

普通高等教育“十二五”规划教材

统计学

TONG J I XUE

主编◎王艳芳 黄永涛 朱越

辽宁大学出版社

普通高等教育“十二五”规划教材

统计学

主 编 王艳芳 黄永涛 朱 越

辽宁大学出版社

图书在版编目(CIP)数据

统计学 / 王艳芳, 黄永涛, 朱越主编. —沈阳 :
辽宁大学出版社, 2013.4 (2015.5重印)
普通高等教育“十二五”规划教材
ISBN 978-7-5610-7243-1

I. ①统… II. ①王… ②黄… ③朱… 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 081675 号

出版者:辽宁大学出版社有限责任公司
(地址:沈阳市皇姑区崇山中路 66 号 邮政编码:110036)
印刷者:北京明兴印务有限公司
发行者:辽宁大学出版社有限责任公司
幅面尺寸:185mm×260mm
印 张:16
字 数:350 千字
出版时间:2013 年 4 月第 1 版
印刷时间:2015 年 5 月第 3 次印刷
责任编辑:张琢石 黄 铮
封面设计:可可工作室
责任校对:齐 悦

书 号:ISBN 978-7-5610-7243-1
定 价:32.00 元
联系电话:86864613
邮购热线:86830665
网 址:<http://WWW.lnupshop.com>
电子邮件:lnupress@vip.163.com

前 言

统计学是一门通过收集、整理、显示和分析统计数据,以达到推断所研究对象的本质,甚至预测研究对象未来的一门综合性科学,其中用到了大量的数学及其他学科的专业知识。它的使用范围几乎覆盖了社会科学和自然科学的各个领域。社会经济管理工作者和科学研究人员应该具备扎实的统计理论和方法。因此,统计学是高等院校财经与管理类各专业学生必修的一门专业基础课程。

为了满足统计学课程教学的需要,在多年统计学教学经验积累的基础上,我们组织编写了这本《统计学》教材。本书系统地阐述了统计学的基本原理和方法,这些方法可用于对社会现象的数量方面的研究,也可用于对自然现象的数量方面的研究,它是各种应用统计学的共同基础。

本书具有以下几个方面的特点:

1. 简明易懂。本书力求强调统计思想,摒弃了不必要的数学证明和公式推导,做到通俗易懂、由浅入深、由表及里。在对各种具体统计方法作必要的阐述之后,均配有相应的具体案例或例题说明其应用过程,并以图形和表格的形式从直观上和逻辑上进行解释,便于读者理解,有利于教师教学。

2. 实践性强。本书在分析统计理论的基础上,在书后附有以 Excel 为统计计算和分析工具的实训教程,结合实例讲述如何应用 Excel 进行各种统计分析,使理论与实践相结合,增强了理论的实践性和可操作性。

3. 整体性强。本书是按照一个完整统计过程中的各个顺序着手编写的。一个完整的统计过程包括对统计资料的收集、整理、显示和分析,并借助这些资料进行推断的决策。其中加入的概率分布和抽样分布是实现描述统计过渡到推断统计的桥梁,使本书的整体框架更严密,统计思想表现得更清晰。

4. 节约导向。本书具备上述 3 个特点的要旨是为读者节约时间。它与一些教材“兼顾不同程度读者需要”不同,本书强调针对性和专业性,服务群体是高等院校财经与管理类各专业本科生、MBA 和相关专业研究生,目的是帮助他们完成课程学习。

本书编写的具体分工为:朱越编写第 1 章、第 2 章和第 3 章,王艳芳编写第 4 章、第 5 章、第 6 章、第 7 章、第 8 章和第 11 章,黄永涛编写第 9 章、第 10 章和实训教程部分。在编写过程中,参考了许多国内外的统计学教材和著作,在此向所有参考文献的作者和编者表示衷心的感谢。

尽管本书编者很想奉献给读者一本满意的教材,但仍然有可能在书中有不当或疏漏之处,恳请同行和读者提出宝贵意见,以便共同努力把统计学教材建设好。

编者

2013 年 2 月

目 录

第 1 章 绪论	(1)
1.1 统计与统计学	(1)
1.2 统计学与统计数据	(5)
1.3 统计学的基本概念	(9)
1.4 统计学的应用	(13)
第 2 章 统计资料的搜集与整理	(17)
2.1 统计资料的搜集	(17)
2.2 统计资料的整理与显示	(22)
2.3 统计表	(33)
第 3 章 数据的概括性度量	(35)
3.1 集中趋势的度量	(35)
3.2 离散程度的度量	(40)
3.3 偏态与峰态的度量	(45)
第 4 章 概率基础	(47)
4.1 概率分布	(47)
4.2 几种常见的离散型分布	(50)
4.3 几种常见的连续型分布	(51)
4.4 大数定律和中心极限定理	(56)
第 5 章 抽样与抽样分布	(57)
5.1 常用的抽样方法	(57)
5.2 抽样分布	(61)
5.3 抽样误差	(66)
第 6 章 参数估计	(68)
6.1 参数估计概述	(68)
6.2 一个总体参数的区间估计	(72)
6.3 样本容量的确定	(77)
第 7 章 假设检验	(81)
7.1 假设检验的基本原理	(81)



