

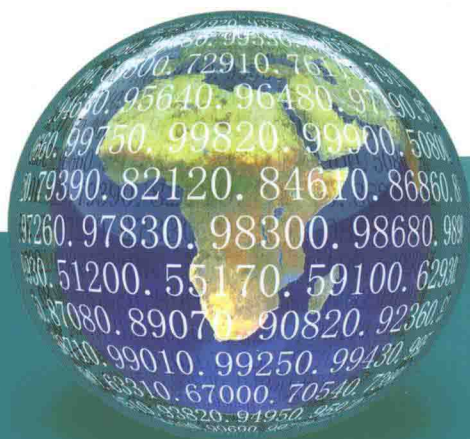
普通高等学校“十三五”规划教材

经管核心课

应用统计学

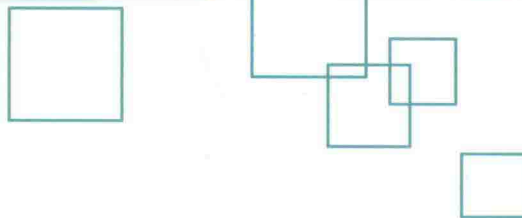
(第4版)

施金龙 吕洁 施然 编著



YING YONG TONG JI XUE

2012 2013 2014 2015



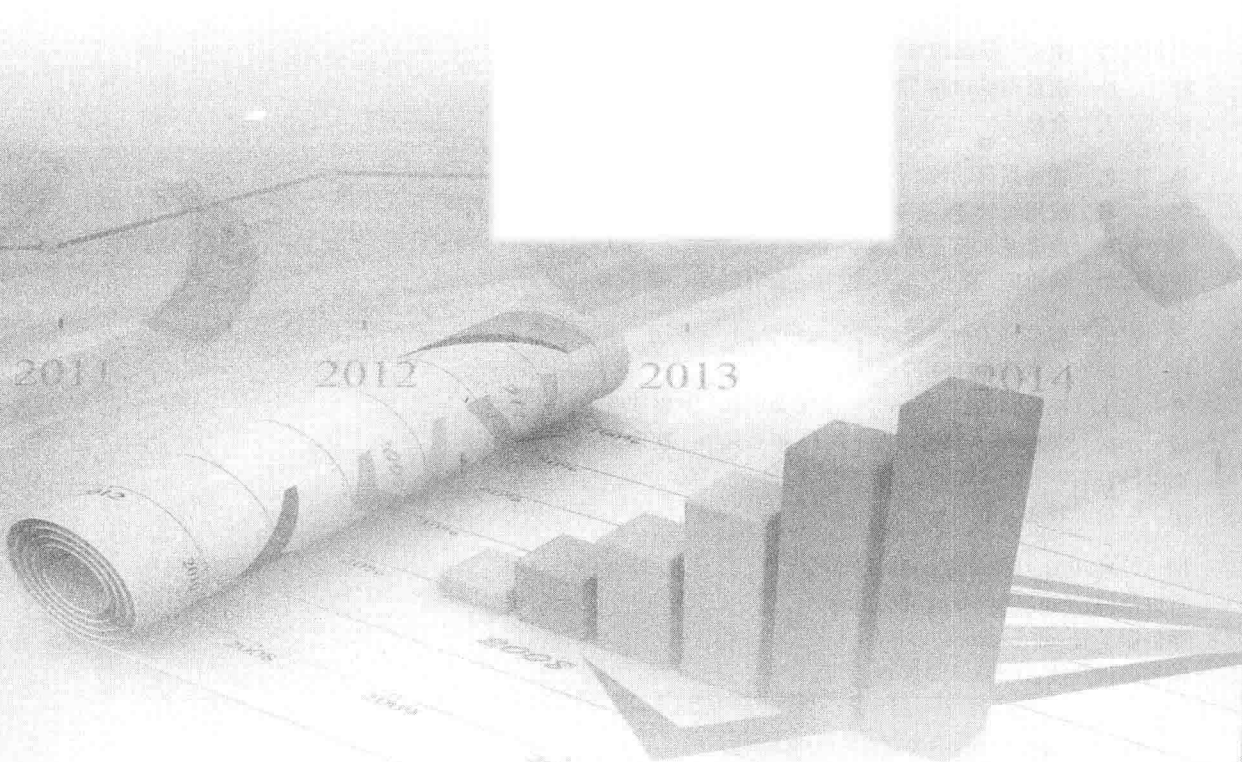
南京大学出版社

普通高等学校

应用统计学

(第4版)

