



全国统计教材编审委员会“十二五”规划教材

统计学

新编

孙玉环 主编

 中国统计出版社
China Statistics Press



全国统计教材编审委员会“十二五”规划教材

统计学

新编

孙玉环 主编

 中国统计出版社
China Statistics Press

图书在版编目(CIP)数据

统计学 / 孙玉环主编. — 北京 : 中国统计出版社,
2016.1

全国统计教材编审委员会“十二五”规划教材

ISBN 978—7—5037—7741—7

I. ①统… II. ①孙… 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(2016)第 004981 号

统 计 学

作 者/孙玉环

责任编辑/姜 洋

封面设计/上智博文

出版发行/中国统计出版社

通信地址/北京市丰台区西三环南路甲 6 号 邮政编码/100073

电 话/邮购(010)63376909 书店(010)68783171

网 址/http://www.zgtjcs.com

印 刷/河北天普润印刷厂

经 销/新华书店

开 本/787×1092mm 1/16

字 数/480千字

印 张/20.5

印 数/1~3000册

版 别/2016年1月第1版

版 次/2016年1月第1次印刷

定 价/43.00元

版权所有。未经许可,本书的任何部分不得以任何方式在世界任何地区
以任何文字翻印、仿制或转载。

中国统计版图书,如有印装错误,本社发行部负责调换。

国家统计局

全国统计教材编审委员会

顾 问 罗 兰 袁 卫 冯士雍 吴喜之

方积乾 王吉利 庞 皓 李子奈

主 任 徐一帆

副主任 严建辉 田鲁生 邱 东 施建军

耿 直 徐勇勇

委 员(按姓氏笔划排序)

丁立宏 万崇华 马 骏 毛有丰 王兆军

王佐仁 王振龙 王惠文 丘京南 史代敏

龙 玲 刘建平 刘俊昌 向书坚 孙秋碧

朱 胜 朱仲义 许 鹏 余华银 张小斐

张仲梁 张忠占 李 康 李兴绪 李宝瑜

李金昌 李朝鲜 杨 虎 杨沛华 杨映霜

汪荣明 肖红叶 苏为华 陈 峰 陈相成

房祥忠 林金官 罗良清 郑 明 柯惠新

柳 青 胡太忠 贺 佳 赵彦云 赵耐青

凌 亢 唐年胜 徐天和 徐国祥 郭建华

崔恒建 傅德印 景学安 曾五一 程维虎

蒋 萍 潘 璠 颜 虹

出版说明

全国统计教材编审委员会是国家统计局领导下的、全国统计教材建设工作的最高指导机构和咨询机构,自1988年成立以来,分别组织编写和出版了“七五”至“十一五”全国统计规划教材。

“十二五”时期,是我国全面实施素质教育,全面提高高等教育质量,深化教育体制改革,推动教育事业科学发展,提高教育现代化水平的时期。“十二五”伊始,统计学迎来了历史性的重大变革和飞跃。2011年2月,在国务院学位委员会第28次会议通过的新的《学位授予和人才培养学科目录(2011)》(以下简称“学科目录”)中,统计学从数学和经济学中独立出来,成为一级学科。这一变革和飞跃将对中国统计教育事业产生巨大而深远的影响,中国统计教育事业将在“十二五”时期发生积极变化。

正是在这一背景下,全国统计教材编审委员会制定了《“十二五”全国统计教材建设规划》(以下简称“规划”)。根据“学科目录”在统计学下设有数理统计学,社会经济统计学,生物卫生统计学,金融统计、风险管理与精算学,应用统计5个二级学科的构架,“规划”对“十二五”全国统计规划教材建设作了全面部署,具有以下特点:

第一,打破以往统计规划教材出版学科单一的格局。全面发展数理统计学,社会经济统计学,生物卫生统计学,金融统计、风险管理与精算学,应用统计5个二级学科规划教材的出版,使“十二五”全国统计规划教材涵盖5个二级学科,形成学科全面并平衡发展的出版局面。

第二,打破以往统计规划教材出版层次单一的格局。在编写出版好各学科本科生教材的基础上,对研究生教材出版进行深入研究,出版一批高水平高层次的研究生教材,为我国研究生教育、尤其是应用统计研究生教育提供教学服务。同时,积极重视统计专科教材出版,联合各专科院校,组织编写和出版适应统计专科教学和学习的优秀教材。