7.2	一个总体参数的假设检验	(86)
7.3	假设检验与参数区间估计的关系	(94)
第8章	相关分析与回归分析	(96)
8.1	相关分析	(96)
8.2	回归分析	(102)
8.3	应用相关分析与回归分析注意的问题	(113)
第9章	时间序列分析	(115)
9.1	时间序列概述	(115)
9.2	时间序列的水平指标分析	(117)
9.3	时间序列的速度指标分析	(123)
9.4	时间序列的构成与分析模型	(126)
9.5	平稳序列的平滑和预测	(128)
9.6	有趋势序列的分析和预测	(135)
9.7	定性预测方法	(146)
第10章	指数	(148)
10.1	指数概述	(148)
10.2	指数编制的基本方法	(150)
10.3	指数因素分析法	(156)
10.4	指数数列	(164)
10.5	几种常用价格指数的编制	(166)
第11章	统计分析	(173)
11.1	统计分析概述	(173)
11.2	统计分析方法	(174)
11.3	统计分析的步骤	(177)
11.4	统计分析报告	(178)
实训一	用 EXCEL 搜集与整理数据	(183)
实训二	用 EXCEL 作统计图与计算描述统计量	(192)
实训三	用 EXCEL 进行区间估计与假设检验	(209)
实训四	用 EXCEL 进行相关与回归分析	(218)
实训五	用 EXCEL 进行时间序列与指数分析	(227)
附录		(238)
参考书目		(249)

第1章 绪论

1.1 统计与统计学

1.1.1 “统计”的含义

在日常生活中“统计”有着多种含义。例如,企业管理人员要掌握生产销售情况和利润额;在生产车间要统计产品生产量;在高考录取中要统计考生的总分;球类比赛时解说员要统计竞赛双方的进攻次数和成功率;开会时主持人要统计一下出席会议的人数,等等。人们也常常从报刊杂志、电视新闻中获悉我国的经济增长速度、消费者价格指数、固定资产投资规模等经济数据资料。那么到底什么是“统计”呢?

请思考:下列资料中“统计”一词的含义是什么?

- (1)张宏是学统计的;
- (2)他已搞了几十年统计了;
- (3)据统计,今年一季度物价指数出现负增长;
- (4)请找统计登记一下;
- (5)请统计一下今天的销售量。

那么,把统计作为一种专业用语,其含义到底是什么?目前,关于统计一词的含义比较趋于一致的解释为:一是统计工作,二是统计数据,三是统计学。

统计工作,即统计实践,是对社会经济现象客观存在的现实数量方面进行搜集、整理和分析预测等活动的总称。如上所列举,要统计进攻次数、统计生产量、统计交易额等,就是统计工作。一个完整的统计工作过程一般包括统计设计、统计调查、统计整理、统计分析等环节。统计工作是统计一词最基本的含义,是人们对客观事物的数量表现、数量关系和数量变化进行描述和分析的一种计量活动。如银行的计划统计科,每月编制项目报表,这个过程就是统计工作。又如:我国进行人口普查时要经过方案设计、入户登记、数据汇总、分析总结和资料公布等一系列过程,这些都是统计工作。在我国,各级政府机构基本上都有统计部门,如统计局,它的职能主要是从事统计数据的搜集、整理和分析工作。

统计资料(统计信息)是统计工作过程中所取得的各项数字资料 and 与之相关的其他实际资料的总称。经济增长速度、价格指数等,就是统计资料。它和前面讲的统计工作是紧密相联的,是统计工作的结果,因而也是很早就有的。根据历史记载,我国夏禹时代就开始有人口统计数字。春秋时期,《商君书》中指出:强国知十三数,这十三数包括粮食储备、人口及其各项分类数、农业生产资料以及自然资源等,不过当时还没有明确叫做统计资料罢了。随着社会的发展,需要的统计数



字也就越来越多,现在只要打开报纸就可以看到各种各样的统计数字。国家统计局每年出统计年鉴,反映国家的经济、文化教育以及科技发展等情况,这些都是在这个意义上的统计。如:

(1)我国国土面积 960 万平方公里,其中山地约 320 万平方公里,高原约 250 万平方公里,平原约 115 万平方公里,丘陵约 95 万平方公里。

(2)2003 年我国全年全部工业增加值 53612 亿元,比上年增长 12.6%,其中规模以上工业企业(即国有工业企业及年产品销售收入 500 万元以上的非国有工业企业)增加值增长 17.0%,工业产品销售率 98.1%,比上年提高 0.1 个百分点。这些由文字和数字共同组成的数字化的信息就是统计资料,是统计提供数据信息的基本表现形式,是统计工作的直接成果。统计资料包括原始资料和整理后的资料即次级资料。例如企业各车间的统计台账、人口普查时初次登记的资料就是原始资料,而统计公报、调查分析报告等现实和历史资料就是次级资料。统计资料的表现形式有统计表、统计图、统计分析报告、统计公报和统计年鉴等。

除了上面所讲的两个含义之外,“统计”一词还有另外的含义,即作为一门科学的“统计学”,它是本书将要探讨的主要内容,作为一门科学的“统计学”,它的出现要比统计工作和统计资料晚得多。

1.1.2 统计学的产生与发展

统计起源很早,是随着社会生产的发展和国家管理的需要而逐步产生和发展起来的,距今已有四五千年的历史。而统计学或统计理论则是在长期统计实践活动基础上形成和发展起来的,距今只有 300 多年的历史。回顾一下统计的渊源及其发展过程,对于我们了解统计学的研究对象和性质,学习统计学的理论和方法,提高我们的统计实践和理论水平,都是十分必要的。

