

高等院校经济管理类主干课程教材



S 统 计 学 *tistics*

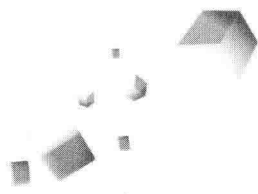


◎主 编 朱喜安
◎副主编 张 敏
班 蕾

 长江出版传媒

 湖北科学技术出版社

高等院校经济管理类主干课程教材

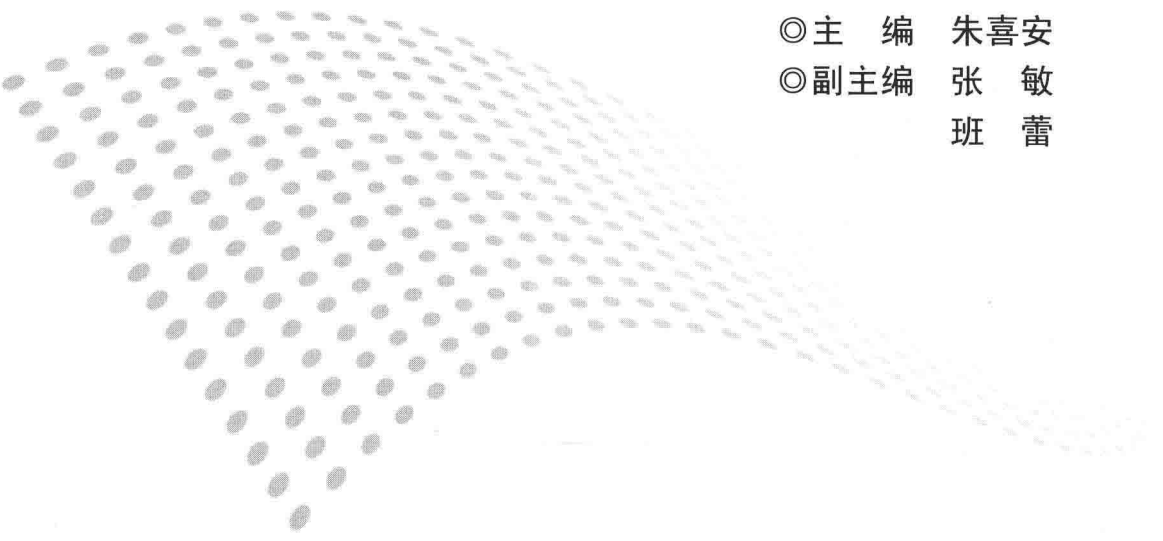


S 统计学

tistics



◎主 编 朱喜安
◎副主编 张 敏
班 蕾



 长江出版传媒
 湖北科学技术出版社

图书在版编目 (CIP) 数据

统计学 / 朱喜安主编; 张敏, 班蕾副主编. — 武汉 :
湖北科学技术出版社, 2013. 12

高等院校经济管理类主干课程教材

ISBN 978-7-5352-6325-4

I. ①统… II. ①朱… ②张… ③班… 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(2013)第 281159 号

责任编辑: 李慎谦

封面设计: 喻 杨

出版发行: 湖北科学技术出版社

电 话: 027-87679468

地 址: 武汉市雄楚大街 268 号

邮 编: 430070

(湖北出版文化城 B 座 13—14 层)

网 址: <http://www.hbstp.com.cn>

印 刷: 武汉市宏达盛印务有限公司

邮政编码: 430000

787 × 1092 1/16

18.5 印张 430 千字

2014 年 1 月第 1 版

2014 年 1 月第 1 次印刷

定价: 46.00 元

本书如有印装质量问题 可找本社市场部更换

总 序

《高等院校经济管理类主干课程教材》是适应我国高等经济管理教育蓬勃发展的需要而组织编写的系列教材。教育部2012年9月发布的《普通高等学校本科专业目录》中,经济学门类计有4小类10个专业,管理学科门类计有9小类32个专业。2006年教育部的数据表明在全国普通高校中经济类专业布点1518个,管理类专业布点4328个。其中绝大多数属于应用型专业。经济管理类应用型专业主要着眼于培养社会主义经济建设发展所需要的德、智、体、美全面发展的高素质专业人才,要求既具有比较扎实的理论功底和良好的发展后劲,又具有较强的专业技能,并且又要求具有较强的创新精神和实践能力。教育是民族振兴和社会进步的基石。我们从推动高等教育内涵式发展,着力提高高等教育质量,努力办好人民满意的教育的高度,确定《高等院校经济管理类主干课程教材》的编写原则与要求如下:

(1)坚持主编负责制。每一本教材的主编,对其主编的教材质量全面负责,并充分发挥参编人员的积极性,吸纳参编人员的智慧。

(2)体系完整、结构严谨。一门独立的课程必然有其区别于其他课程的内容,课程的知识体系决定该课程教材的体系。在确定教材编写大纲及章节安排时,应做到思路清晰、逻辑严密、主次分明、重点突出、承前启后、前呼后应、首尾一贯。