第三,打破以往统计规划教材出版品种单一的格局。鼓励内容创新,联系统计实践,具有教学内容和教学方法特色的、各高校自编的相同内容选题的精品教材出版,促进统计教学向创新性、创造性和多样性发展。

第四,重视非统计专业的统计教材出版。探讨对非统计专业学生的统计教学问题,为非统计专业学生组织编写和出版概念准确、叙述简练、深入浅出、表

达方式活泼、练习题贴近社会生活的统计教材,使统计思想和统计理念深入非统计专业学生,以达到统计教学的最大效果。

第五,重视配合教师教学使用的电子课件和辅助学生学习使用的电子产品的配套出版,促进高校统计教学电子化建设,以期最后能形成系统,提高统计教育现代化水平。

第六,重视对已经出版的统计规划教材的培育和提高,本着去粗存精、去旧加新、与时俱进的原则,继续优化已经出版的统计教材的内容和写作,强化配套课件和习题解答,使它们成为精品,最后锤炼成为经典。

“十二五”期间,编审委员会将本着“重质量,求创新,出精品,育经典”的宗旨,组织我国统计教育界专家学者,编写和编辑出版好本轮教材。本轮教材出版后,将能够形成学科齐全、层次分明、品种多样、配套系统的高质量立体式结构,使我国统计规划教材建设再上新台阶,这将对推动我国统计教育和统计教材改革,推动我国统计教育事业科学发展,提高我国统计教育现代化水平产生积极意义。

让教师的教学和学生的学习事半功倍,并使学生在毕业之后能够学以致用用的统计教材,是本轮教材的追求。编审委员会将努力使本轮教材好教、好学、好用,尽力使它们在内容上和形式上都向国外先进统计教材看齐。限于水平和经验,在教材的编写和编辑出版过程中仍会有不足,恳请广大师生和社会读者提出批评和建议,我们将虚心接受,并诚挚感谢!

国家统计局
全国统计教材编审委员会
2012年7月

前 言

本书是编者在原版《统计学》(东北财经大学出版社,2011年)基础上,根据最新统计学教学大纲的要求,在总结多年教学经验并参阅国内外相关资料的基础上修订编写完成的。参编人员均长期致力于统计学教学与研究工作,具有丰富的案例教学、实验教学和实践教学经验。本书主要面向具有一定概率论与数理统计基础的经济管理类本科生以及实际从事数据分析工作的人员。

本书在编写过程中力求体现以下特点:一是注重基础理论知识的讲解。在系统介绍统计学基本理论与基本方法的同时,强调正确应用统计方法的条件。二是强化应用案例。通过结合经济管理实际应用实例,介绍统计方法的具体应用,提高学生学以致用和解决实际问题的能力。三是实现统计方法讲解与软件操作的有机结合。在讲解统计学基本方法时,同时给出利用统计软件(以SPSS20.0为主,并辅之以EXCEL2010基础操作)实现数据处理和计算过程的操作,弱化繁杂的数据计算过程,强调对输出结果的准确解释,力求避免以往统计学基本原理教学与软件教学相脱节的情况,培养和提高学生基于实际数据解决现实问题的能力。

本教材设计学时为54学时。教材中各个章节内容的编排遵循以下主线:引例→方法原理→软件操作→习题,同时给出配套的例题、习题所用数据集、习题参考答案以及电子版PPT课件,以方便教师备课和帮助学生自学和复习使用(下载网址:www.zgtjcb.com)。课后习题采用单选题、多选题以及基于实际数据集设计的计算题和操作题等多种形式,以适应今后实现统计学上机考试改革的需求。

全书共分十一章。第一章和第十一章由孙玉环编写,第二章、第五章、第八章和第九章由王雪妮编写,第三章由刘沈忠和王雪妮编写,第四章由王皓编写,第六章和第七章由徐建邦编写,第十章由马晓君编写;另外,研究生韩月、刘婷和吴姝对全书各章的习题进行了整理和补充。全书由孙玉环总纂并定稿。