1. 统计实践的产生和发展

在原始社会,人类最初的一般计数活动蕴藏着统计萌芽。在奴隶社会,统治阶级为了对内统治和对外战争,需要征兵征税,开始了人口、土地和财产的统计。例如,公元前 3050 年,埃及建造金字塔,为征集建筑费,就有对全国的人口与财产调查;罗马皇帝恺撒·奥古斯都曾下过一道命令,要全世界向他纳税,于是每个人都要向就近的收税人登记;英国的威廉大帝下令测量英国的土地,其目的是为了征税和征兵役;我国在夏朝就有了关于人口和土地的数字记载:夏朝时分中国为九州,人口约 1355 万人,土地约 2438 万顷;春秋时期齐桓公任用管仲为相使齐国大治,在反映管仲思想的重要著作《管子》一书中就有这样的论述:“不明于计数,而欲举大事,犹无舟楫而欲经于水险也”,这就是说不善于利用计数而进行宏伟事业,犹如没有船和桨而想渡过激流险滩一样。

由于生产力水平所限,奴隶社会的统计只属于初级阶段。到了封建社会,统计有了一定的发展,封建君主和精明的政治家日益意识到统计对于治国强邦的重要性,统计范围有所扩大。但由于封建经济的封闭割据和保守性,统计活动的范围受到限制,统计方法也很不完善。到了资本主义社会,随着社会生产力的迅速发展和社会分工的愈益精细,统计得到了很大的发展,统计的应用领域越来越广泛,不仅仅只局限于经济管理领域,在军事、医学、生物、物理、化学等领域中也大量地运用统计方法。而电子计算机技术的应用为统计活动的现代化进程提供了重要手段。正是在这样的历史背景下,统计学应运而生。从 17 世纪中末期开始,经过 300 余年的发展,形成了今天的统计学。

2. 统计学的产生和发展

在统计学作为一门科学逐渐形成的过程当中,由于历史和社会的原因,形成了很多学派。

(1)国势学派。国势学派也称记述学派,是伴随德意志的兴盛而产生的。其代表人物是康令(H. Conring, 1606—1681)和阿亨瓦尔(G. Achenwall, 1719—1772)。国势学以叙述国家显著事项和国家政策关系为内容,并给出了统计学这一名词,但国势学主要用文字表述,缺乏数字内容,和现代统计学比较起来,有一些实不副名,即没有太多的实际内容符合统计学的名称。

(2)政治算术学派。政治算术学派产生于17世纪的英国,代表人物是威廉·配第(W. Petty, 1623—1687)和约翰·格朗特(John Graunt, 1620—1674)。威廉·配第著有《政治算术》一书,在书中提出“不用比较级、最高级进行思辨或议论而是用数字来表达自己的问题,借以考察在自然中有可见的根据的原因”。该书用数量分析的方法对比了英国、法国和荷兰三国的“财富和力量”,以批驳当时英国国内的悲观论调。在书中他用数字来表述,用数字、重量和尺度来计量,并配以直观的图表,为现代统计学提供了一个完美的开端。马克思称威廉·配第为“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”。

约翰·格朗特对英国伦敦人口的出生率和死亡率分别进行统计,并编制出世界上第一张“死亡率”统计表。他们的工作和使用方法,在现代统计学中,也是重要的,但遗憾的是,他们没有使用“统计学”一词,后人戏称“名不符实”,即没有用适当的名称称呼其实际的统计工作的内容。

国势学派和政治算术学派同时开始于17世纪中叶,直到19世纪末,从不同的国度、地域做了统计学的奠基工作。从统计学的内容上看,描述统计学的基本内容和方法已经给出;从时间上划分,这两个半世纪的统计学研究,又可称为“古典统计学”。并且随着微积分的完善、概率论的发展,为统计学注入新的内容,共同为推动统计学的形成打下了基础。

(3)数理统计学派。产生于19世纪中叶,创始人是比利时的天文学家、数学家和统计学家凯特勒(1796—1874),其著作有《统计学的研究》、《关于概率论的书信》等。他是当时统计学界的中心人物,担任过比利时中央统计局局长,主持过第一次国际统计会议(1853年),他最先将概率论应用于人口、人体测量和犯罪等问题的研究,完成了统计学和概率论的结合。从此,统计学开始进入更为丰富发展的新阶段。许多学者从各个角度研究统计学,不断增加新内容,相继提出和发展了相关和回归理论、t分布以及抽样理论等,使数理统计学很快发展成为一门比较系统、完善的学科。国际统计学界称凯特勒为“近代统计学之父”,就在于他发现了大量现象的统计规律和开创性地应用了许多统计方法,促使统计学向新的境界发展。由于这一学派主要在英美等国发展起来,故又称英美数理统计学派。

数理统计学派在理论上混淆了自然现象和社会现象之间的本质区别,过分夸大了概率论的作用,认为统计学就是数理统计学,是现代数学的一个分支,是通用于研究自然现象和社会现象的方法体系,否认社会经济统计学的存在,因而又导致了与社会经济统计学派的长期争论。