施金龙 吕洁 施然 编著



南京大学出版社

图书在版编目(CIP)数据

应用统计学 / 施金龙, 吕洁, 施然编著. —4 版

—南京: 南京大学出版社, 2016. 6

普通高等学校“十三五”规划教材. 经管核心课

ISBN 978-7-305-17024-9

I. ①应… II. ①施… ②吕… ③施… III. ①应用统计学—高等学校—教材 IV. ①C8

中国版本图书馆 CIP 数据核字(2016)第 120221 号

出版发行 南京大学出版社

社 址 南京市汉口路 22 号 邮 编 210093

出 版 人 金鑫荣

丛 书 名 普通高等学校“十三五”规划教材·经管核心课

书 名 应用统计学(第 4 版)

编 著 施金龙 吕 洁 施 然

责任编辑 何永国 编辑热线 025-83686596

照 排 南京理工大学资产经营有限公司

印 刷 南京人民印刷厂

开 本 787×1092 1/16 印张 20.75 字数 596 千

版 次 2016 年 6 月第 4 版 2016 年 6 月第 1 次印刷

ISBN 978-7-305-17024-9

定 价 42.00 元

网 址: <http://www.njupco.com>

官方微博: <http://weibo.com/njupco>

微信服务号: njyuxue

销售咨询热线: (025)83594756

* 版权所有, 侵权必究

* 凡购买南大版图书, 如有印装质量问题, 请与所购
图书销售部门联系调换

第 1 版序

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、通用性方法论学科。从统计方法的构成来看,统计学可以分为描述统计学和推断统计学;从统计方法的研究与应用来看,统计学可以分为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,它主要研究统计学的一般理论和统计方法的数学理论。应用统计学是指统计学的专业应用,它主要研究如何应用统计原理和方法去解决各专业领域的实际问题。本书紧密结合企业管理和经济管理的实际问题,阐述统计原理和方法及其应用。

本书按篇幅精简、内容丰富、适用面广的原则编写,是一本经济类、管理类的非统计专业的通用性、导论性教材。各章配有习题,并附有参考答案;在部分章节中,穿插补充 Excel 的统计应用。

本书由施金龙任主编,吕洁任副主编;王晓晖、王云、黄雪丽参加部分章节的编写。

编者学习、吸收了大量统计学著作和相关文献的理论成果。在此,向这些著作和文献的作者,表示我们的深深谢意。

编 者

2005 年 10 月

第 2 版序

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、通用性方法论学科。从统计方法的构成来看,统计学可以分为描述统计学和推断统计学;从统计方法的研究与应用来看,统计学可以分为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,它主要研究统计学的一般理论和统计方法的数学理论。应用统计学是指统计学的专业应用,它主要研究如何应用统计原理和方法去解决各专业领域的实际问题。本书紧密结合企业管理和经济管理的实际问题,阐述统计原理和方法及其应用。

本书按篇幅精简、内容丰富、适用面广的原则编写,是一本经济类、管理类的非统计专业的通用性、导论性教材。各章配有习题,并附有参考答案。在部分章节中,穿插补充 Excel 的统计应用。

本书第 1 版 2005 年出版以来,受到多方专家、教师和学生的好评。2007 年 4 月、9 月,本书分别获得“江苏科技大学精品教材”“镇江市哲学社会科学优秀成果三等奖”等荣誉。

本书第 2 版主要修正了第 1 版中的遗漏和错误,改写了第 1 版中的部分过时的统计数据和习题资料。施金龙、吕洁、王晓晖、黄雪丽教师和李绍丽、张艳同学参加了本书改版工作。

为方便教学,我们编写了随书赠送的配套教学辅助材料(光盘),内含教学课件和模拟试卷。我们真诚期待广大教师和学生,就本书使用及统计教学中的相关问题,与我们交流、切磋。我们热忱欢迎广大教师和学生,查阅我们的网页、发送邮件到我们的电子邮箱:<http://202.195.195.69/lesson/teacher/mainteacher.jsp,jkdsjl@yahoo.com.cn>。

编者学习、吸收了大量统计学著作和相关文献的理论成果。在此,向这些著作和文献的作者,致以我们的深深谢意。

编 者
2008 年 3 月

第3版序

统计思维是现代人必须具备的素质之一。正如英国学者威尔斯(H. G. Wells)所说:统计思维,如同读写能力一样,总有一天会成为讲求效率的公民所必需的本领。教育部也一直将《统计学》列为经济和管理类大学本科教育的核心基础课程。