(3)内容充实、新颖。借鉴国内外最新教材,融会当前有关经济管理学科的最新理论和实践经验,用最新的知识充实教材内容。

(4)应特别注重能力培养。全面实施素质教育,教材建设具有重要作用。教材建设不仅要关注知识的现代化与传授,而且要注重学生能力的培养。经济管理类教材建设,更应重视培养学生社会责任感、创新精神和实践能力。

(5)锐意改革与创新。教材建设应秉承锐意改革与创新的精神,及时反映教学改革与创新的成果,同时用教材改革与创新推动整体教育教学改革。

党的十八大明确提出坚定不移沿着中国特色社会主义道路前进,为全面建成小康社会而奋斗的宏伟目标,实现这一宏伟目标需要培养千百万经济管理高素质人才。我们认真着力编好这套《高等院校经济管理类主干课程教材》,既是我们的追求,也是我们的责任。

《高等院校经济管理类主干课程教材》的顺利出版,参加本系列教材编写和审稿的各位老师付出了大量卓有成效的辛勤劳动;湖北科学技术出版社领导和编辑室、市场部的同仁奔波四方做了大量组织协调工作,对此,编委会表示由衷的敬意与谢意!由于该套书编写、编辑工程浩大,参编单位与人员众多,加之编者的学术水平所限,在编写要求、写作风格、学术思想等诸多方面的协调难度较大,本套教材肯定存在一些不足与错漏,诚恳欢迎同行专家和广大读者批评指正,以便再版时修改与完善。

《高等院校经济管理类主干课程教材》编委会

2014年1月于武汉

前 言

统计学,是一门收集、整理和分析统计数据的方法科学,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。不论是国家宏观经济运行与管理还是公司企业微观经济运行与管理,不论是社会经济生产活动还是人们的日常消费活动,都会产生巨大的数据。要从巨量数据中提取有用的信息,找出数据中存在的数量关系和数量规律,必须借助和依赖于统计理论和方法。目前,所有学科研究都已认可了统计证据的重要性。无论从事自然科学研究还是人文社会科学研究,都需要统计学的知识和技术。统计学已成为各学科不可或缺的研究工具,统计知识已成为各类专业人员的必备知识。

为此,我们为了满足各方面对统计学知识的需要,我们特编写了这本《统计学》。本教材编写中总结了我们长期的教学经验,并参阅了国内外许多同类的优秀教材,是所有编写人员共同劳动的结晶。教材的出版极大地方便了统计教学活动,同时必将进一步促进统计学教学活动的开展。

本教材在编写中力求体现以下几个特点:

(1)内容体系完整。本教材的内容涵盖了描述统计和推断统计,系统地介绍了统计数据的收集与整理、数据分布特征的描述、概率与概率分布、抽样与参数估计、假设检验、方差分析、相关与回归分析、时间数列分析、统计指数、非参数统计等统计理论与方法。

(2)阐释说明简单。尽管统计学是一门理论性很强的学问,本书尽量做到深入浅出,对已有的理论叙述尽量简洁,含义清晰。

(3)案例导入为先。读者学习统计学,其目的在于了解统计思想,运用统计方法分析实际问题。本教材以案例导入为特点,使本教材介绍内容更加自然。

(4)统计软件相辅。目前常用的统计软件有 SAS、SPSS、STATA、STATISTICA、EViews、MINITAB 等,此外,Office 办公系统中的 Excel 软件也具有基本的统计计算功能。本教材鼓励学生使用统计软件,并在教材中加以介绍。

本教材由朱喜安教授任主编,张敏、班蕾老师任副主编。本书各章编写的分工如下:第一、四、五、六章由张敏老师编写;第二、三、八、九章由班蕾老师编写;第七章由朱喜安、王晓娟老师编写,第十章由朱喜安老师编写。稿件编写完成以后,朱喜安教授审阅和总纂全部书稿。

本书的出版得到了各方面的支持和帮助。感谢华中科技大学武昌分校阎德玉教授,本书的出版是他亲自关怀和支持的结果。感谢湖北科学技术出版社李慎谦编审,是他对全书精心筹划,才使本书得以顺利出版。

本书编写过程中,多次开会讨论写作大纲、内容和基本规范,力图奉献给读者一本较满意的教材,但由于统计理论和方法处于不断更新与发展之中,加之作者的水平有限,仍然难以达到各个方面的要求。书中疏漏或错误之处,恳请同行和读者多提宝贵意见,以便我们进一步修改和完善。