(4)社会经济统计学派。这一学派于19世纪后半叶兴起于德国,即原来政治算术意义下的统计学。但由于它在理论上比政治算术学派更加完善,在时间上比数理统计学派提前成熟,因而它很快占领了“市场”,对国际统计学界影响较大,流传较广。其主要代表人是恩格尔(1821—1896)和稍后的梅尔(1841—1925)。他们主张统计学是研究社会现象的社会科学。这一学派融会了记述学派和政治算术学派的观点,并把政府统计和社会调查融合起来,进而形成社会经济统计学。

数理统计学派与社会经济统计学派共存并争论至今已有100多年,目前,虽然数理统计学派在国际统计学界占据着优势,但二者已出现了融合的趋势。

统计发展史表明,统计学是从设置指标研究社会经济现象的数量开始的。随着社会的发展与实践的需要,统计学家对统计方法的不断丰富和完善,统计学也不断发展和演变。从当前世界各国统计研究状况来看,统计学已不仅为研究社会经济现象的数量方面,也为研究自然技术现象的数量方面提供各种统计方法;它既研究确定现象的数量方面,又研究随机现象的数量方面。从统计学的发展趋势来看,它的作用与功能已从描述事物现状、反映事物规律,向抽样推断、预测未来变化方向发展。它已从一门实质性的社会性学科,发展成为方法论性质的综合性学科。

1.1.3 统计学的定义

古往今来的统计学者对统计学给予了不同的定义。根据美国统计学家 David Freedman 等著的《统计学》(魏宗舒等译,中国统计出版社,1997 年版)中的定义:统计学是对令人困惑的问题作出数字设想的艺术。

根据《蓝登书屋大字典》(the random house college dictionary),统计学是“对用数字表示的事实或数据进行收集、整理、分类、分析以及解释的科学”。简而言之,统计学就是统计数据的科学。

统计学的定义众多,比较有代表性的是如下定义:

定义 1.1 统计学:统计学是收集、整理、显示和分析统计数据的科学,其目的是探索数据内在的数量规律性。

统计学的定义告诉我们,统计学是关于数据的科学,其内容包括数据收集、整理、显示和分析等。数据收集也就是取得统计数据,数据整理和显示就是将数据用图表等形式展示出来,数据分析则是通过统计方法分析数据,并对分析的结果进行说明。

把统计学称为艺术显然有些夸张,但这一说法的目的正在于提示统计工作者,应当创造性地提出和解决统计问题,不应局限于某些条条框框去理解统计这门科学。

案例:在一个水库中养着许多鱼,管理人员希望了解鱼的大致数量。这就是一个实践中的统计学问题。由于鱼是不听从指挥,会在各处自由游动的,因此,在进行统计时,必须创造性地提出解决方案。一种解决方法是先从水库的不同位置一共捕上来 1000 条鱼,在每条鱼的尾部作上一个标记,应当保证标记不会影响鱼的自由游动。然后,将鱼全部放回水库。几天后,从水库中再捕上来 2000 条鱼,检查其中尾巴上有标记的鱼的数目。假定在第二次捕上来的 2000 条鱼中,有 20 条尾巴上做了标记,则可以推断,水库中鱼的总数大致为

$$1000/(20/2000)=10 \text{ 万条}。$$

上述这个案例在实践中是经常见到的,对于一个统计工作者来说,作出一个 10 万条鱼的估计是不够的,他还应当对这一估计的精度作出判断。但这种搜集统计数据的方法,更多的是一种艺术,是很难从书本上学到的。

统计学是一门艺术。它提供一种归纳推理的方法,推理就是一种艺术。统计的应用方面是十分复杂的,只有将统计理解为一种艺术,创造性地提出新的方法去解决新的问题,才是真正地掌握了统计的精髓。既然是归纳推理,就不能保证结论百分之百正确,就不能没有争议。怎样让别人看懂并理解统计理论,就要看统计表达这些结论的技巧和艺术性了。

统计学是一门科学。它提供一套方法和技术,这些方法和技术并不是一成不变的,使用者在给定的情况下必须根据所掌握的专门知识选择使用这些方法,而且,如果需要还要进行必要的修正。统计方法是通用的数据分析方法,这些方法不是为某个特定的问题领域而构造的。

统计学是一种工艺。如同工业生产过程中的质量控制程序一样,统计方法是在为保证产品达到所希望的质量和保持其稳定性的管理系统中建立起来的。统计方法也能用于控制、减少和

考察不确定性。

统计学是科学、工艺和艺术这三者的组合。

统计学是研究大量社会现象(主要是经济现象)的总体数量方面的方法论的科学。这里所指的方法论包括指导统计活动的原理原则、统计过程所应用的核算和分析的方法以及组织方法。人们通过对社会现象中各种数量关系的研究来认识社会现象发展的规律性。值得注意的是,统计学在研究社会现象时,首先从定性研究开始,即在搜集原始统计资料(统计调查)之前,就要根据所要研究的对象的性质和研究任务、目的,确定调查对象的范围,规定分析这个对象的统计指标、指标体系和分组方法,这种定性工作是下一步定量分析的必要准备。在定量分析的基础上再达到认识社会经济现象的本质、特征或规律,这就是质—量—质的统计研究过程和方法。

