

TONGJIXUE

YUANLI YU YINGYONG

统计学

原理与应用

张瑜 牟晓云 等·编著



东南大学出版社
SOUTHEAST UNIVERSITY PRESS

014037614

C8
257

图书在版编目(CIP)数据

统计学原理与应用

TONG JI XUE
YUAN LI YU YING YONG

张瑜 牟晓云 等·编著



 东南大学出版社
SOUTHEAST UNIVERSITY PRESS

·南京·



北航

C1726091

C8
257

014032814

图书在版编目(CIP)数据

统计学原理与应用 / 张瑜等编著. —南京: 东南大学出版社, 2014. 4

ISBN 978-7-5641-4815-7

I. ①统… II. ①张… 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(2014)第 059704 号

用前已服煎者书能

TONG JI XUE
YUAN LI YU YING YONG

審識·第 三 册 余 第 一 卷

统计学原理与应用

出版发行 东南大学出版社
出版人 江建中
社 址 南京市四牌楼 2 号(邮编:210096)
网 址 <http://www.seupress.com>
责任编辑 孙松茜(E-mail:ssq19972002@aliyun.com)
经 销 全国各地新华书店
印 刷 大丰市科星印刷有限责任公司
开 本 787mm×1092mm 1/16
印 张 21
字 数 560 千字
版 次 2014 年 4 月第 1 版
印 次 2014 年 4 月第 1 次印刷
书 号 ISBN 978-7-5641-4815-7
定 价 45.00 元

(本社图书若有印装质量问题,请直接与营销部联系。电话:025-83791830)

目 录

第一章 总论	1
第一节 统计的产生和发展	1
第二节 统计学的研究对象、性质和方法	4
第三节 统计基本要素	8
第四节 统计职能	14
第五节 统计活动的基本环节	14
第六节 统计学在经济学和经济管理中的应用	16
本章小结	17
思考与习题	18
第二章 统计调查	21
第一节 统计调查的基本理论	21
第二节 统计调查方案	25
第三节 问卷设计	27
第四节 统计调查的组织形式	31
本章小结	38
附录 设计调查问卷	39
思考与习题	42
第三章 统计数据的整理	44
第一节 统计整理的基本理论	44
第二节 统计分组	45
第三节 次数分布	51
第四节 统计表	56
第五节 统计图	58
本章小结	60
附录 利用 Excel 进行数列分布实验	61
思考与习题	64
第四章 总量指标和相对指标的计算	67
第一节 总量指标	67
第二节 相对指标	70
第三节 相对指标的运用	77
本章小结	79
思考与习题	80
第五章 统计数据的描述	86
第一节 集中趋势	86
第二节 离散程度	98
第三节 偏态与峰度	101
本章小结	104
附录 利用 Excel 和 SPSS 进行描述性统计	104
思考与习题	107
第六章 参数估计	110
第一节 参数估计的基本方法	110
第二节 总体均值的区间估计	113

第三节	总体比例的区间估计	119
第四节	样本容量的确定	121
本章小结		123
附录	利用 Excel 和 SPSS 进行参数估计	124
思考与习题		129
第七章	假设检验	132
第一节	假设检验的一般问题	132
第二节	总体均值的假设检验	138
第三节	总体比例的假设检验	144
第四节	总体方差的假设检验	145
第五节	配对样本的假设检验	148
本章小结		150
附录	利用 Excel 和 SPSS 进行假设检验	151
思考与习题		156
第八章	方差分析	160
第一节	方差分析概述	160
第二节	单因素方差分析	163
第三节	双因素方差分析	168
本章小结		176
附录	利用 Excel 和 SPSS 进行方差分析	176
思考与习题		185
第九章	相关与回归分析	188
第一节	相关分析	188
第二节	一元线性回归分析	193
第三节	多元线性回归分析	204
本章小结		212
附录	利用 Excel 和 SPSS 进行相关与回归分析	212
思考与习题		224
第十章	统计指数	227
第一节	统计指数概述	227
第二节	总指数的计算	230
第三节	指标体系与因素分析	240
第四节	指数的应用	246
本章小结		250
附录	利用 Excel 计算统计指数	251
思考与习题		254
第十一章	时间序列分析	257
第一节	时间序列概述	257
第二节	时间序列水平指标分析	259
第三节	时间序列速度指标分析	265
第四节	时间序列因素构成分析	268
第五节	时间序列分析与预测	281
本章小结		284
附录		285
思考与习题		304
参考答案		308
附表		326
参考文献		331
后记		332

第一章 总论

第一节 统计的产生和发展

一、统计的概念

在日常生活中,人们对于“统计”这一术语常常有不同的理解。例如,企业每年要“统计”产品的产量和产值;了解股票的交易状况要看有关成交额和股票指数“统计”;解说球赛时解说员要不断统计竞赛双方的进攻次数和成功率;报刊上定期或不定期地公布诸如物价指数、人口增长率、国内生产总值等统计数据。“统计”一词,在不同的场合,人们赋予它不同的含义。

