。常住地为“广州”的众多家庭户就构成了本研究项目的总体，它们在常住地这个属性上是相同的。那些临时来广州全家旅游的家庭户，就不能作为总体单位。当然，当对总体单位——家庭户问及其它属性时，总有不尽相同的回答。

总体单位的各种属性，或者表现为数量不等，或者表现为状态不同。前者如家庭人口，后者如是否参加了家庭财产保险。这些个体属性的观察结果汇集到总体，就是总体的特征。这种特征都可以用数值来描述。全市每户平均人口数是一个数值，全市参加了家庭财产保险的户数也是一个数值。

表示总体单位属性的具体名词称为标志 (attribute)，对总体单位属性的观察结果称为标志表现。凡其表现可用数值计量的标志称为数量标志 (quantitative attribute)，凡其表现只能用文字描述状态的标志称为品质标志 (qualitative attribute)。对于保险业务研究一例，家庭人口是数量标志，保险状态是品质标志。表示总体特征的名词称为指标。对于本例，全市家庭平均人口是一指标，全市已参加家庭财产保险的户数又是一指标。

总体具有同质性，变异性和大量性的特点。同质性是指总体单位至少应具有一个相同的属性，而又要求具有该相同属性的个体都得纳入这一总体。总体失去同质性，就必然使统计研究陷入迷津，导致错误的认识。变异性是指统计研究感兴趣的标志表现存在的差别。倘若总体单位的标志表现划一，则实无统计研究之必要。大量性是指总体单位数目众多，非如此不能进行大量观察，得不出关于总合现象数量特征的正确认识。

大量观察并不意味着必须对每一总体单位都无一遗漏地观察。许多时候，只要有足够多的总体单位得到观察，就可以说明该总体的数量特征。这些被观察的单位的集合，称为样本 (sample)。组成样本的单位，即是样本单位。至于究竟有多大数目的总体单位才称得上“足够多”，要根据另外一些原则和方法来确

定。

前边已提到，观察可以理解为统计资料的搜集过程。自然现象的观察，往往要使用测量仪器和仪表。有些自然现象研究课题，还要进行实验（experiment），即用人为的方法控制某些现象的发生和人为地创造一些通常不易出现或不能出现的条件，使被观察现象多次重复出现，取得感兴趣的数据。社会现象统计资料的取得，一般通过调查（survey）的途径。因为社会现象比自然现象更加复杂，更为多变，不易采用人工控制条件的实验途径，一般多通过现场观察、记录，或事后向当事人询问、记录，或当事人自报的方法。居民住户调查、集贸市场价格调查和产品质量跟踪调查等，都属于这种情况。对于有些现象，可以借助于行政管理记录取得感兴趣的数据。譬如研究交通事故统计，可以去交通安全管理部门查阅交通事故登记，编制货物进出口统计，则依据海关对进出口货物的监管记录。由于行政管理登记具有法律强制性，其记录较为全面、真实，故此成为某些社会经济统计资料的重要来源。

大量观察法在哲学上的根据就是必然性与偶然性的对立统一。恩格斯说：“被断定为必然的东西，是由纯粹的偶然性构成的，而所谓偶然的东西，是一种有必然性隐藏在里边的形式。”^① 恩格斯又指出：“在表面上是偶然性起作用的地方，这种偶然性始终是受内部隐藏着的规律性支配着，而问题只是在于发现这些规律。”^② 大量观察法的意义在于，在同质总体的基础上，经过大量观察，把个别的、偶然的差异性相互抵消，而将总合的、必然的规律性显示出来。例如对新生儿性别比例进行观察，如果只观察某10个新生儿，其性别比例可能是7:3，或5:5，或其他数值。但是经过对新生儿性别的大量观察，发现这一性别比例稳定在

① 《马克思恩格斯全集》第4卷第240页，人民出版社，1972年。

② 《马克思恩格斯全集》第4卷第243页，人民出版社，1972年。

105 : 100 上,从而显现出新生儿男女性别大体平衡而男略多于女的自然规律。又就一个城市的各家庭人均收入而言,各户高低不一,甚至有些户之间相差悬殊。但是通过大量观察,可以找到全市人均收入的一般水平。这个一般收入水平,是由该市经济发展水平、就业情况和工资政策等多种宏观因素决定的。尽管各户收入不等,但是上述宏观因素决定了全市的一般收入水平必然如此;反过来,这个一般收入水平又体现在千差万别的各户收入之中。

客观总合现象数量方面的必然性与偶然性的对立统一,体现在总体的集中趋势与离散趋势的辩证关系上。而集中趋势和离散趋势的测度又无不通过平均法 (method of mean)。保利 (A. L. Bowley) 说“统计学就是平均法科学”。这种说法在一定意义上是正确和透彻的。平均法是统计学的基本分析方法。总体的同质性、变异性和大量性是平均法的应用前提。在同质的基础上,通过大量观察,使用平均法,可以将总体单位标志表现的差异抽象掉,得出总体集中趋势的测度值。离散趋势的测度值也建立在总体单位标志表现差异平均化基础上。其他许多统计分析方法,如动态现象趋势分析,指数法,期望值标准决策,等等,也都是以平均法为基本分析方法的。

平均法对于认识社会经济规律是非重重要的。马克思在《资本论》研究中,对于平均社会必要劳动量、平均利润率、平均工资、平均地租、平均资本构成、平均市场价格等等,无不以平均数作为规律的表现形式。那些牵动着全球人大大小小每一家公司的中枢神经、可以引发全球经济“大地震”的著名股票价格指数,如纽约道一琼斯指数,东京日经指数,那一轮轮关税贸易总协定多边谈判旨在不断削减的关税税率,其计算也都是平均法的运用。当然社会现象在量上的差异,必须以总体的同质性为前提。例如社会必要劳动量,指“在现有社会正常的生产条件下,在社会平均