编 者

2014 年 1 月

目 录

第一章 导论	(1)
第一节 什么是统计学	(1)
第二节 统计数据的基本类型	(4)
第三节 统计学的产生与发展	(5)
第四节 统计学的分科及与其他学科的关系	(7)
第五节 统计学中的几个基本概念	(9)
思考与练习	(13)
第二章 统计数据的收集、整理与展示	(14)
第一节 统计数据的收集	(15)
第二节 调查设计	(23)
第三节 统计数据的整理与展示	(31)
第四节 统计表	(44)
思考与练习	(47)
第三章 数据分布特征的测度	(49)
第一节 集中趋势——数值平均数	(50)
第二节 集中趋势——位置平均数	(57)
第三节 离散程度的测度	(63)
第四节 偏态和峰度的测度	(70)
思考与练习	(72)
第四章 抽样与抽样分布	(75)
第一节 三种不同性质的分布	(75)
第二节 大数定律与中心极限定理	(76)
第三节 常用抽样分布及其特点	(77)
第四节 常用统计量的抽样分布	(82)
思考与练习	(87)
第五章 参数估计	(89)
第一节 参数估计的基本问题	(89)
第二节 一个总体参数的区间估计	(91)
第三节 两个总体参数的区间估计	(97)
第四节 样本容量的确定	(105)
思考与练习	(107)
第六章 假设检验	(111)
第一节 假设检验的基本问题	(111)
第二节 一个总体参数的检验	(117)

第三节 两个总体参数的检验·····	(125)
第四节 非参数假设检验·····	(142)
思考与练习·····	(150)
第七章 方差分析·····	(154)
第一节 方差分析的基本原理·····	(154)
第二节 单因素实验数据的方差分析·····	(157)
第三节 双因素实验数据的方差分析·····	(164)
思考与练习·····	(170)
第八章 相关与回归分析·····	(172)
第一节 相关分析·····	(173)
第二节 一元线性回归分析·····	(179)
第三节 多元线性回归分析·····	(191)
思考与练习·····	(198)
第九章 时间数列分析·····	(204)
第一节 时间数列分析概述·····	(205)
第二节 时间数列的水平分析·····	(208)
第三节 时间数列的速度分析·····	(217)
第四节 时间数列的影响要素及模型·····	(222)
第五节 时间数列趋势分解与预测·····	(223)
第六节 季节变动与循环变动分解与预测·····	(229)
思考与练习·····	(233)
第十章 统计指数·····	(237)
第一节 统计指数概述·····	(237)
第二节 综合指数法·····	(239)
第三节 平均指数法·····	(245)
第四节 指数体系和因素分析法·····	(251)
思考与练习·····	(263)
附表 A 统计表·····	(267)
表 A1 正态分布表·····	(267)
表 A2 t 分布的临界值表·····	(270)
表 A3 χ^2 分布的上侧临界值(χ_α^2)表·····	(272)
表 A4 F 检验的临界值(F_α)表·····	(275)
表 A5 多重比较中的 Duncan 表·····	(279)
表 A6 由百分率坐标变为概率坐标表·····	(282)
表 A7 相关系数检验表·····	(284)
参考文献·····	(287)

第一章 导 论

学习目的与要求

本章主要学习统计学的基本涵义,要求了解统计学的产生与发展历史,熟悉统计数据的基本类型、统计学的应用领域,理解统计学的学科性质,掌握统计学中的几个基本概念。

【导入案例】

众所周知,《红楼梦》一书共 120 回,自从胡适作《红楼梦考证》以来,一般都认为前 80 回为曹雪芹所写,后 40 回为高鹗所续,然而长期以来这种看法一直都饱受争议。从 1985 年开始,复旦大学的李贤平教授带领他的学生从统计角度作了考证。一般认为,同一情节大家描述的都差不多,但由于个人写作特点和习惯的不同,所用的虚词(之、其、或、亦、呀、吗、咧、罢、可、便、就……)出现的次数(频率),作为《红楼梦》各个回目的标志,利用统计方法果然能将 120 回分成两类,即前 80 回为一类,后 40 回为一类,很形象地证实了《红楼梦》不是出自同一人的手笔。之后又进一步分析前 80 回是否为曹雪芹所写。这时又找了曹雪芹的其他著作,做了类似计算,结果证实了用词手法完全相同,断定前 80 回为曹雪芹一人手笔,而后 40 回不是高鹗一个人所写。这个论证在红学界轰动很大,使红学界大为赞叹。

——《红楼梦》作者考证

(<http://www.qintai.net/forum.php?mod=viewthread&tid=6254>)

国家教育部把统计学列为经济管理类大学本科学生的专业核心课程。为什么要学习统计学?什么是统计学?统计数据有哪些类型?统计学与其他学科之间有什么关系?在这一章中,我们将学习一些统计学的基本问题。

第一节 什么是统计学

一、统计与统计学的含义

“统计”一词由来已久,其含义在历史上是不断发展和变化的。“统计”最早源自中世纪拉丁语“status”,意思是指各国的国家结构和国情知识的总称。“统计”最早作为学名使用是在 1749 年,德国哥丁根大学政治学教授 G. Achenwall 将课程“国势学”定为“statistik”(统计)。此后,各国相继沿用“统计”一词,并将其译为各国文字,法国译为 statistique,意大利译为 statistica,英国译为 statistics。该词不断被赋予新的内容并逐渐传播到各国,在 20 世纪初由日本传入我国。“统计”一词成为记述国家和社会状况数量关系的总称。

“统计”一词在各种实践活动和科学研究领域中经常出现。然而,不同的人或在不同的场合,对其理解存在差异。归纳起来,一般有三种涵义: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活动,是指收集、整理和分析统计数据,并探索数据的内在数量规律性的工作过程。例如,为了获得粮食产量而进行的抽样调查活动,为了获得全国人口的数量和构成等而进行的人口普查活动等。