本书《应用统计学》定位于经济类、管理类的大学本科非统计专业开设“统计学”“统计学原理”“应用统计学”“管理统计学”等课程的通用性、导论性教材。本书按篇幅精简、内容丰富、适用面广的原则编写。各章配有习题,并附有参考答案。在部分章节中,穿插补充 Excel 的统计应用。

本书 2008 年再版以来,受到多方专家、教师、学生等读者的广泛认可和好评,销量不断增加。这给予了我们极大的鼓励,也使我们更有信心对本书不断改进和完善。

本书第 3 版主要修正了第 2 版中的遗漏和错误;改写了第 2 版中的部分过时的统计数据 and 习题资料;增设了“知识链接”栏目(留给课程教学和读者阅读的广阔知识空间);补充了各章的习题。

本书第 3 版的改版工作,有幸邀请江西农业大学南昌商学院三位教师加盟。参加改版的这三位教师和江苏科技大学四位教师的角色分工是:施金龙和吕洁任主编;潘求丰(江西农业大学)和杨君(江西农业大学)任副主编;施然、王育霞和杨娟(江西农业大学)参与编写。

为方便教学,我们编写了随书赠送的配套教学辅助材料(光盘),内含教学课件和模拟试卷。我们真诚期待广大师生和读者,就统计教学及本书使用中的相关问题,与我们交流、切磋。我们热忱欢迎广大师生和读者,查阅我们的网站(<http://202.195.195.21>),点击“网络教学”;发送邮件到我们电子邮箱(jkdsjl@yahoo.com.cn)。

编者学习、吸收了大量统计学著作、相关文献及网络资料的理论成果。在此,向这些著作、文献和资料的作者,致以我们的深深谢意。

编 者

2012 年 6 月

第 4 版序

本书第 4 版对第 3 版各章做了修正、补充和完善。第一,以“十二五”我国经济社会发展的最新统计数据,更新了第 3 版中的部分过时的统计数据和习题资料。第二,较大幅度地补充了第 3 版部分章节的习题,补齐了各章单选题和多选题。第三,在保留第 3 版各章“知识链接”内容的基础上,增加了“英文选读”栏目(作为学生课外阅读和教师双语教学的可选素材)。第四,删节了第 3 版中一些不必要的数学证明。

为方便教学,我们编写了随书赠送的配套教学辅助材料,内含教学课件和模拟试卷。我们热忱欢迎广大师生和读者选用、阅读本书,真诚期待与各位交流本书使用及统计学教学中的相关问题。交流邮箱:JSKDSJL@163.com。

编者学习、吸收了大量统计学著作、相关文献及网络资料的理论成果。在此,向这些著作、文献和资料的作者,致以我们的深深谢意。

编 者

2016 年 3 月

目 录

第 1 版序	1
第 2 版序	1
第 3 版序	1
第 4 版序	1
第一章 绪 论	1
第一节 统计学的产生和发展	1
第二节 统计学的对象与方法	3
第三节 统计学的基本概念	5
第四节 统计工作的任务和职能	8
习 题	12
第二章 统计调查	15
第一节 统计调查概述	15
第二节 统计调查方案	17
第三节 统计调查体系	26
习 题	33
第三章 统计整理	35
第一节 统计整理概述	35
第二节 统计分组	36
第三节 分布数列	43
第四节 统计资料汇总	50
第五节 统计图表	51
习 题	61
第四章 综合指标	66
第一节 总量指标	66
第二节 相对指标	70
第三节 平均指标	77
习 题	94
第五章 变异与均衡指标	98
第一节 变异指标	98
第二节 偏度指标	108

第三节	峰度指标	111
第四节	均衡指标	114
第五节	Excel 在统计描述中的应用	118
习 题		124
第六章	时间数列	127
第一节	时间数列概述	127
第二节	时间数列指标	130
第三节	时间数列分析	140
第四节	时间数列预警	151
第五节	Excel 在时间数列分析中的应用	156
习 题		160
第七章	指 数	164
第一节	指数概述	164
第二节	指数计算	166
第三节	指数分析	175
第四节	指数数列	182
习 题		187
第八章	抽样分布	193
第一节	抽样概述	193
第二节	抽样设计	197
第三节	抽样分布	200
习 题		209
第九章	参数估计	211
第一节	参数估计概述	211
第二节	总体平均数的区间估计	213
第三节	总体比率的区间估计	219
第四节	总体方差的区间估计	221
第五节	样本容量的确定	223
习 题		226
第十章	假设检验	229
第一节	假设检验的基本问题	229
第二节	总体平均数的假设检验	233
第三节	总体比率的假设检验	239
第四节	总体方差的假设检验	242
第五节	Excel 在统计推断中的应用	245
习 题		253