那么究竟何为“统计”,这里有必要给出一个准确的科学定义。在中国,“统计”一词是由英语“statistics”一词翻译过来的,具体有统计工作、统计资料和统计学三种含义。

(一) 统计工作

统计工作是搜集、整理、分析和研究统计数据资料的工作过程,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、户登记、数据汇总、分析总结和资料公布等一系列过程,这就是统计工作。

(二) 统计资料

统计资料是统计工作所取得的各项数据资料及有关情况的总称,如统计表、统计图、统计分析报告和各种统计资料汇编等。

(三) 统计学

统计学是对统计工作及统计规律进行的科学总结和理论概括。

统计的三种含义具有密切联系:统计工作是统计的基础;统计资料是统计的成果;统计学是统计工作的经验总结与理论概括。反过来,统计学又是指导统计工作的基本原理、原则与方法,并使统计资料更加准确、更加及时和更加全面;统计工作是先于统计科学发展起来的。

二、统计学的产生和发展

统计活动已经有了几千年的历史,但在学术上作为一门学科的统计学的历史却没有这么长。一般认为,统计学产生于17世纪中叶的欧洲,距今已有300多年。其发展主要可分为三个阶段:

（一）古典统计学时代

这个时代大致是从 17 世纪中叶至 19 世纪初叶,其代表学派是“政治算术派”和“国势学派”。

“政治算术派”产生于英国,后人称其为统计学中的政治算术学派。其主要代表人物是英国的威廉·配第(W. Petty)和约翰·格朗特(J. Graunt)。

配第于 1676 年出版了《政治算术》一书,以一系列分析和大量计算手段清晰地描述了英格兰、荷兰、法兰西和爱尔兰等地的经济、军事、政治等方面的情况,为英国称霸世界提供了各种有说服力的实证分析资料。

《政治算术》的意义主要表现在研究问题的方法方面。配第在该书的“序言”里写道:“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辨式的议论,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的问题,只进行诉诸人们的的感觉的议论,借以考察在自然中有可见的根据的原因。”

配第用“数字、重量和尺度”研究现象的方法为统计学的产生奠定了基础。自配第之后的 200 年间,以用数量方法研究社会经济问题为基本特征的“政治算术”模式成为统计学发展的主流。《政治算术》的出版,标志着统计学的诞生。马克思对威廉·配第和他的《政治算术》评价很高,他说“配第创造‘政治算术’,即一般所说的统计”,^①还称配第是“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”。^②

英国人约翰·格朗特于 1662 年出版了《关于死亡表的自然观察与政治观察》。他根据伦敦市发表的人口自然变动公报,通过大量观察的方法,对人口的出生和死亡率作了许多分类、计算和研究,发现了人口与社会现象中重要的数量规律性。如新生儿的性别比例稳定在 14:13;男性在各年龄组中死亡率高于女性;新生儿的死亡率较高,一般疾病与事故的死亡率较稳定,而传染病的死亡率波动较大;等等。在研究中,格朗特不但探索了人口变化和发展的一些数量规律,而且还对伦敦市总人口数量做出了较科学的估计。如果说配第是政府统计的创始人,则格朗特可被认为是人口统计的创始人。

国势学派又称记述学派或国情学派,产生于 17 世纪封建制的德国,其主要代表人物是海尔曼·康令(H. Gonring)和高特弗里德·阿亨瓦尔(G. Achenwall)。这一学说最早提出了“统计学”的名称。在康令之前,欧洲各国已出版有记述各国情况的著作。从 1660 年开始,康令在西尔姆斯特大学以“国势学”为题讲述一门课程,内容是各个国家的显著事项,方法则是文字叙述,目的是为了满 足政治家所必需的知识。阿亨瓦尔是国势学的主要继承人和最有名的代表人物,一生在大学任教。他在 1749 年出版的《近代欧洲各国国势学论》中,首次使用“统计学”(Statistik)这个名称代替了国势学(阿亨瓦尔说过,Statistik 的语源是拉丁语 status 和意大利语 state,前者是“现状”或“现势”的意思,后者是“国家的”意思)。阿亨瓦尔对统计学的性质作了解释。他认为统计学是关于各国基本制度的学问,其研究对象是一个国家显著事项的整体。这里的“国家显著事项”是指一个国家的领土、人口、财政、军事、

① 《马克思恩格斯选集》第 3 卷,北京:人民出版社,1972:273。

② 《马克思恩格斯选集》第 3 卷,北京:人民出版社,1972:302。

政治和法律制度等等,用这些来说明和比较国家的形势,因此称为国势学。

(二) 近代统计学时代

这个时代大致是从 18 世纪末到 19 世纪末。著名的大数法则、最小平方法、相关与回归分析、指数分析法、时间数列分析法以及正态分布等理论都是在这个时期建立和发展起来的。其代表学派主要有数理统计学派和社会经济统计学派。