统计实践活动有很久远的历史,早在公元前 2 000 多年,统治者为了征兵、征税、管理奴隶的需要,就有了人口、土地、财产的统计活动。《通典》记载了我国历史上最早的统计数据:“禹平水土,为九州”;“九州之地凡 24 388 024 顷,人口 13 553 923 人”。到中世纪,西欧各国都有了人口、军队、领地、财产等的统计活动。

2. 统计数据

统计数据又称为统计资料,是通过统计工作所获得的能够说明现象特征的相关数据,是统计实践活动的成果。统计数据的表现形式有统计图、统计表及有关分析报告。具体载体有统计公报、统计年鉴、统计手册及各种纸质和电子出版物等。

3. 统计学

统计学是处理统计数据的一门科学。统计学家们给统计学下的定义众多,例如,“统计学是收集、分析、表述和解释数据的科学”;“统计学是一门收集、分析、解释和提供数据的科学”;“统计学是一门科学,它研究怎样以有效的方式收集、整理、分析带随机性的数据,并在此基础上对所研究的问题作出统计学推断,直至对可作出的决策提供依据或建议”。

总结上述定义,统计学的含义概括如下:

统计学是一门研究收集数据、整理数据、表现数据、分析数据和解释数据,从而认识现象数量规律的方法论科学。统计学源于统计实践活动,是对统计实践活动的理论概括和总结,又用于指导统计实践活动。

统计工作、统计数据和统计学之间有着密切联系。统计工作与统计数据是过程与成果的关系,即统计数据是统计工作的成果。统计学与统计工作是理论与实践的关系,即统计学是统计工作的经验总结与理论概括,统计学对统计工作具有指导作用。统计工作、统计数据和统计学相互依存、相互联系,共同构成了一个整体,即通常所说的统计。

二、统计学的性质

统计学就在于研究客观现象数量方面的规律性并为其提供科学的方法,其核心内容是统计数据的收集、整理、描述、分析的原理和方法,因此,统计学也被称为“数据的科学”。这些方法论构成了统计学的科学体系,所以统计学是一门认识客观现象总体数量特征和数量关系的方法论科学。统计学具有数量性、总体性和具体性三个特征。

1. 数量性

所谓“数量性”,指统计学是一门定量科学,通过分析客观现象的数量方面进而认识客观事物及其发展变化的“本质”。统计学是数据的科学。它的研究对象是客观事物的数量方面。它的本质是分析数据,认识客观事物数量特征。使用的方法是数量分析方法。

2. 总体性

统计学是研究总体现象规律的科学。它通过对大量事物进行观察,或对一个事物的变化多次观察,才能得出反映总体的数量特征、反映事物必然性的结论。这是因为客观现象的个别现象通常有其偶然性、特殊性,而现象总体则具有相对的普遍性、稳定性,是有规律可循

的。虽然统计研究是从个别现象数据调查开始,对个别单位的具体事实进行调查研究,但其目的是为了认识总体的数量特征。例如,城镇居民调查,虽然是对每户居民进行调查,但目的不在于研究个别居民户的家计状况,而是通过大量的调查来反映一个城市、一个地区、一个国家的居民收入水平、收入分配、消费水平、消费结构等。

3. 具体性

统计学的研究对象是客观现象某一具体事物的数量方面,而不是像数学那样研究抽象的“纯数量”。客观现象的具体事物,都是在一定时间、地点、条件下的数量表现,它总是与时间、空间、事物紧密地联系在一起,具体地、历史地描述客观现象的发展过程,由此反映其本质和规律性。当然,由于统计学是研究客观现象总体的数量特征及关系的科学,因而它也要遵循数学的一些基本规律,并运用许多相应的数学方法进行运算及统计分析。

三、统计学的应用领域

统计方法是适用于所有学科领域的通用数据分析方法,只要有数据的地方就会用到统计方法。目前,随着定量研究的日趋重要,统计方法已被应用到自然科学和社会科学的众多领域,统计学已发展成为由若干分支学科组成的学科体系。可以说,几乎所有的研究领域都要用到统计方法,比如政府部门、学术研究领域、日常生活中、公司或企业的生产经营管理中都要用到统计。

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地了解 and 把握这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中的一项重要应用。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。比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,并作为企业制定工资计划、奖惩程度的依据。

第二节 统计数据的基本类型

统计数据是对客观现象进行测量的结果。任何现象都有其属性或数量表现,现象之间都有内在的关系,这些属性或数量表现及内在关系的表现,就是统计数据。例如,对经济活动总量的测量可以得到国内生产总值(GDP)的数据;对股票价格变动水平的测量可以得到股票价格指数的数据。由于使用的测量尺度不同,统计数据可以分为不同的类型。按所采用的计量尺度不同,可以将统计数据分为分类数据、顺序数据和数值型数据。按统计数据的收集方法,可以将其分为观测数据和试验数据。按被描述的现象与时间的关系,可以将统计数据分为截面数据、时序数据和混合数据。