第十一章 方差分析	256
第一节 方差分析概述	256
第二节 单因素方差分析	257
第三节 双因素方差分析	262
第四节 Excel 在方差分析中的应用	267
习 题	270
第十二章 相关分析	274
第一节 相关分析概述	274
第二节 相关系数	275
第三节 回归方程	284
第四节 相关与回归	291
第五节 Excel 在相关与回归分析中的应用	293
习 题	299
附录一 常用统计表	303
附表 1 标准正态分布表	303
附表 2 t 分布表	304
附表 3 χ^2 分布表	306
附表 4 F 分布表	308
附录二 各章习题(部分)参考答案	312
主要参考书目	320

第一章 绪 论

那些默默无闻的统计学家们已经改变了我们的世界,不是发现了新的事实或技术,而是改变了我们推理和试验的方法,以及我们对这个世界的观念的形成方式。

哈克英

统计,原本是统而计之、合而计之的意思。后来,译自英语 statistics 的“统计”一词延伸为包含下面三种含义:统计工作,从事具体的数字资料搜集、分析的实践活动;统计数据,统计工作的成果——数字资料及相关信息;统计学,统计工作实践的理论总结和提炼,一门方法论科学。本章介绍统计学、统计数据及统计工作的一般问题。

第一节 统计学的产生和发展

人类的统计活动,特别是搜集社会经济数字资料的活动,在中国可以上溯到夏、商时代,在外国可以追及古埃及、古罗马时期。而统计学的历史,最早也只能从 17 世纪算起。统计学的产生和发展,大体上经历了三个阶段。

一、古典统计学

古典统计学指的是 17 世纪中末叶至 18 世纪中末叶萌芽时期的统计学。差不多同时,在德国和英国分别产生了“国势学”和“政治算术”,史称“国势学派”和“政治算术学派”。

1. 国势学派

所谓国势学是指以文字来记述国家的显著事项的学说(故国势学派又称记述学派)。最早在德国大学里讲授国势学的是康令(H. Conring)。康令的后继者阿亨华尔(G. Achenwall)讲授国势学时,于 1749 年把“国势学”称为“Statistika”,即“统计学”。

国势学派在研究各国的显著事项时,主要是用对比分析的方法研究关于国家组织、人口、军队、领土、财产等基本国情、国力,用以比较各国实力的强弱。国势学的研究方法侧重文字描述、性质解释,缺乏数量、数值分析。

2. 政治算术学派

这一学派是以英国人威廉·配第(W. Petty)的著作《政治算术》(1671 年写出,1690 年正式出版)命名的。这里的“政治”是指政治经济学,“算术”是指统计方法。《政治算术》运用大量实际资料,对英、法、荷三国的国情国力作了系统的数量对比分析,阐明英国的国际地位并不悲观。他用“数字、重量、尺度来说话”的方法,奠定了统计学的方法论基础。对此,马克思评价他是“政治经济学之父,在某种程度上也可以说是统计学的创始人。”

政治算术学派的另一创始人是约翰·格朗特(J. Graunt),他的代表作《对死亡率公报的自然观

察和政治观察》与《政治算术》齐名。当时,伦敦人口死亡情况严重,格朗特对此作了分类计算和分析,揭示出人口出生率、死亡率、性别比例等人口现象的某些规律性事实。

政治算术学派的人物及著作,一直未正式使用“统计学”这一命名。因此,被称为无统计学之名却有统计学之实。这恰与国势学相反。一般认为,政治算术是统计学的正宗起源。

国势学所研究的是历史学的组成部分,政治算术也还未从政治学中分化出来。它们都属实质性的社会科学。

二、近代统计学

近代统计学指的是18世纪末到19世纪末期间发展起来的数理统计学派和社会统计学派及其学说。

1. 数理统计学派

1869年比利时人凯特勒(A. Quetelet)发表了《社会物理学》。他的社会物理学是“要给政治科学和精神科学附加上一种以观察和计算为基础的方法,而支配着社会现象的法则和方法则是概率论。”他把作为数学分支的概率论引入统计学,使统计方法在“政治算术”所建立的“算术”基础上,在准确化的道路上大大跨进一步。凯特勒不仅是近代统计学的先驱者,也是数理统计学的奠基人。因为数理统计学就是在概率论基础上,经过许多英美统计学者的努力,发展起来的一门兼有数学和统计学双重性质的新生科学。因此,英美统计学者将凯特勒誉为“近代统计学之父”,数理统计学派又称英美数理统计学派。