数理统计学派产生于 19 世纪中叶,创始人是比利时学者阿道夫·凯特勒(A. Quetelet)。他在统计理论上的主要贡献是把概率论引入了统计学,从而提出了关于统计学的新概念。凯特勒根据大数定律的原理提出了大量观察法,利用统计观察资料计算和研究社会现象和自然现象的数量规律性,并用于预测未来的情况。他创立大数法则,认为统计学就是数理统计学。

凯特勒开创了统计理论和实际应用的一个新领域,即应用概率论认识随机现象数量规律性的理论和方法。这个新领域起初没有确定的名称,1867 年德国数学家威特斯坦(T. Wittstein)发表了题为《数理统计学及其在经济学和保险学中的应用》的论文,因而定名为数理统计学。数理统计学产生较晚,但发展很快。后来经过葛尔顿、皮尔逊、鲍莱、友尔、戈塞特、费雪等人的研究和实践,发展成为一门完整系统的新学科。

社会统计学派产生于 19 世纪末期,首创者是德国人克尼斯(Knies),主要代表人物有梅尔、恩格尔,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) 现代统计学时代

20 世纪初,大工业的发展对产品质量检验问题提出了新的要求,即只抽取少量产品作为样本对全部产品的质量好坏做推断。因为大量产品要做全面的检验,既费时、费钱,又费人力,加之有些产品质量的检验要做破坏性检验,全部检验已不可能。1907 年,“学生”(W. S. Gosset 戈塞特的笔名)发表 t 分布的论文,创立了小样本代替大样本理论,利用 t 统计量就可以从大量的产品中只抽取较小的样本完成对全部产品质量的检验和推断。费雪(R. A. Fisher)又对小样本理论进一步研究,给出了 F 统计量、最大似然估计、方差分析等方法 and 思想,标志着现代统计学的开端。1930 年,尼曼(J. Neyman)与小皮尔逊(E. S. Pearson)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,同时尼曼又创立了区间估计理论。美国统计学家瓦尔德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。

从 20 世纪 50 年代以来,统计理论、方法和应用进入了一个全面发展的新阶段。一方面,统计学受计算机科学、信息论、混沌理论、人工智能等现代科学技术的影响,新的研究领域层出不穷,如多元统计分析、现代时间序列分析、贝叶斯统计、非参数统计、线性统计模型、探索性数据分析、数据挖掘等。另一方面,统计方法的应用领域不断扩展,几乎所有的科学研究都离不开统计方法。因为不论是自然科学、工程技术、农学、医学、军事科学,还是社会科学都离不开数据。只要对数据进行研究和分析就必然要用到统计方法,现在文科领域的法律、历史、语言、新闻等都越来越重视对统计数据进行分析,国外的人文与社会学科普遍开设统计学的课程,因而可以说统计方法与数学、哲学一样成为所有学科的基础。

三、现代统计学的发展趋势

在最近半个世纪统计活动的发展中,国民经济账户体系的建立,概率论和其他数学方法的广泛应用,统计机构的进一步完善,信息处理手段的自动化,被称为“现代统计”的四大标志。随着信息社会的到来,依据建立社会主义市场经济和现代管理科学发展的需要,统计面向决策部门、面向社会、面向基层、面向世界的服务领域和内容逐渐扩展,统计信息涉及的范围不断扩宽,统计在国家管理、企业预测与决策、市场营销调查、家庭投资决策以及人类的一般认识活动和科学研究等各个领域将被广泛地应用。从学科发展趋势看,统计学必将由描述统计向管理统计的方向发展,必将与信息科学紧密结合。

第二节 统计学的研究对象、性质和方法

一、统计学的研究对象

统计学的研究对象是指统计所要认识的客体,是大量社会现象(主要是经济现象)总体的数量表现,即社会经济现象总体的数量特征、数量关系及其规律性表现。具体地说就是采用科学的方法,搜集、整理、分析实际数据,并通过统计指标和指标体系来表明现象的规模、水平、速度、比例和效益等。

统计学和数学都是研究数量关系的,但两个学科有着不同的性质特点。数学撇开具体的对象,以最一般的形式研究数量的关系和空间形式,数学的分析方法主要是逻辑和演绎论证的方法;而统计分析的方法,本质上是归纳的方法,根据试验和调查,观察到大量的个别情况,加以归纳以判断总体的情况。

二、统计学研究对象的特点

统计学研究对象的特点主要有:数量性、总体性、具体性、社会性、广泛性、差异性。

(一) 数量性

统计学研究的是大量社会经济现象总体的数量方面的特征,包括:①社会经济现象的数量多少;②各种现象之间的数量关系;③事物与量互变的界限和规律性。数量性是统计学区别于其他经济科学(如政治经济学)最根本的特点,可以说“数字是统计的语言”。但必须注意,它所研究的事物的量是从社会现象的定性认识开始,以质的规定性为基础的。