一、分类数据、顺序数据、数值型数据

1. 分类数据

分类数据(categorical data)是按照现象的某种属性对其进行分类或分组而得到的反映事物类型的数据,又称定类数据。例如,按照性别将人口分为男、女两类;按照经济性质将企业分为国有、集体、私营、其他经济等。“男”、“女”,“国有”、“集体”、“私营”和“其他经济”就是分类数据。为了便于计算机处理,通常用数字代码来表述各个类别,比如,用1表示“男性”,0表示“女性”,但是1和0等只是数据的代码,它们之间没有数量上的关系和差异。

分类数据没有优劣、大小、顺序之分,只能区分事物的同类或不同类,所以其数学特性是“=”或“ $\neq$ ”,其计算功能是能够计算每一个类别的次数和比重、众数和异众比,进行列联表分析和卡方检验等。

2. 顺序数据

顺序数据(rank data)是按照现象的某种属性对其进行有等级差异或顺序差异的分组或分类得到的反映事物顺序类别的数据,又称定序数据。例如,将满意程度分为非常满意、满意、基本满意、不满意、非常不满意五类;将产品等级划分为一等品、二等品、三等品和次品。

利用顺序数据不仅可以研究对象分成不同的类别,而且还可以反映各类的优劣、量的大小或顺序。所以顺序数据的数学特性除了“=”或“ $\neq$ ”外,还有“ $>$ ”或“ $<$ ”。但是不能具体测量类别之间的差异,不能具体说一个类别大于或小于另一个类别多少,其计算功能除了分类数据的功能外,还可以计算中位数、分位数、等级相关等,但是不能进行加、减、乘、除运算。

分类数据和顺序数据统称为品质数据或定性数据,它是统计数据中的重要组成部分。品质数据或定性数据的统计分析具有独特性和专有性。随着“大数据”时代的来临,统计数据的构成越来越复杂,品质数据或定性数据在所有数据中比重会不断增加,分析方法会越来越复杂。

3. 数值型数据

数值型数据(metric data)是使用自然、物理和货币等单位,以及以上单位的复合单位对客观现象进行测量的结果,其结果表现为具体的数值。目前,现实中我们所处理的大多数都是数值型数据。数值型数据又可细分为定距数据、定比数据。

定距数据是对现象类别或顺序之间的间距进行的测度得到的数据,又称为区间数据、间隔数据。定比数据是对现象进行观测计数或计算得到的数据。

二、观测数据、试验数据

1. 观测数据

通过调查或观测而收集到的数据,称为观测数据(observational data)。观测数据是在没有对现象人为控制的条件下而得到的,有关社会经济现象的统计数据几乎都是观测数据。

2. 试验数据

在试验中控制试验对象而收集到的数据,称为试验数据(experimental data)。比如,对一种新药疗效的试验数据,对一种新的农作物品种的试验数据。自然科学领域的大多数数据都是试验数据。

三、截面数据、时序数据、混合数据

1. 截面数据

截面数据(cross-sectional data)是在同一时间不同空间上的数据,它所描述的是现象在某一时刻或某一时间段不同空间的数量变化状况。如 2012 年我国各省国内生产总值数据就属于截面数据。

2. 时序数据

时序数据(time series data)是同一空间在不同时间上的数据,它所描述的是现象随时间而变化的情况。如 1992—2012 年我国历年国内生产总值数据就属于时序数据。

3. 混合数据

混合数据(panel data)又叫面板数据,是不同空间在不同时间上的数据。如 1992—2012 年历年我国各省国内生产总值数据就属于面板数据。

第三节 统计学的产生与发展

自从有了国家,便有了统计实践活动。统计实践活动的出现已有几千年的历史,由于社会生产的发展和管理的需要而产生和发展的。我国在原始社会末期,在奴隶制形成的过程中,就已经出现了统计实践活动的萌芽。由于封建社会生产力十分低下,经济落后,统计发展缓慢,还处于对事物原始调查登记和简单计数汇总分析的阶段。资本主义生产方式的建立,使社会生产力水平得到不断提高,包括人口、工业和农业在内的“国情普查”逐步成为制度。18 世纪末,西欧各国相继建立了专门的统计机构负责全国的统计工作,并定期举行人口普查和专门调查。

统计学创始于 17 世纪中叶,经历了以下时期,产生了众多的统计学派。

一、古典统计学时期

古典统计学时期,是统计学产生和萌芽时期。约自 17 世纪中叶至 19 世纪初,其代表学派是“国势学派”、“政治算术学派”和“概率论学派”。

1. 国势学派

国势学派又称记述学派或国情学派,该学派用记述的方法研究一国的地理、人口、财政、军事、政治和法律制度等国家大事。其创始人是德国人康令(H. Corning, 1606—1682)。