1.2 统计学与统计数据

1.2.1 统计学的研究对象及特点

统计学的研究对象是指统计研究所要认识的客体。只有明确了研究对象,才可能根据它的性质特点指出相应的研究方法,达到认识对象客体规律性的目的。由统计学的发展史可知,统计学是从研究社会经济现象的数量开始的,随着统计方法的不断完善,统计学得以不断发展。因此,统计学的研究对象为大量现象的数量方面。

统计学的特点可以归纳为以下几个方面:

1. 数量性

社会经济统计学最基本的研究特点就是以数字为语言,用数字说话。具体地说,是用规模、水平、速度、结构和比例关系等,去描述和分析社会经济现象的数量表现、数量关系和数量变化,揭示事物的本质,反映事物发展的规律,推测事物发展的前景。

但应注意,统计学研究现象的数量方面,不同于数学上研究的纯数量,它不是抽象的数量,它是以现象质的规定性为基础的,是带有一定具体内容的数量。因为任何事物都是质和量的辩证统一,没有质也就没有量。

例如:要了解哈尔滨市重工业产值,首先要明确什么是重工业。所谓重工业是为国民经济各部门提供技术装备、动力和原材料的工业,包括采掘工业、原材料工业和制造工业。然后要确定重工业产值的含义和统计口径以及哈尔滨市哪些企业属于重工业企业,这些都是质的规定。在此基础上,还要解决怎样搜集、整理和汇总重工业产值资料,最后才能得到哈尔滨市重工业产值的具体数值。

2. 总体性

总体性又称大量性或综合性。统计研究的着眼点是大量社会经济现象总体,而不是少量或个别现象,它是通过对个别事物大量观察,占有丰富材料,加以分析综合,来反映现象总体的数量特征,揭示现象的本质和规律性。例如,2004年全年居民消费价格总水平比上年上涨3.9%,这个数量反映的是550多种消费商品及服务项目的价格总的平均上涨水平,而不是指哪一种具体消费商品或服务项目的价格上涨水平。而要对这550多种消费商品及服务项目的价格上涨情况进行调查,就必须先对每一种个别消费商品及服务项目的价格情况进行调查,然后进行汇总



综合,从而达到对 550 多种消费商品及服务项目价格的总体认识。

统计研究并不排除从个别现象入手,但统计研究个体是为了综合个体而认识总体,是手段而不是目的,其最终目的是要认识总体。例如,2000 年 11 月 1 日进行的第五次全国人口普查,逐一登记了全国大陆 31 个省、自治区、直辖市(不包括香港特别行政区、澳门特别行政区、台湾省)的每个人的性别、年龄等特征,但人口普查的目的并不是要了解关于某个人的特征,而是为了通过对全国人口情况进行汇总计算,得出关于我国人口总体的特征资料,从而达到对全国人口现象总体的认识。汇总后结果显示,祖国大陆 31 个省、自治区、直辖市(不包括福建省的金门、马祖等岛屿,下同)和现役军人的人口共 126583 万人。同第四次全国人口普查 1990 年 7 月 1 日 0 时的 113368 万人相比,十年零四个月共增加了 13215 万人,增长 11.66%。平均每年增加 1279 万人,年平均增长率为 1.07%。同 1990 年第四次全国人口普查相比,0—14 岁人口的比重下降了 4.80 个百分点,65 岁及以上人口的比重上升了 1.39 个百分点。从总体着眼,从个体入手,体现了统计工作中总体和个体之间的辩证关系。

3. 社会性

社会经济统计学通过研究大量社会经济现象总体的数量方面,来认识人类社会活动的条件、过程和结果,反映物质资料的占有关系、分配关系、交换关系以及其他的社会关系。其定量研究是以定性分析为前提的,而定性使其在客观上就有了社会关系的内涵。社会经济现象与自然科学技术问题是不同的,对于同一社会经济现象,站在不同的立场,持有不同的观点,运用不同的方法,可以得出差别较大的结论。这些都体现出社会经济统计活动的社会性。

4. 变异性

变异性又称差异性。统计学研究同类现象总体的数量特征,它的前提是总体各单位的表现存在着差异,而且这些差异并不是由某种固定的原因事先给定的。例如一个地区的居民人口有多有少,居民的文化程度有高有低,住户的生活消费水平有升有降,等等,正是各单位之间这种差异的存在,才需要研究地区的人口总数、居民文化结构、住户平均生活消费水平等统计指标。如果各单位不存在这些差异,也就无需进行统计,如果各单位之间的差异是按已知条件事先可以推定的,也就无需进行统计调查研究。

5. 具体性

统计研究的总体数量是一个有具体时间、具体地点、具体条件限定的数量。如利润额 800 万元,单独看来,它只是一个毫无意义的抽象数量。如果说 2004 年 12 月某企业利润额 800 万元,这就是统计中所说的具体数量了。可见具体性就是指在时间、地点、条件三方面有着明确的规定性。

统计工作虽然是研究具体的数量,但为了进行复杂的定量分析,还需要借助抽象的数学模型和数理统计方法,遵循一定的数学规则。以抽象方法为手段,以具体数量为目的,体现了统计研究中具体和抽象的辩证关系。

1.2.2 统计学的学科构成

统计学是一个复杂的体系,不同的学术机构对于这门学科的构成进行了不同的阐述。

美国数学学会出版的《数学评论》中对统计学的分类:

A. 基础;B. 充分性和信息;C. 决策理论;D. 抽样理论和抽样调查;E. 分布理论;F. 参数推

断;G. 非参数推断;H. 多元分析;I. 线性推断;J. 试验设计;K. 序贯分析;L. 随机过程推断;M. 工程统计学;N. 应用;O. 统计表。

从非统计专业的学生学习的角度来看,统计学可以分为四个大的组成部分:

1. 调查与实验设计

调查与实验设计涉及统计中获得原始数据的各种方法。调查是在社会经济统计中获得原始数据的主要手段。随着市场经济的发展,调查在经济活动中所起的作用越来越大,企业的经营,政府的决策,都离不开来自调查的第一手数据。

在科学研究过程中,获得统计数据的手段还包括实验方法。实验是在研究对象进行一定控制的情况下获得数据的方法。

2. 描述统计

描述统计包括整理、显示和分析数据的一系列方法。调查或者实验中所获得的有关事物整体的原始资料,往往是零乱和不系统的,需要经过一系列的统计处理,才能转化为人们可以直接阅读和理解的信息。这种针对事物整体数据的统计处理工作,被称为描述统计。

3. 推断统计

在有些情况下,人们获得的统计资料并非事物整体的状况,而是来自事物的一个局部。如果利用局部的数据去推断整体的情况,以及这种推断的有效性和可靠性如何,即是推断统计所要研究的内容。

4. 多元统计分析

在统计课程设计中,多元统计分析是一个独立的部分,主要涉及到对多变量情况的研究。例如,描述一个人的能力,需要从科研能力、动手能力、组织能力等多个方面进行综合判断,如果对涉及多个变量的统计问题进行研究,即为多元统计的内容。多元统计根据掌握信息的不同,也可分为多元描述统计和多元推断统计,但基本方法大多需要涉及到矩阵等工具,属于统计学中要求较高的部分。

1.2.3 统计数据

1. 数据的计量尺度

统计数据是对客观现象进行计量的结果。对客观现象进行计量,就必须弄清楚数的计量尺度问题。根据对研究对象计量的不同精确程度,将计量尺度由低到高、由粗略到精确分为四个层次:定类尺度、定序尺度、定距尺度和定比尺度。

(1)定类尺度。定类尺度亦称为列名尺度,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平行的分类。此时,若用数字表示,该数字仅作为各类的代码,度量各类之间的类别差,不反映各类的优劣、量的大小或顺序。例如,人口按性别分为男女,用“1”表示男性,用“0”表示女性。定类尺度的主要数学特征是“=”或“ $\neq$ ”。在统计处理中虽然可以计算单位数,但它不能表明第一类的一个单位可以相当于第二类的几个单位。

(2)定序尺度。定序尺度亦称为顺序尺度,它是对客观现象各类之间的等级差或顺序差的一种测度,是比定类尺度更高一级的计量尺度。定序尺度不仅可以研究将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,学生成绩可以分为优、良、中、及格和



不及格等五类,在这里,定序尺度虽然无法表明一个优等于几个良,但却能确切地表明优高于良,良又高于中。定序尺度的主要数学特征是“ $<$ ”或“ $>$ ”。在统计的变量数列中可以确定其中位数、分位数等指标的位置。

(3)定距尺度。定距尺度亦称为间隔尺度,它是对现象类别或次序之间间距的测度,是比定序尺度更高一级的计量尺度。定距尺度不但可以用数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量方面的差异。定距尺度使用的计量单位一般为实物单位(自然或物理)或者价值单位。反映现象规模水平的数据必须以定距尺度计量,例如,产品产量、人口数、企业数、国内生产总值、气象的温度湿度及各种试验数据都以定距尺度为计量尺度。定距尺度的主要数学特征是“ $+$ ”或“ $-$ ”。定距尺度在统计数据中,占据重要的地位,统计中的总量指标就是运用定距尺度为计量尺度。

(4)定比尺度。定比尺度亦称为比率尺度,它是在定距尺度的基础上,确定相应的比较基数,将两种相关的数加以对比而形成相对数(或平均数),用于反映现象的结构、比重、速度、密度等数量关系。例如,将一国的国内生产总值与该国的人口数对比,计算人均国内生产总值,以此反映该国的经济能力。定比尺度的主要数学特征是“ $\times$ ”或“ $\div$ ”。在统计的对比分析中,广泛地运用定比尺度进行计量。

2. 数据的类型

数据是对现象进行测量的结果。例如,对经济活动总量的测量可以得到国内生产总值(GDP)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,对人口性别的测量可以得到男或女这样的数据,等等。由于使用的测量尺度不同,统计数据可以分为不同的类型。下面从不同角度说明数据的分类。

(1)按照所采用的计量尺度不同,可以将数据分为分类数据、顺序数据和数值型数据。

定义 1.2 分类数据:只能归于某一类别的非数字型数据。

分类数据是对事物进行分类的结果,数据表现为类别,是用文字来表述的。它是由定类尺度计量形成的。例如,人口按性别分为男、女两类,企业按照经济性质分为国有、集体、私营、合资、独资企业等,这些均属于分类数据。为便于统计处理,对于分类数据,可以用数字代码来表示各个类别,例如,用1表示“男性”,0表示“女性”等。