2. 社会统计学派

19世纪中叶,正当英美数理统计刚刚开始发展的时候,在德国兴起了社会统计学派。

社会统计学派由德国人克尼斯(K. G. A. Knies)首创,主要代表人物还有德国人恩格尔(C. L. E. Engel)和梅尔(G. V. Mayr)。他们认为统计学是一门社会科学,是研究社会现象变动原因和规律性的实质性学科。社会统计学派一方面研究社会总体现象(不是个别现象),另一方面在研究方法上采用大量观察法,这两方面构成这一学派的两大特点。

社会经济的发展,社会科学不断细化和定量化,要求统计学提供更多的搜集资料、分析资料的方法。因此,社会经济统计学逐步从实质性科学向方法论科学转化。

三、现代统计学

现代统计学指自20世纪初至今的推断统计学。

统计学是一种以随机抽样为基础,推论(估计)总体数量特征的方法体系,导源于英国人戈塞特(W. S. Gosset)的小样本理论,后经费希尔(R. A. Fisher)、尼曼(J. Neyman)和皮尔逊(E. S. Pearson)等人而发展。

20世纪50年代以后,统计学的发展呈现三个明显的特点: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,如多元统计分析,探索性数据分析、现代时间序列方法等;其三是统计研究与电子计算机应用密切结合。

我国于20世纪初由日本传入统计学。新中国建立前,基本上是介绍、研讨英美数理统计学。新中国建立后,学习苏联的社会经济统计学(苏联学者定义的一门研究社会经济规律的实质性社会科学)。1978年以后,出现了是一门统计学还是两门统计学的争论。20世纪90年代以来,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,尽管分支繁多,但它们的基本理论和基本方法都是共同的。

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、通用性方法

论学科。从统计方法的构成来看,统计学可以分为描述统计学和推断统计学;从统计方法的研究与应用来看,统计学可以分为理论统计学和应用统计学。

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搜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。而推断统计学则是研究如何根据样本数据去推断总体数量特征的方法,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做出以概率形式表述的推断。描述统计是推断统计的前提,推断统计是描述统计的发展;描述统计是整个统计学的基础,推断统计是现代统计学的核心。

理论统计学是指统计学的数学原理,它主要研究统计学的一般理论和统计方法的数学理论。应用统计学是指统计学的专业应用,它主要研究如何应用统计原理和方法去解决各专业领域的实际问题。本书紧密结合企业管理和经济管理的实际问题,阐述统计原理和方法及其应用。

知识链接 1:“统计”词语的产生

统计已经有几千年的历史。不过在早期还没有出现“统计”这样的用语。

统计语源最早出现于中世纪拉丁语的 Status,意思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 18 世纪德国政治学教授阿亨华尔在 1749 年所著《近代欧洲各国国家学纲要》一书绪言中,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。“统计”一词,就成了记述国家和社会状况的数量关系的总称。

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(清光绪二十九年)钮永建、林卓南等翻译了横山雅南所著的《统计讲义录》一书,把“统计”这个词从日本传到我国。1907 年(清光绪三十三年)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。

第二节 统计学的对象与方法

由以上统计学的发展史来看,统计学是从研究社会经济现象数量特征开始的。随着统计方法的不断发展和丰富,统计学的面貌发生了根本变化。时至今日,多数学者认为:统计学是在质与量的辩证统一中,通过大量观察和平均分析,对自然、社会客观总体现象的数量规律性进行研究的方法论科学。

一、统计学的对象

统计学的研究对象一直是统计学研究中的争论问题之一。学者们的分歧集中于三点:其一是认为统计学与统计工作的研究对象是一致的,两者均以实质性的社会科学规律为对象,称为实质性科学派;其二是认为统计学是专门研究关于社会经济总体现象的方法论的科学,强调统计方法的对象性,称为特定对象方法论派;其三是认为统计学研究的总体现象的数量特征和规律性,既适用于自然现象,也适用于社会现象,称为通用方法论派。

我们持通用方法论观点。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,使人们有可能透过对客观总体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,而这种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具体给出。

统计学的研究对象是自然、社会客观现象总体的数量表现、数量界限和数量关系。统计学研究对象具有以下特点:

1. 数量性

统计学的研究对象是自然、社会经济领域中现象的数量属性,这一特点是统计学(定量分析学科)与其他定性分析学科的分界线。数量性是统计学研究对象的基本特点,因为,数字是统计的语言,数据资料是统计的原料。一切客观事物都有质和量两个方面,事物的质与量总是密切联系、共同规定着事物的性质。没有无量的质,也没有无质的量。一定的质规定着一定的量,一定的量也表现为一定的质。而从认识的角度,可以在一定的质的情况下,单独研究量,通过认识事物的量进而认识事物的质。