1660年以后,康令开始定期地、系统地用对比的方法讲授国家之间比较方面的知识,他不仅讲述事实,而且试图探讨事实的因果关系。他将这个课程叫做《欧洲最近国势学》,于是“国势学”由此产生。当时,康令的学说在学术界产生了很大的影响,德国大学中的许多教授都称赞并追随他的学说。其中法国人戈特弗里德·阿亨瓦尔(G. Achenwall, 1719—1772)发扬了康令的学术思想,将这门课程定名为“统计学”,开始有了“统计学”这个名称。但是国势学派只是对各国情况作一般性的比较记述,如“某国人口众多”、“土地辽阔”之类,而没有进行数据分析和描述,与当代统计学的“数据分析”主流特征不符。所以,国势学派有统计学之名而无统计学之实。

2. 政治算术学派

“政治算术学派”的创始人是英国人威廉·配第(William Petty, 1623—1687),配第首先使用“数量”科学地研究阐述了社会经济现象。在他的名著《政治算术》序言中,配第明确指出:“我进行这项工作所使用的方法,在目前还不是常见的。因为和只使用比较级或最高级的词语以及单纯作思维的论证相反,我却采用了这样的方法,即用数字、总量和尺度的词汇来表达我自己想说的内容,只进行诉诸人们感官的论证和考察在性质上有可见的根据的原因。”他还说:“用数字、重量、尺度来表示的展望和论旨,都是真实的,即使不真实,也不会有明显的错误。……因为,能够证明为确实的东西,也就是确实的。”他在研究社会经济现象的规律时,还应用推算法、分组法,编制了原始数据的图表,计算了一系列的重量指标、相对指标和平均指标。政治算术学派的另一代表人物是英国人约翰·格朗特(J. Graunt, 1620—1674),格朗特于1662年出版了《关于死亡公报的自然和政治观察》一书,发表了对人口出生率研究的结果,并观察到一切疾病和事故在全部死亡原因中占有稳定的百分比等。政治算术学派虽未以“统计学”命名其学说,但其所用于探索社会和经济现象数量规律性的方法却具有“统计学”之实,因此,后人也将政治算术学派看作是统计学产生和萌芽时期的重要统计学流派之一。

3. 概率论学派

概率论学派以法国的学者为主。这一学派最早起源于对赌博中投骰子输赢问题的研究。其创始人是法国人B. Pascal(1623—1662)和P. de Fermat(1601—1665),他们以通信方式讨论赌博时的概念问题。同一时期,瑞士数学家I. Bernoulli(1654—1705)提出二项分布理论。此后,法国人P. S. Laplace(1748—1827)在1814年发表《概率分析论》一书,构筑了古典概率理论的完整体系,并用于自然和社会现象的研究。法国人S. D. Poisson(1781—1840)提出Poisson分布。德国人K. F. Gauss(1775—1855)提出最小平方法,还从观察天象中发现误差正态曲线。这一时期,概率论学派与其他统计学流派似乎并无交集,他们独立发展,各显特色。

二、近代统计学时期

近代统计学时期,约自19世纪初至19世纪末。代表学派主要有数理统计学派和社会统计学派。

1. 数理统计学派

数理统计学派继承概率论学派的主要理论遗产,将概率论的一些思想运用于自然、社会及其他各个领域,形成了数理统计学派。数理统计学派是当时主要的统计学学派之一,较早的代表人物是比利时著名的统计学家、数学家、物理学家、天文学家和人类学家阿道夫·凯

特勒(A. J. Quetelet, 1796—1874)。他融会贯通各家各派的统计思想,博采众长,在统计研究方面,先研究天文、气象方面的统计资料,后又用统计数字研究植物界和人类社会,他以概率论作为理论基础,用大量观察和综合平均的方法进行研究,从而把概率论、国势学派和政治算术学派观察群体现象进行数量分析的方法,融合为一门统计学,奠定了近代统计学的基础。凯特勒是古典统计学的终结者,近代统计学的先驱者,在统计学发展史上起着承前启后、继往开来的作用。

2. 社会统计学派

社会统计学派产生于 19 世纪末期,该学派主要继承了国势学派和政治算术学派的部分有益的学术遗产。首创者是德国人克尼斯,主要代表人物有梅尔、恩格尔,他们认为统计学的研究对象是社会现象,研究方法是大量观察法,提出统计学是一门实质性的社会科学。

三、现代统计学时期

现代统计学时期一般认为是从 19 世纪末 20 世纪初开始,主要继承和发扬数理统计学派的学术成就。其标志是推断统计学的进一步发展和精确小样本理论的产生。英国生物学家、统计学家卡尔·皮尔逊(K. Pearson, 1857—1936)在 1894—1895 年提出包括正态分布、矩形分布、J 型分布、U 型分布等 13 种曲线及其方程式,这一研究成果打破了以往次数分布曲线的“唯正态”观念,推进了次数分布曲线理论的发展和应用,为大样本理论奠定了基础;他还提出了卡方检验和标准差概念,发展了相关回归理论等。此外,英国人戈塞特(W. S. Gosset, 1876—1937)创立了 t 分布,开创了小样本理论的先河。英国人费歇尔(R. A. Fisher, 1890—1962)提出 Z 分布、显著性水平、假设检验、自由度、实验设计和方差分析等方法 and 概念。1930 年,尼曼和小皮尔逊共同对假设检验理论作了系统的研究,创立了“尼曼—皮尔逊”理论。美国统计学家瓦尔德(A. Wald, 1902—1950)把统计学中的估计和假设理论予以归纳,创立了“决策理论”。这些研究和发现大大充实了现代统计学的内容。