定义 1.3 顺序数据:只能归于某一有序类别的非数字型数据。

顺序数据也是对事物进行分类的结果,但这些类别是有顺序的。它是由定序尺度计量形成的。例如,将产品分为一等品、二等品、三等品、次品等;考试成绩分为优、良、中、及格、不及格等;一个人对某一事物的态度分为非常满意、满意、保持中立、不满意、非常不满意等。同样,对顺序数据也可以用数字代码来表示,例如,1表示非常满意,2表示满意,3表示保持中立,4表示不满意,5表示非常不满意。

定义 1.4 数值型数据:按数字尺度测量的观察值。

数值型数据是使用自然或度量衡单位对事物进行测量的结果,其结果表现为具体数值。现实中处理的大多数都是数值型数据。

(2)按照数据的收集方法,可以将其分为观测数据和实验数据。

定义 1.5 观测数据:通过调查或观测而收集到的数据。

观测数据是在没有对事物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。



定义 1.6 实验数据:在实验中控制实验对象而收集到的数据。

例如,对一种新药疗效的实验数据,对一种新的农作物品种的实验数据。自然科学领域的大多数数据都为实验数据。

(3)按照被描述的对象与时间的关系,可以将统计数据分为截面数据和时间序列数据。

定义 1.7 截面数据:在相同或近似相同的时间点上收集的数据。

截面数据所描述的是现象在某一时刻的变化情况。例如,2012年我国各地区的国内生产总值数据就是截面数据。

定义 1.8 时间序列数据:在不同时间上收集到的数据,称为时间序列数据。

时间序列数据所描述的是现象随时间而变化的情况,例如2003年至2012年我国的国内生产总值数据就是时间序列数据。关于时间序列,本书将在第9章中进行详细介绍。

3. 数据的表现形式

统计数据通常表现为:绝对数、相对数和平均数。

(1)绝对数。现象的规模、水平一般以绝对数形式表现,例如,国内生产总值、人口数、进出口额等。绝对数的计量单位一般为实物单位或价值单位,有时也采用复合单位。实物单位可以是自然计量单位,也可以是物理计量单位,如人口数用人计量,粮食产量用吨计量,耕地面积用公顷计量等,对于一些化工产品,常常折合成为标准实物单位。价值单位是以货币形式进行计量,如国内生产总值、进出口总额是以价值单位为计量单位。复合计量单位是由两种或两种以上计量单位复合而成的,如以“吨公里”为货物周转量的计量单位,以“千瓦时”为用电量的计量单位。

绝对数按其反映的时间状态不同,分为时期数据和时点数据。时期数据是反映现象在一段期间内发展过程的总量,它具有连续统计和可加性的特点,其数值大小与所属的时间长短有直接关系,如国内生产总值、进出口总额。时点数据是反映现象在某一特定时点所处的状态,它是采用间断登记方式取得资料的,不具有可加性,其数值大小与时点间隔长短没直接关系,如期末人口数、期末在建工程投资额等。

(2)相对数。相对数是由两个绝对数对比而得的,常用的相对数有:结构相对数、动态相对数、比较相对数、比例相对数、强度相对数、利用程度相对数、计划完成相对数等。相对数的计量单位大部分是无名数,但也有一些是采用有名数为计量单位。把对比基数抽象为100而计算的相对数为百分数,把对比基数抽象为1000而计算的相对数为千分数,这些都是无名数。如果把某地区的人口数与该地区的土地面积对比所计算的相对数是一种强度相对数,被称为人口密度相对数,其计量单位为“人/平方公里”,这就是有名数。

(3)平均数。统计平均数是用于反映现象总体的一般水平或分布的集中趋势,数值平均数是由总体标志总量对比总体单位数而计算的。关于这部分内容,本书将在第3章中进行介绍。

1.3 统计学的基本概念

1.3.1 总体和样本

1. 总体

定义 1.9 总体:总体是由客观存在的、具有某种共同性质又有差别的许多个别单位所构



成的整体,当这个整体作为统计研究对象时称统计总体,简称总体。

例如,研究某个工业部门的企业生产情况时,该部门的所有工业企业可以作为一个总体,因为它是由许多客观存在的工业企业组成的,而每个工业企业都是进行工业生产活动的基层单位,具有同质性。

如果一个统计总体中包括的单位数是无限的,称为无限总体,例如,连续大量生产某种零件时,其总产量是无限的,构成一个无限总体。总体中包括的单位数是有限的,称为有限总体。例如,在特定时点上的人口总数、工业企业总数等,都是有限总体。对于有限总体,既可以进行全面调查,也可以抽样调查。对于无限总体来说,只能进行抽样调查,根据样本数据推断总体特征。此外,统计总体还可以分为静态总体和动态总体,前者所包含的各个单位属于同一个时间,后者所包含的各个单位则属于不同时间。根据一定的目的,针对这两类总体就可以分别进行静态研究或动态分析。

综上所述,可见总体和总体范围的确定、取决于统计研究的目的要求。而形成统计总体的必要条件,亦即总体必须具备三个特性:大量性、同质性和差异性。

(1)大量性。大量性是总体的量的规定性,即指总体的形成要有一个相对规模的量,仅仅由个别单位或极少量的单位不足以构成总体。因为个别单位的数量表现可能是各种各样的,只对少数单位进行观察,其结果难以反映现象总体的一般特征。统计研究的大量观察法表明,只有观察足够多的量,在对大量现象的综合汇总过程中,才能消除偶然因素,使大量社会经济现象的总体呈现出相对稳定的规律和特征,这就要求统计总体必须包含足够多数的单位。足够多数,是指足以反映规律的数量要求。当然,大量性也是一个相对的概念,它与统计研究目的、客观现象的现存规模以及总体各单位之间的差异程度等都有关系。