例如,我们观察一定时期人们的物质消费水平,就要统计一定范围一定时期的“社会商品零售额”,并与相应的人口数进行比较。统计“社会商品零售额”,首先要明确这个指标的含义,然后再去统计它的“量”。而确定“社会商品零售额”的含义,需要以政治经济学的理论为指导,与实践相联系,并解决什么是商品流转、什么是零售、社会商品零售额的统计范围以及如何搜集、整理和汇总社会商品零售额统计资料等问题。在确定这一统计指标含义的同时,还必须考虑指标的可操作性。当我们统计了“社会商品零售额”以后,把它与相应的人口联系起来对比分析,就可以观察到相应的物质消费水平。我们还可以把“社会商品零售额”

与“社会商品购买力”进行对比,观察购买力的实现程度,分析两者之间的比例关系。至于这两者之间应该维持一个什么样的比例关系,才能稳定市场、满足人民的物质文化需要,还应进一步分析它们之间的数量界限。由此可见,统计虽然是研究社会经济现象数量方面的问题,但任何时候都不能离开社会经济现象的质。

(二) 总体性

统计学研究社会现象数量方面的问题不是指个别现象的数量特征,而是指由许多个别现象所构成的总体的数量特征,是通过对许多性质相同的个别现象所组成的总体进行大量观察和综合分析,来反映现象总体的数量特征,揭示社会经济现象的一般状况。

统计学研究对象的总体性特点,是由社会经济现象的特点和统计学研究的目的来决定的。由于社会经济现象错综复杂,个别现象所处的时间、地点和条件不同,表现出明显的偶然性和不确定性,难以说明社会经济现象总体的本质和规律。只有以社会经济现象的总体为研究对象,即以构成总体的全部或足够多的单位作为研究对象时,才能消除偶然性因素的影响,防止“只见树木,不见森林”的片面性,从而正确地揭示出社会经济现象的本质和规律性。当然,任何一个总体都是由个体所构成的,要认识社会经济现象总体的数量特征,必须从调查个体的表现入手,从个体到总体。

例如,研究中国农民的生活水平,就需要把全国各省、市、自治区的所有农民组成一个总体来统计,不论是哪一个地区、哪一个民族,也不论是高收入还是低收入,只要是农民都要包括在内。这样就可以消除地理环境、民族特征、收入高低等方面的差异,反映出中国农民生活水平的一般情况。当然,为了深入分析农民的生活水平,还可以就高收入、中收入、低收入农民的典型进行调查分析,探究不同典型的差异和形成原因,从而全面客观地说明被研究对象的情况。

(三) 具体性

统计学研究的对象是社会经济现象中具体事物数量方面的问题,而不是抽象的数量及其相互关系,这是统计学和数学的重要区别。这是由于社会经济现象中的事物都是具体的,都是在一定的地点、时间、条件下发生的,所以其量的表现就必然带有特定场合和特定历史的痕迹,离开具体地点、时间和条件,是无法说明社会经济现象的本质及其运行规律的。

例如,中国农民的生活水平,既存在地区上的差异,也存在时间上的不同。因此,在研究时,除考虑地区因素以外,还必须明确是哪一年的生活水平,并联系我国农业的生产条件、科学技术在农业生产中的作用,联系农村消费品价格等因素进行分析,来说明中国农民的生活水平是高还是低,是好还是差。单凭一个孤立的统计数字是难以说明问题的。

(四) 社会性

统计学属于社会科学,这一点与自然科学不同。社会科学研究的是社会经济问题,而自然科学研究的是自然现象。除了研究对象不同以外,社会科学由于认识主体所站立场、所持观点、所用方法的不同,会得出差别很大、甚至完全不同的结论。

统计学通过社会经济现象总体数量的调查研究,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。统计学研究的社会经济现象与各种利益关系是密切联系的。其定量研究是以定性分析为前提的,而定性分

析使统计学在客观上就有了社会关系的内涵。所以,统计学在研究社会经济现象时,就必须注意正确处理好这些涉及人与人之间关系的社会矛盾。

例如,在研究劳动者的收入时,可以根据劳动者在社会再生产过程中的地位和作用,将劳动者分为经营者、管理者、技术员、普通工人等,然后再统计不同类型劳动者的收入,分析他们之间的相互关系,研究社会分配的合理性,从而制定或修改劳动报酬分配政策,以调整不同类型劳动者之间的相互关系,达到稳定社会秩序、调动广大劳动者积极性的目的。如果不加区别地把经营者、管理者、技术员和一般工人的收入混为一谈,势必得出错误的结论,以致制定出错误的政策,从而引起不必要的各种矛盾和社会问题。所以说,社会性是统计学区别于其他自然科学的主要特征之一。

(五) 广泛性

统计学研究的数量方面的问题非常广泛,包含全部社会现象的数量方面的问题。这个特点,是统计学区别于研究某一领域的其他社会科学(如政治学、经济学、社会学、法学等)的特征之一。