20 世纪 60 年代以后,随着计算机技术和网络技术的不断完善和各种新技术的不断创新,统计学的发展有如下趋势:首先统计学从面对小批量的数据转变为面对海量数据、“大数据”,因此使用计算机统计分析软件对数据进行处理成为必然,统计学及其应用与计算机技术的联系越来越紧密;其次,统计学从有关领域中吸取的养分也越来越多,如卫星技术的发展催生了空间统计学及其他统计学的应用边缘学科的发展。越来越多的数学方法被引进来,又被越来越多地应用到各个领域,如医学界的新药研制,企业中的过程控制,以及金融资产定价及风险统计分析等。第三,统计学所研究的方法可以用于各行各业的数据分析,这使得它几乎成为一门“万能”的方法论学科,统计数据的范围越来越广泛。美国著名期刊“Science”上有一篇文章列出近百年来最有用的科学,统计学位居前 10 名。

第四节 统计学的分科及与其他学科的关系

一、统计学的分科

统计方法已广泛应用于自然科学和社会科学的众多领域,统计学也发展成为由若干分支组成的学科体系。出于不同的视角或不同的研究重点,人们常对统计学科体系做出不同

的分类。一般而言,有两种基本的分类:从方法的功能来看,统计学可以分成描述统计学和推断统计学;从方法研究的重点来看,统计学可分为理论统计学和应用统计学。

1. 描述统计学和推断统计学

描述统计学研究如何取得反映客观现象的数据,并通过图表形式对所收集的数据进行加工处理和显示,进而通过综合、概括与分析得出反映客观现象的规律性数量特征。描述统计学的内容包括统计数据的收集方法、数据的加工处理方法、数据的显示方法、数据分布特征的概括与分析方法。

推断统计学研究如何根据样本数据去推断总体数量特征,它是在对样本数据进行描述的基础上,对统计总体的未知数量特征作出以概率形式表述的推断。

描述统计学与推断统计学的划分,还反映了统计方法发展的前后两个阶段和使用统计方法探索客观事物数量规律性的不同过程。统计研究过程的起点是统计数据,终点是探索出客观现象内在的数量规律性。在这一过程中,如果收集到的是总体数据(例如普查数据),那么运用描述统计就可以达到认识总体数量规律性的目的;如果获得的只是研究总体的一部分数据(样本数据),那么要找到总体的数量规律性,就要运用概率论的理论,并根据对样本特征的描述,对总体进行科学的推断。因此,描述统计是整个统计学的基础,推断统计则是现代统计学的主要内容。推断统计在现代统计学中的地位越来越重要,已成为统计学的核心内容,这是因为在对现实问题的研究中,所获得的数据主要是样本数据。但这并不等于说描述统计不重要,没有描述统计收集可靠的统计数据并提供有效的样本信息,再科学的统计推断方法也难以得出切合实际的结论。从描述统计学发展到推断统计学,是统计学发展的巨大成就,也是统计学成熟的重要标志。然而,随着“大数据”时代的来临,全面统计数据增多,描述统计也将进一步发挥其作用。

2. 理论统计学和应用统计学

理论统计学,主要探讨统计学的数学原理和统计公式的来源。理论统计学是统计方法的理论基础,没有理论统计学的发展,统计学也不可能发展成为一个完善的科学知识体系。概率论是统计推断的数学和理论基础,所以广义的统计学也应包括概率论在内。统计学中一系列的统计方法及其数理原理,构成了理论统计学的基本内容。

应用统计学探讨如何运用统计方法去解决实际问题。其中,将理论统计学的原理应用于各个学科领域,就形成了各种各样的应用统计学。例如,统计方法在医学中的应用形成了医疗卫生统计学,在生物学中的应用形成了生物统计学,在农业试验、育种等方面的应用形成了农业统计学。统计方法在社会科学领域的应用也形成了若干分支学科,在经济领域的应用形成了经济统计学及其若干分支,在管理领域的应用形成了管理统计学,在社会学研究和社会管理中的应用形成了社会统计学,在人口学中的应用形成了人口统计学,等等。应用统计学除了包括各领域通用的方法,如参数估计、假设检验、方差分析等之外,还包括某领域所特有的方法,如经济统计学中的指数法、现代管理决策法等。应用统计学着重阐明这些方法的统计思想和具体应用,而不是统计方法数学原理的推导和证明。

二、统计学与其他学科的关系

1. 统计学与哲学的关系

哲学是人类认识世界的最一般的方法论科学,统计学既然是认识社会经济现象的科学,

必然以哲学作为它的方法论基础。统计学必须在质与量的密切联系中认识事物的本质和规律性。统计学的研究方法和原理必须遵循事物认识的一般规律。

2. 统计学与经济学的关系

经济学是一门实质性科学,其任务在于揭示经济现象发展变化的规律,以指导人们按照客观经济规律的要求去改造和认识世界。经济学在揭示社会经济现象规律时,必须运用大量统计学的基本方法和原理,处理和分析经济领域中的统计数据,以统计规律为指导。同时统计学的重要分支——经济统计学是以社会经济现象为认识对象的方法论科学,就必须以科学的经济理论为指导。