(2)同质性。总体的同质性,是指构成总体的各个单位至少有一种性质是共同的,同质性是将总体各单位结合起来构成总体的基础,也是总体的质的规定性。例如,全国工业企业作为统计总体,则每个总体单位都必须具有从事工业生产活动的企业特征,而不具有这些特征的就不能称之为工业企业。如果违反同质性,把不同性质的单位结合在一起,对这样的总体进行统计研究,不仅没有实际意义,甚至会产生虚假和歪曲的分析结论。

同质性的概念是相对的,它是根据一定的研究目的而确定的,目的不同,同质性的意义也就不同。例如,研究全国工业企业的生产状况时,所有工业企业都是同质的,而研究民营工业企业生产状况时,那么,民营工业企业与国有工业企业就是异质的。可见,同质性是相对研究目的而言的,当研究目的确定后,同质性的界限也就确定了。

(3)差异性。总体各个单位除了具有某种或某些共同的性质以外,在其他方面则各不相同,具有质的差别和量的差别,这种差别称为变异。正因为变异是普遍存在的,才有必要进行统计研究,是统计的前提条件。总体中各个单位之间具有变异性的特点,这是由于各种因素错综复杂作用的结果,所以有必要采用统计方法加以研究,才能表明总体的数量特征。

定义 1.10 总体单位:构成总体的每一个事物或基本单位。

原始资料最初就是从各个总体单位取得的,所以总体单位是各项统计数字最原始的承担者。例如,研究某个工业部门的生产情况时,该工业部门的所有工业企业可以作为一个总体,每个工业企业则是总体单位,将每个工业企业的某些数量特征加以登记汇总,就取得该工业部门的统计资料。

总体和总体单位是相对而言的,在一次特定范围、目的的统计研究中,统计总体与总体单位是不容混淆的,二者的含义是确切的,是包含与被包含的关系。但是随着统计研究目的及范围



的变化,统计总体和总体单位可以相互转化。同一事物在不同情况下,可以作为总体,也可以作为总体单位。例如,在上述某一工业部门所有工业企业的统计总体中,每个企业是一个总体单位。但为了要研究一个典型企业的内部问题时,则被选作典型的某一企业又可作为一个总体。

2. 样本

定义 1.11 样本:从总体中抽取的一部分元素所组成的集合称为样本。构成样本的元素数目,称为样本容量。

例如,一家公司正在接受审计,审计人员没有必要对该公司年度内的所有 55400 张发票全部审查,只需随机抽查一个 100 张发票的样本即可,审计人员通过这 100 张发票计算的差错率可对全部 55400 张发票的差错率进行推断,其中 100 即为样本容量。

样本是从总体抽取出的、作为总体的代表、由部分单位组成的集合体。在抽样推断中,总体又被称为母体,相应地,样本也被称为子样。抽取样本应注意如下几个问题:

(1)样本的单位必须取自总体,这是因为抽取样本的目的是为了推断总体,所以,不允许以总体外部的单位作为该总体的样本;

(2)一个总体可以抽取许多样本,样本个数的多少与抽样方法有关;

(3)样本的抽取必须排除主观因素的影响,以确保样本的客观性与代表性。

1.3.2 标志和指标

1. 标志

定义 1.12 标志:用来说明总体单位属性和特征的名称。

标志按其表现形式有数量标志与品质标志两种。凡是表示总体单位数量特征的标志,称数量标志。它能用数量来表示,如企业的职工人数、产量、产值;职工的年龄、工龄、工资等。凡是表示总体单位“质”的特征的标志,称品质标志。如职工的性别、企业的经济类型、工人的工种等。标志的具体表现是在标志名称之后所表明的属性或数值,如某职工的性别是女,民族是汉族。这里的“性别”和“民族”是品质标志名称。而“女”和“汉族”是这类标志的属性的具体表现。又如该工人的年龄是 35 岁,工资是 1960 元,则“年龄”和“工资”是数量标志的名称,而“35 岁”和“1960 元”则是它们的数值表现。

按照标志在总体中各单位的具体表现是否相同,标志可分为不变标志和可变标志。标志在同一个总体中的各单位具体表现都相同,我们称之为不变标志。例如,在教师总体中,职业这一标志在各单位的表现都是相同的,都是教师,在此,职业就是不变标志。一个总体中,至少要有不变标志,才能把各单位结合成为一个总体。如果没有不变标志,那么,总体也就不存在。由此可见,不变标志是总体同质性的基础。标志在同一个总体中的各单位具体表现有可能不同,我们称之为可变标志。可变标志的属性或特征的具体表现是由一种状态变为另一种状态,统计上称之为变异,因而可变标志也称为变异标志。例如,在教师总体中,教龄这一标志在各单位的表现可能不同,在此,教龄就是可变标志。在一个总体中,如果不存在可变标志,或者说所研究的现象总体在各单位之间不存在着任何差异,这就无须做调查,无须进行统计研究了。所以,总体的同质性是问题研究的基础,而总体的差异性则是问题研究的本质。

2. 指标

定义 1.13 指标:反映统计总体的数量特征的概念和数值称为指标。