统计学研究的领域涵盖整个社会,它既研究生产关系,也研究生产力以及生产关系和生产力之间的关系;它既研究经济基础,也研究上层建筑以及经济基础和上层建筑之间的关系。此外,还研究生产、流通、分配、消费等社会再生产的全过程以及社会、政治、经济、军事、法律、文化、教育等全部社会现象数量方面的问题。

(六) 差异性

统计研究对象的差异性,是指总体各单位的特征表现存在着差异。统计学研究同类对象总体的数量特征,它的前提就在于这个特征在总体各单位的具体表现各不相同,而且这种差异并不是由固定的原因事先给定的。例如,研究一个地区居民家庭的收入水平,就是因为各家庭的收入有高有低,参差不齐,这样才有必要研究该地区的人均收入水平及其分布状况。

三、统计学的性质

统计学究竟是属于方法论科学,还是属于实质性科学?这个问题在理论界至今没有一个统一的明确说法,回答这个问题对于全面认识这门学科和为以后的论述铺平道路具有重要意义。国内有三种观点:

第一种,规律派。这一派认为统计学是研究社会经济现象发展规律的,即统计学是通过研究在一定时间地点条件下的社会经济现象的数量表现,来揭示社会经济发展规律的独立的社会科学,是一门实质性科学。

第二种,数理统计学派。这一派认为统计学的研究对象是随机现象,是以概率论为基础的应用数学,是一门通用的数理方法学科。这一派认为并不存在独立的社会经济统计学,它只不过是数理统计方法在研究社会经济现象时的应用。

第三种,方法论派。这一派认为统计工作和统计科学是不同的。统计工作研究的是大量的社会经济现象数量方面的问题以及社会经济现象与自然技术因素相互影响的数量变化。而统计学研究的则是社会经济统计活动的规律和方法,即社会经济统计工作的方法论。

本书的观点与第三种观点是一致的。我们认为,统计学属于认识社会经济总体现象数量方面的问题的方法论科学,或者说是对一定社会经济总体现象的定量认识方法论。从研究领域来讲它属于社会科学,从研究对总体现象数量方面的认识方法来讲它属于认识方法论科学。因此它有两方面的基本特性:一方面是包括阶级性在内的社会性;另一方面是与社会性结合的数量方法性。

四、统计学的研究方法

统计学研究的基础工作是收集数据,目的是探索大量数据的数量特征与规律性,而贯穿于这一过程的统计研究方法主要有大量观察法、描述统计法和推断统计法。

(一) 大量观察法

大量观察法的数学依据是大数定律。大数定律是随机现象的基本规律。大数定律的一般概念是:在观察过程中,每次取得的结果不同,这是由偶然性所导致的,但大量、重复观察结果的平均值却几乎接近确定的数值。大量观察的本质意义,在于通过大量观察,把个别的、偶然的差异性相互抵消,而必然的、集体的规律性便显示出来。例如,随机地投掷一枚硬币或骰子出现正面、反面或某个点数是不确定的,完全是偶然的。但我们进行多次的重复投掷,就会发现投一枚均匀硬币出现正面和反面的次数大体相同,即比值接近于 $1/2$ 。投掷的次数越多,就越接近于 $1/2$ 这一稳定的数值。同样,在投掷骰子时,出现1至6点任一点数的比例随着投掷次数的增加也逐渐接近于 $1/6$ 。这里的 $1/2$ 和 $1/6$ 就是掷硬币和掷骰子出现某一特定结果的概率,也就是投掷硬币或骰子时所呈现的数量规律性。这个例子说明,通过多次观察或试验得到大量的统计数据,利用统计方法是可以探索出其内在的数量规律性的。

(二) 描述统计法

描述统计法,就是运用各种表格、图形和数字来概括总体数量特征的方法。在对观察到的大量原始数据进行整理汇总后,可以绘制统计表和统计图来显示总体数量的分布特征,也可以计算得出平均数、变异指标等一系列统计测度值,这些测度值均可以揭示现象在一定条件下的集中趋势和离散趋势等。

近年来,随着计算机技术的普及和发展,描述统计法的各种方法都可以借助计算机来实现。

(三) 推断统计法

推断统计法,就是在人们可以控制的范围内,根据样本资料的特征,对总体的特征做出估计和观测的方法。当面对的总体范围很大甚至是无限总体时,或者由于经费、时间或破坏性检验等因素,我们只能在客观上根据局部观察的结果来推断总体的特征。例如,要观察一批灯泡的平均使用寿命,只能从该批灯泡中抽取一小部分进行检验,推断这一批灯泡的平均使用寿命。在研究对象的总体数量关系中,推断统计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,也可以用于对总体某些假设的检验,因而它被广泛地应用于统计研究的许多领域。