2. 总体性

统计学的研究对象是自然、社会经济领域中现象总体的数量属性,即统计的数量研究是对总体普遍存在着的事实进行观察和分析,认识现象总体的数量特征和规律性。自然、社会经济现象的数据资料和数量对比关系等,一般是在一系列复杂因素的影响下形成的。在这些因素当中,既有发挥决定和普遍作用的主要因素,也有发挥偶然和局部作用的次要因素。由于种种原因,在不同的个体中,它们相结合的方式和实际发生的作用都不可能完全相同。所以,对于每个个体来说,就具有一定的随机性质,而对于有足够多数个体的总体来说又具有相对稳定的共同趋势,显示出一定的规律性。例如,对企业员工工资的统计分析,并不着重分析和研究个别员工的工资数额,而是着重分析和研究一个企业总体的工资水平。

3. 具体性

统计学研究对象是自然、社会经济领域中具体现象的数量属性,这正是统计学与数学的分水岭。数学是研究事物的抽象空间和抽象数量的科学,而统计学研究的数量大多是客观存在的、具体实在的数量表现。统计数据作为主观对客观的反映,必然是存在第一性、意识第二性,存在决定意识。只有如实地反映具体的已经发生的客观事实,才能为统计分析、研究提供可靠的基础,才能分析、探索和掌握事物的统计规律性。

4. 变异性

统计学研究对象的变异性是指构成统计学研究对象的总体各单位,除了在某一方面保持同质以外,在其他方面存有差异,而且这些差异并不是由某种特定的原因事先给定的。假如高等院校作为统计学的研究对象,则它除了都是从事高等教育的教学活动这一共同性质之外,各高等院校在隶属主管部门、院校性质、招生规模、专业设置等各方面都有差异。企业员工作为统计学的研究对象,每个员工在性别、年龄、工龄、工作岗位、工资福利等方面也都会有不同表现。正是这种变异性,才使从观察和分析对象的差异表现中探索其特征及规律性的统计方法大有用武之地。

二、统计学的方法

统计学是采集数据和分析数据的方法论科学,是方法论的集合体。其中基本的方法有:实验设计法、大量观察法、平均分析法和归纳推断法。

1. 实验设计法

实验设计,就是对实验进行合理安排,以较小的实验规模(实验次数)、较短的实验周期和较低实验成本,获得理想的实验结果(实验数据)以及得出科学的结论。

英国统计学家费希尔(R. A. Fisher)总结在农业生产中使用实验设计方法(研究育种)的实践,于

1935 年出版了他的名著《实验设计》。书中提出了实验设计应遵循的三个原则:随机性、区组化和重复性。随机性是指在实验设计中,对实验对象的分配和实验次序都是按随机安排的。这种安排可以使可控的影响因素作用均匀化,突出不可控影响因素的作用。区组化是指利用类型分组技术,对实验对象按有关标志顺序排队,然后依次将各单位随机地分配到各处理组,使各处理组组内标志值的差异相对扩大,而处理组组间的差异相对缩小,以提高处理组的估计精度。重复性是指允许在相同条件下重复多次实验。重复实验是为了降低随机误差的影响,以保证实验结果的重现性。

2. 大量观察法

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、实验和调查,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。

政治算术学派的苏斯密尔斯(J. P. Süßmilch),被视为大量观察法的倡导者。他从人们的统计研究中得出结论:“事实若多一分,人事现象的规律则多发现一分……因此,不能用太少的事实,要尽可能地多,而且更好的是要尽可能包含更多的年份。”

3. 平均分析法

英国学者鲍莱(A. L. Bowley)早在 20 世纪 20 年代就指出:“统计学为平均法科学。”将客观总体现象的内部差异加以抽象,以一简明的数值(平均数)概括反映总体的典型的数量特征,这就是平均分析法。统计学中的其他许多分析方法,例如总体差异的测度、总体动态趋势分析、指数计算和回归分析等,都体现了平均分析思想和原理,都是平均分析方法的延伸和运用。

平均分析方法,也是认识经济规律的基本方法。马克思在《资本论》里,对于平均社会必要劳动量、平均利润率、平均地租等,无不以平均数作为规律的表现形式。

4. 归纳推断法

归纳推断法可细分为归纳法和推断法。归纳法是指由观察各单位的数量表现,总结得出有关全部单位的数量特征的从个别到一般的推理方法。推断法是指根据部分单位的观察结果,对全部单位的特征做出推论或估计的方法。