3. 统计学与数学的关系

统计学在研究现象的数量关系和数量特征时,日益依靠数学和概率论基本原理和方法,需要数学和概率论提供的原理和方法进行严密的定量分析。例如,在社会经济现象中,有许多数量关系可以利用一定的数学模型来研究。这种模型,有些表现为函数关系,也有不少表现为随机性的统计关系。例如统计学中的平均指标、标志变异指标、抽样调查、相关分析和回归分析及统计预测等很多都需要数学原理作支撑,都是数学方法在统计学研究中的应用。因此,数学已成为现代统计和统计理论方法研究的重要基础。

第五节 统计学中的几个基本概念

统计学中的概念众多,其中有几组概念经常用到,熟悉这几组概念是掌握统计学的基础。

一、总体、个体与样本

(一) 总体

1. 含义

总体(population)是在一定的研究目的下,所要研究事物的全体,它是由客观存在的、具有某种共同性质的众多个别事物构成的整体(本章的“总体”指实物总体,在推断统计学中,称随机变量为总体)。例如,要研究某地区民营企业生产经营状况,则该地区全部民营企业就构成了一个总体,调查设计、调查方案、数据处理与分析都要围绕这一对象来进行;要研究我国的人口状况,则全国人口就构成了统计总体,从普查方案的设计、普查登记、资料汇总到公布普查数据,都要围绕总体进行。

2. 特点

总体具有如下几个特点:

一是同质性,是指构成总体的总体单位在某一方面性质是相同的,只有性质相同的人、单位、物等才能集合在一起,研究其数量表现和数量联系才有意义,同质性是构成总体的基础。

二是大量性,是指构成总体的总体单位必须足够多。总体单位是总体数量特征最原始的承担者,总体的数量特征很多时候是无法直接观测到的,只能通过对总体单位的数量特征进行观测得到。而总体单位的数量特征可能各不相同,没有规律可循,只有对大量总体单位的数量特征进行综合,才能体现总体的数量特征。因此,大量性是构成总体的条件。

三是总体具有差异性,是指构成总体的总体单位在某一方面性质是相同的,而在其他方面都是不尽相同的。例如,关于某地区民营企业生产经营状况,如果该地区每一家民营企业的生产经营情况都一样,我们就无需总体了,只要了解一家民营企业就知道了该地区所有的民营企业的生产经营情况。因此,差异性构成是构成总体的前提。

3. 分类

按构成总体的总体单位数是否有限,总体可以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,而且总体单位的数目是有限可数的。比如,由若干个企业构成的总体就是有限总体;再如全国人口普查,总体单位多达十几亿人,但它也是有限的,是有限总体。无限总体是指总体所包含的总体单位是无限的。比如,在抛硬币试验中,每一次抛掷硬币的结果可以看成是一个总体单位,而试验可以无限地进行下去,由该试验数据构成的总体就是一个无限总体。社会经济现象绝大多数是有限总体,而在统计推断中,往往将总体视为无限总体。

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,采用不放回抽样时,每次抽取一个单位,并不会太多地影响下一次的抽样结果,因此每次抽取可以近似地看作是独立的。对于有限总体,采用不放回抽样时,抽取一个单位后,总体元素就会减少一个,前一次的抽样结果会影响后一次的抽样结果,因此每次抽取样本是不独立的,将会影响到抽样推断的结果。

(二)个体

构成总体的个别事物是个体(individual)或总体单位,它是构成总体的基本元素。个体是所要研究具体问题的承担者。在统计调查中,常常将总体称为调查对象,将个体称为调查单位。个体是统计研究的起点和基本单位。

(三)样本

样本(sample)是从总体中抽取的一部分个体所构成的子集合,又称作子样。构成样本的个体的数目称为样本容量。从总体中随机抽取一部分个体作为样本,目的是要根据样本提供的有关信息去推断总体的特征。

二、参数和统计量

1. 参数

总体参数(parameter)是出现在总体分布中某个未知常数,用来刻画总体特征的量称为参数。参数是研究者想要了解的总体的某种特征值或者是这些特征值的反应。比如总体平均数用 μ ,总体标准差用 σ ,总体比例用 π 等等均为总体参数。

由于总体数据通常是不知道的,所以参数是一个未知的常数。比如,不知道某一地区所有人口的平均身高,不知道一个城市所有家庭的消费支出情况,等等。正因为如此,才进行抽样调查,根据样本统计量去估计总体参数。

2. 统计量

样本统计量(statistic)是不包括任何未知总体参数,用来描述样本特征的函数称为统计量。统计量是根据样本数据计算出来的一个量,它是样本的函数。比如,样本平均数用 $\bar{x}$,样本标准差用 s ,样本比例用 p 等等都是不包含总体未知参数的统计量。