上述统计研究方法在统计研究过程中的关系可以通过图1-1得到反映。

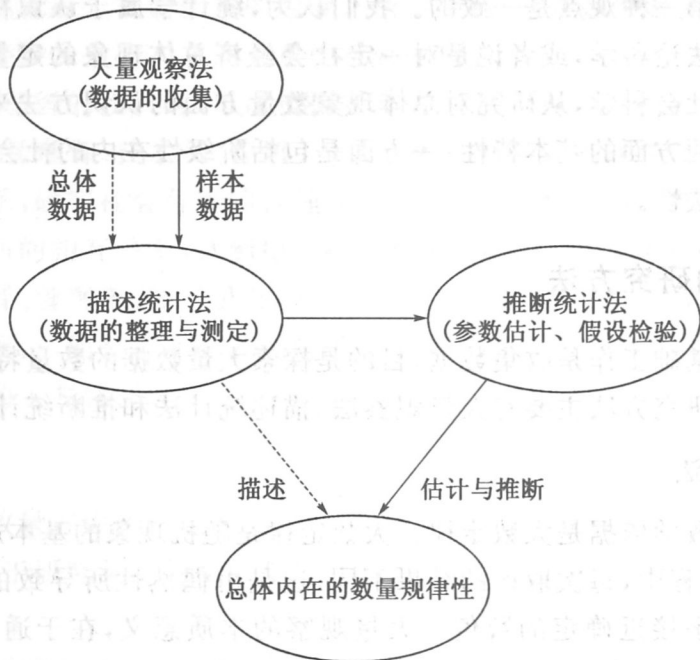


图 1-1 统计方法在统计研究过程中的关系

由图 1-1 可以看出:统计研究过程的起点是个体数据,终点是客观现象总体的数量规律性。如果我们收集到的是总体数据(如普查),则经过描述统计之后就可以达到统计研究的目的了;如果我们获得的是样本数据,经过大量观察以及对样本数据特征进行描述后,还要用概率论的理论并结合样本统计量的信息对总体做出科学的推断,显然推断统计法是在描述统计法的基础上得以进行的。

尽管描述统计法可以在获得总体数据时直接探索出总体的数量规律性来,但这种情况在实际工作中很少见到。大量的管理和研究工作不可能组织普查,例如城市居民每月的家庭收入支出调查、某种商品的市场调查、某个事件的民意测验等都只能是先进行抽样调查,然后再对总体的数量规律性进行科学的推断。另外在实践中总体单位很多,加之要考虑时间、费用和人力的投入,实际上难以一一进行调查,所以推断统计法被广泛应用于管理活动的各个领域。

第三节 统计基本要素

一、统计总体和总体单位

(一) 统计总体

如前所述,统计学是从总体上研究大量社会经济现象数量特征的,这就产生了统计总体的概念。凡是客观存在的、在至少有某一性质相同的基础上结合起来的许多个别事物(单位)组成的整体,都称为统计总体(简称总体)。

总体是统计研究的具体对象。例如,要研究 A 市民营企业的生产经营情况,那么该市的所有的民营企业就构成了一个总体,统计设计、统计调查、统计整理和统计分析则都要围

绕这一对象来进行。再如,要研究我国的人口状况,则全国人口就构成一个总体,从设计普查方案、普查登记、资料汇总到最后公布普查数据等等,也都要围绕这一对象来进行。

总体可分为有限总体和无限总体。有限总体是指总体中的总体单位数可以计数或穷尽的总体。例如一个企业的全体职工、一个国家的全部人口等都是有限总体。如果总体中的单位数是一个无穷大量,或准确地度量它的单位数是不经济或没有必要的,这样的总体称为无限总体。例如在连续生产的生产线上产出的全部零件数,一片树林中生长的林木数,江河湖海中生长的鱼的尾数等。对于有限总体既可以进行全面调查,也可以进行非全面调查。但对于无限总体则只能抽取一部分单位进行非全面调查,据以推断总体。

总体具有三个特点:

(1) 同质性。所谓同质性是指构成统计总体的各个单位必须在某些方面而且至少在某一个方面具备共同的性质。同质性是构成总体的前提。

(2) 大量性。大量性是指总体是由许多总体单位组成的,只有一个单位的总体是不存在的。当然,研究目的不同,统计总体就不一样,总体中所包含的总体单位的数量也就不同,一个总体究竟包含多少总体单位,最终取决于统计研究的目的。

(3) 变异性。简言之,变异就是事物之间的差异或不同。从统计研究的角度来说,变异性是指构成总体的各个单位之间存在的差别。例如,学生的性别具体表现为男、女,学生的成绩表现为78分、85分、92分等。

在此,有三个问题需要特别说明:首先,变异是客观的,没有变异的事物是不存在的;其次,变异对于统计非常重要,没有变异就没有统计,这是因为如果总体单位之间不存在变异,我们只需要了解一个总体单位的资料就可以推断总体情况了;第三,变异性和同质性之间相互联系、相互补充,是辩证统一的关系。用同质性否定变异性或用变异性否定同质性都是错误的。

(二) 总体单位

构成统计总体的个别单位称为总体单位,或称个体。例如,全部A市民营企业中的每一家民营企业、该市所有人口中的每个人都是总体单位。

(三) 总体与总体单位的相互关系

总体和总体单位的关系是整体与个体、集合与元素的关系,如果说总体是集合的概念,那么总体单位就是集合的元素。两者相互依存、相互联系,不存在没有总体的总体单位,也不存在没有总体单位的总体。