推断法是现代统计学的基本方法。它既可以用于对总体参数的估计,也可以用作对总体的某些假设进行检验,广泛应用于农业产量调查、工业产品质量检测和控制等方面的估计和检验。

第三节 统计学的基本概念

概念是人们对客观事物的不同方面分析归类而得出的基本名词。每门科学都有自己的特有概念,作为本学科的科学基础。概念之间的联系和发展引申出规律,规律的总和构成科学体系。因此,掌握一个学科的基本概念,是全面认识这门学科科学体系的基础环节。

统计学是研究客观总体现象数量特征的方法论,总体由总体单位组成,总体单位互有差异的特征通过标志以说明,标志经综合成指标,指标相互结合形成指标体系。因此,总体、总体单位、标志、指标和指标体系等,是统计学的基本概念。

一、总体,总体单位

总体就是统计研究对象的全体,它是由客观存在的、具有某种共同性质的许多个别事物构成的整体。简言之,总体就是同质个体集团。组成集团的个体,称为总体单位。

总体具有四个性质,即大量性、变异性、同质性和相对性。大量性是指总体内包含的总体单位有许许多多。如果总体是由个别或少数个体构成,就无从对其实施大量观察,就无法揭示总体的数量特征。变异性是指构成总体的各个总体单位之间互有差异。例如工业企业总体,其中的各个企业在

所有制形式、企业规模、行业、产值等方面都有差异。同质性是指构成总体的各个总体单位之间虽互有差异,但至少具备一种共同的性质。同质性是总体的前提,而变异性则是统计的前提。相对性是指总体与总体单位是相对于统计研究目的而言的。例如,在研究目的是研究机械行业发展时,某一机械工业企业为总体单位;在研究目的是研究某一机械工业企业发展时,该企业为总体。

二、标志,标志表现

总体内各个总体单位之间的差异(变异),可以从总体单位的某些属性或特征方面加以描述和说明。说明总体单位属性或特征的名称,叫作标志。前述的所有制形式、企业规模、行业、产值等,即为标志。

标志按其性质可分品质标志和数量标志。品质标志表明总体单位属性方面的特征,它不能用数值表示,只能用文字说明。数量标志表明总体单位数量方面的特征,它用各种不同的数值表示。按变异情况,标志可分不变标志和可变标志。当一个标志在各个总体单位的具体表现都相同时,这个标志称为不变标志。所谓总体的同质性,即总体单位至少具有一个不变标志。当一个标志在各个总体单位的具体表现有可能不同时,这个标志称为可变标志或变异标志。所谓总体的变异性,也即总体单位具有可变标志。不变的数量标志称为常量或参数,可变的数量标志称为变量。

标志表现是标志在各总体单位的具体体现。例如,品质标志“职业”的标志表现为工人、农民、教师等,数量标志“工龄”的标志表现为3年、5年、10年等。数量标志的标志表现称为标志值或变量值。

三、指标,指标体系

根据一定的统计方法对总体各单位的标志表现进行登记、核算、汇总、综合,就形成能说明总体某一数量特征的指标。指标这一概念,有两种理解:一是说明总体特征的名称,一是说明总体特征的名称和数值。例如,粗钢产量或粗钢产量8亿吨,都可以理解为一个指标。若将指标仅理解为名称,指标的具体数值就相当于指标的具体表现。

指标按其说明总体特征内容的不同,分为数量指标和质量指标。数量指标是说明总体外延量大小的指标,如总人口、国内生产总值、粗钢产量等。质量指标是说明总体内涵量多少的指标,如人口出生率、全员劳动生产率、人均国内生产总值等。指标按其作用功能不同,分为描述指标和分析指标。客观反映总体数量特征的指标称为描述指标;对描述指标加以提炼、综合、重组,用以评价、判别、分析总体数量规律的指标,称为分析指标。

一个指标说明总体某一方面的特征,多个指标才能说明总体的全面特征。说明总体全面特征(全面说明总体特征)的多个相互联系的指标组成一指标群,这个指标群称为指标体系。例如,一个工业企业的生产经营活动是人力、物质、资金、生产、供应、销售相互联系的整体运作过程,以一群指标说明和分析企业的全面情况,就形成工业企业统计指标体系;又如工业产品产量可以选用实物产量、定额工时产量、价值产量(总产值、净产值、增加值)指标加以统计描述,这就可以称为产量统计指标体系。而产值=产量×价格,总产量=亩产量×面积等,也可以称为指标体系。