总体和总体单位的具体形式随着统计研究目的的不同而不同,可以是人,也可以是物,还可以是组织(企业或家庭)或时间、空间、行为等。

总体和总体单位的关系不是一成不变的,随着研究目的的变动,两者可以相互转化。在一定研究目的下,一个事物可以作为总体而存在,然而当研究目的发生变化后,这个事物可能就成为总体单位了。例如,当研究A市民营企业的生产经营情况时,全部该市的民营企业就是一个总体,其中的每一家民营企业就是一个总体单位;如果要研究一家民营企业的生产经营情况,那么这家民营企业就成为总体了;而如果要研究全国的民营企业的生产经营情况,全国的民营企业组成总体,而其中A市的民营企业又变成了总体单位了。

二、样本

统计研究的目的是要确定总体的数量特征,但是当总体单位数量很多甚至无限时,不必要也不可能对构成总体的所有单位都进行调查。这时,需要采用一定的方式,从总体(又称母体)中抽取一部分单位,作为总体的代表加以研究。这种由总体的部分单位组成的集合称为样本(又称子样)。样本也是由一定数量的单位构成的,样本所包含的总体单位数称为样本容量。

三、标志和指标

(一) 标志

统计是从对个体的观察开始,逐步过渡到对总体数量特征的认识的。标志是指说明总体单位特征或属性的名称。例如,A市一家民营企业作为总体单位,其“企业性质”、“企业类型”、“生产能力”、“年产值”、“销售收入”、“职工人数”、“工资总额”等都是标志。

标志按是否可以用数值表示来看,可分为品质标志和数量标志两种类型。品质标志是说明总体单位属性特征的名称,如“企业性质”、“企业类型”等,品质标志只能用文字而不能用数值表示。数量标志是说明总体单位数量特征的名称,如:“生产能力”、“年产值”、“销售收入”、“职工人数”等,数量标志既可以用文字也可以用数值表示。

由于数量标志的具体表现为一个数值,因此又称为标志值。在同一总体内,对于名称相同的标志,无论总体单位数有多少个,只能算是一个标志,而标志的具体表现则与总体单位数相同。如所有的A市民营企业为总体时,共有1.2万家企业,对于每家企业(总体单位)而言,产值是数量标志,且对该总体内的所有总体单位而言,也只能算是一个数量标志,但其具体表现(标志值)则有1.2万个。

如果按总体单位在标志上的具体表现是否存在差异来看,标志可分为不变标志和可变标志。不变标志体现为总体的同质性。组成一个总体的各个总体单位必须有一个或几个不变标志,不变标志是使许多个别单位组合成一个总体的前提。例如,以A市所有的中小民营企业为一总体,这里的各民营企业均有所有制、企业规模这两个不变标志。一家A市的中小民营企业如果不具备这两者中任何一个方面特征的话,就不能成为这个总体的一个单位。

可变标志是指具体表现在总体各个单位上不相同或不完全相同的那些标志。一般来说,组成总体的各个总体单位具有许多可变标志。例如,把A市所有的中小型民营企业作为一个统计总体,那么厂址、隶属关系、职工人数、资金额、生产能力、工业增加值、工业总产值、劳动生产率、平均工资、利税额等就是这个总体各单位的可变标志。

和标志相联系的另外一个概念是标志表现。所谓标志表现,是指总体单位特征在某一标志上的具体表现。如一个人的性别是“男”、年龄“50岁”、民族“汉族”;某企业是“股份公司”、年产值“1亿元”。和标志一样,标志表现也分为品质标志表现和数量标志表现两种类型。

(二) 统计指标

统计指标(简称指标)是反映社会经济现象总体数量特征的概念和具体数值。例如,要

表明 A 市全部民营企业这个总体的数量特征,其数量表现可以有:该市 2005 年底有民营企业 3.9 万家,全年总产值 870 亿元,职工人数 52.04 万人,人均产值 3.6 万元,总产值比上年增长 14%。

单就指标本身而言,其构成主要有两部分:指标名称,指标内容和所包括的范围,即指标质的规定性。指标数值,指数量的特征,是指标量的规定性。统计指标离不开数值。

然而,由于社会经济现象中的事物都是具体的,都是在一定的地点、时间、条件下发生的,其量的表现就必然带有特定场合和特定历史的痕迹。所以,一个完整的统计指标除了包括指标名称、指标数值外,还应包括计量单位、指标的时间范围、指标的空间范围及指标的计算方法等方面的要素。例如,“按可比价格计算,2005 年 A 市实现财政收入总额 182 亿元人民币”。在这个例子中,财政收入总额是指标名称,182 是指标数值,亿元人民币是指标的计量单位,2005 年是指标的时间范围,A 市是指标的空间范围,按可比价格计算是指标的计算方法。显然,上述六个要素在说明总体数量特征方面都是不可缺少的,否则就失去了作为一个统计指标的意义,也就不称其为指标了。