根据统计所研究问题的范围大小,指标体系大体分为基本指标体系和专题指标体系两大类。基本指标体系是反映国民经济和社会发展及其各个组成部分的基本情况的指标体系。这类指标体系能反映社会生产和生活的全过程,其组成部分又可以分成经济、社会和科技三个子体系(子系统)。专题指标体系是针对某一个社会经济、工商管理研究课题而专门设计的指标体系。例如,经济效益指标体系,人民生活水平指标体系等。

四、数据,数据计量

统计数据是总体单位标志或统计指标的具体数量表现。要对客观现象进行计量,首先必须弄清

数据的计量尺度问题。根据对研究对象自身特点及对其计量的精确程度不同,美国社会学家、统计学家史蒂文斯(S. S. Stevens)提出了由低到高、由粗略到精确的四个层次计量尺度:定类尺度、定序尺度、定距尺度和定比尺度。

1. 定类尺度

定类尺度是最粗略、计量层次最低的计量尺度,它是按照客观现象的某种属性对其进行分类。作为定类尺度的数值只是作为各种分类的代码,并不反映各类的优劣、量的大小或顺序。例如,人口按性别分为男和女;本科学生按专业分为经济学、统计学、会计学等。定类尺度的主要数学特征是“=”或“ $\neq$ ”。在统计处理中,对于不同的类别,虽然可以计算单位数,但它不能表某一类的一个单位可以相当于另一类的几个单位。

2. 定序尺度

定序尺度是对客观现象各类之间的等级差或顺序差的一种测度。利用定序尺度不仅可以研究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。例如,本科学生可以分为大一、大二、大三和大四四类;学生成绩可以分为优秀、良好、中等、及格和不及格五类。在这里,定序尺度虽然无法表明大四学生(学到的知识)等于几倍大三学生(学到的知识),成绩优秀等于几倍成绩良好;但却能确切地表明大四学生高于大三学生、大三学生又高于大二学生,优秀高于良好、良好又高于中等。定序尺度的主要数学特征是“ $<$ ”或“ $>$ ”。

3. 定距尺度

定距尺度是对现象类别或次序之间间距的测度。定距尺度不但可以用数字表示现象各类别的不同和顺序大小的差异,而且可以用确切的数值反映现象之间在绝对量方面的差异。定距尺度没有真正的零点(绝对零点)。在定距尺度中,“0”表示某一个数值,并不表示“没有”或“无”。温度是典型的定距尺度,因为在摄氏温度中, 0°C 表示在海平面高度上水结冰的温度,并不表示没有温度。定距尺度的主要数学特征是“+”或“-”。

4. 定比尺度

定比尺度不仅具有定距尺度的全部性质,还能通过对比运算来反映现象之间在相对量方面的差异。定比尺度具有真正的零点(绝对零点)。在定比尺度中,“0”表示“没有”或“无”。例如,产品产量、生产成本、销售利润等都是定比尺度。定比尺度的主要数学特征是“ $\times$ ”或“ $\div$ ”。

由定类尺度计量形成的定类数据,表现为类别,但不区分顺序;由定序尺度计量形成的定序数据,也表现为类别,但有顺序;由定距尺度计量形成的定距数据,表现为数值,可进行加、减运算;由定比尺度计量形成的定比数据,也表现为数值,可进行加、减、乘、除运算。前两类数据说明的是事物的品质特征,不能用数据表示,其结果均表现为类别,可统称为定性数据或品质数据;后两类数据说明的是现象的数量特征,能够用数值来表现,可统称为定量数据或数值数据。

根据对客观现象观察的角度不同,统计数据可分为:横截面数据和时间序列数据。横截面数据又称为静态数据,它是指在同一时间对同一总体内不同单位的数量进行观察而获得的数据。例如,2015年我国城镇居民人均可支配收入31 195元,农村居民人均可支配收入11 422元。时间序列数据又称为动态数据,它是指在不同时间对同一总体的数量表现进行观察而获得的数据。例如,2015年、2014年、2013年我国城镇居民人均可支配收入31 195元、28 844元、26 955元,2015年、2014年、2013年我国农村居民人均可支配收入11 422元、10 489元、8 896元。

从广义上说,不仅可变数量标志可称变量,而且所有指标也可称变量。因此,统计数据也就是统计变量的具体表现。变量的具体表现即为变量数值或变量值。根据变量值连续出现与否,变量可分为连续型变量和离散型变量。连续型变量是指变量的取值在数轴上连续不断,无法一一枚举,即在一个区间内可取任意实数值。比如,人体身高、体重,天气温度、湿度,零件尺寸、寿命等。离散型变量是指变量的数值可以一一枚举,且通常取整数值。比如,企业数、设备数和员工数等。