标志和指标,两者既有区别又有联系。区别有以下四点:

第一,标志是说明总体单位(个体)特征的;而指标是说明总体特征的。

第二,标志中的数量标志是可以数值表示,品质标志不能用数值表示;而所有的指标都是用数值表示的,不存在不能用数值表示的指标。

第三,标志中的数量标志不一定经过汇总,可以直接取得;而指标是由数量标志汇总得来的。

第四,标志一般不具备时间、地点等条件;而作为一个完整的统计指标,一定要有时间、地点、范围。

另外,指标和数量标志之间存在着一定的变换关系。由于研究目的的不同,当原来的总体变成总体单位时,相应的统计指标也就变成数量标志了(这时,指标名称变成数量标志名称,指标数值变成标志值);反之亦然。例如,在研究 A 市某一民营企业时,该民营企业为总体,其产值为指标,每一个车间为总体单位,车间的产值为数量标志;而当研究 A 市的民营企业时,所有的该市的民营企业则构成一个总体,该市的总产值为指标,而此企业则变成一个总体单位,其产值则变为数量标志了。

(三) 统计指标的分类

(1) 按其反映的事物性质不同,统计指标可分为实体指标和行为指标两类。实体指标是指它所反映的是具有实物形态、客观存在的具体事物数量特征,如产品产量指标、职工人数指标、固定资产价值指标等;行为指标是指它所反映的是某种行为的数量特征,如工伤事故指标、犯罪行为指标等。

(2) 按其数据的依据不同,统计指标可分为客观指标和主观指标两类。客观指标是指其取值依据是对统计对象的实际度量或计数的指标,又称为显性指标,如产品产量、职工人数等都是客观指标;主观指标是指不可能或难以直接度量或计数取值而只能凭人们的感受、评价确定其量的指标,又称为隐性指标,如民意测验、对事物综合评价等指标就属于主观指标。

(3) 按其反映社会经济的功

描述指标是反映社会经济现象的现实状况、变化过程和运行结果的统计指标,如反映生产经营条件的物质技术设备、职工人数、生产总值、总销售收入、利润总额等。评价指标是用于考核、评估、比较社会经济活动质量及其效果的统计指标,如设备利用率、资金周转率、职工劳动效率等。预警指标是对社会经济活动过程中的关键点进行监测,通过与正常值的比较而发出警示的统计指标,如宏观经济中的通货膨胀率、失业率、物价指数、社会积累率,微观经济中的资金利用率、成本利润率、工资利润率等。

(4) 按其反映总体内容的不同,统计指标可分为数量指标和质量指标。数量指标是反映总体范围、总体规模、总体水平的统计指标,也称为外延指标,其表现形式一般为绝对数。如A市民营企业数、投资总额、固定资产总值、总销售收入等。质量指标是反映总体内部结构、比例以及相互数量关系或发展变化的指标,也称为内涵指标,一般表现为相对数或平均数等,如该市民营企业生产效率、流通费用率、销售收入的增长率等。

(5) 按其数量对比关系的不同,统计指标可分为总量指标、相对指标和平均指标三类。这是统计学上常用的指标,也是最重要的分类。总量指标,又称绝对指标或绝对数,是反映总体的规模和现象发展结果的指标,其表现形式为绝对数,一般用以反映总体的总规模、总水平和工作总量。相对指标,又称相对数,是两个有联系的统计指标的对比形成的比率。其表现形式为相对数,一般用来反映总体的内部结构、现象间的数量对比关系和相对水平等。平均指标,又称平均数,是指总体中某一数量标志的一般水平。其表现形式为平均数,一般用来反映总体内某一数量标志的集中趋势等。

此外,统计指标还可以按其他标志进行分类。这些分类可以结合统计工作的实践进行,主要是为了便于在实际工作中有效运用。

四、变异和变量

(一) 变异

变异是指统计中的标志或指标间的差别,也就是标志和指标的具体表现各不相同。如人的性别有男女之别,各时期、各地区、各部门的产值不同,耕地面积不同等,这些差别被称为变异或变差。变异有属性的变异和数值的变异。变异是普遍存在的,是统计的前提条件。有变异才有统计,没有变异就用不着统计了。

(二) 变量

变量是可变的数量标志和统计指标。变量的数值表现就是变量值,也就是可变的数量标志和统计指标的不同取值。变量与变量值不能误用。例如,有工资2 250元、2 500元、2 750元、3 000元四个数值,要求计算其平均工资,不能说求这四个“变量”的平均数,因为这里只有“工资”这一个变量,并没有四个变量,所以平均的是“工资”这个变量的四个数值,即四个变量值。

(三) 变量的分类

1. 变量按其取值是否连续,可分为离散变量和连续变量

离散变量:变量值只能表现为整数。如工人数、工厂数、机器台数等。

连续变量:指其数值连续不断,在相邻的两值之间可无穷分割,表现为无穷小数。如粮