由于时间和水平有限,本书的错误与疏漏之处在所难免,敬请读者提出宝贵意见和建议。

编者

2015年11月

目 录

第一章 导论	(1)
第一节 统计及其应用领域	(1)
第二节 统计学中的基本概念	(4)
第三节 SPSS 及 EXCEL 基本操作	(9)
本章小结	(21)
习题	(22)
第二章 数据的收集、整理与显示	(26)
第一节 数据的收集	(27)
第二节 数据整理	(32)
第三节 品质型数据的图表描述	(35)
第四节 数值型数据的图表描述	(42)
本章小结	(52)
习题	(53)
第三章 统计数据的描述性分析	(57)
第一节 相对程度的描述	(57)
第二节 集中趋势的描述	(61)
第三节 离中趋势的描述	(73)
第四节 分布形态的描述	(78)
第五节 运用 SPSS 进行统计量描述	(80)
本章小结	(87)
习题	(88)
第四章 概率与概率分布	(91)
第一节 概率基础	(91)
第二节 随机变量及其概率分布	(96)
第三节 大数定律与中心极限定理	(106)
本章小结	(109)
习题	(110)
第五章 参数估计	(112)
第一节 抽样分布	(113)

第二节	点估计	(115)
第三节	单样本总体均值的区间估计	(117)
第四节	两个独立样本总体均值差的区间估计	(121)
第五节	两个匹配样本总体均值差的区间估计	(126)
第六节	样本容量的确定	(130)
第七节	总体方差的区间估计	(132)
第八节	总体比率的区间估计	(134)
第九节	抽样组织方式	(139)
	本章小结	(141)
	习题	(142)
第六章	假设检验	(147)
第一节	假设检验基础知识	(148)
第二节	总体均值的检验	(157)
第三节	总体比率的检验	(167)
第四节	总体方差的检验	(171)
	本章小结	(175)
	习题	(176)
第七章	方差分析	(181)
第一节	方差分析引论	(181)
第二节	单因素方差分析	(184)
第三节	多因素方差分析	(191)
	本章小结	(197)
	习题	(197)
第八章	列联分析	(203)
第一节	列联表	(204)
第二节	列联表中的 χ^2 检验	(207)
第三节	列联表中的相关系数	(212)
	本章小结	(215)
	习题	(215)
第九章	相关与回归分析	(219)
第一节	相关系数	(220)
第二节	回归模型与回归方程	(225)
第三节	估计的回归方程	(227)
第四节	判定系数	(229)
第五节	F 检验	(231)

第六节 回归预测	(233)
第七节 残差分析	(238)
第八节 基于 SPSS 实现回归分析过程	(244)
本章小结	(249)
习题	(250)
第十章 时间序列	(254)
第一节 时间序列构成	(255)
第二节 波动性的描述	(258)
第三节 长期趋势的测定	(263)
第四节 季节波动的测定	(272)
第五节 循环趋势的测定	(279)
本章小结	(281)
习题	(282)
第十一章 指数	(286)
第一节 统计指数的一般问题	(286)
第二节 综合指数	(288)
第三节 平均指数	(291)
第四节 指数体系与因素分析法	(295)
第五节 几种常用经济指数的编制	(302)
本章小结	(306)
习题	(307)
附录 常用统计表	(311)
主要参考文献	(317)

第一章 导论

实践中的统计

理解统计对每个人都是必要的

统计学是一种用来制定决策的重要工具。100多年前,英国作家、历史学家 H. G. Wells 就曾预言说,对于追求效率的公民而言,统计思维总有一天会和读写能力一样必要(Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write)。现在,这一天已经到来。统计学不仅对商业人士有用,每个人在日常生活中也会自觉或不自觉地用到统计概念。了解一些统计知识对每个人来说都是必要的。比如,在外出旅游时,你需要关心一段时间内的详细天气预报;在投资股票时,你需要了解股票市场价格的信息,了解某只特定股票的有关财务数据;在观看世界杯足球赛时,除了关心进球的多少外,你可能还会关心各支球队的技术统计结果,等等。进一步地,要正确阅读并理解下面的一些统计研究结论,就更需要具备一些统计知识:

- 吸烟对健康是有害的。
- 身材高的父亲,其子女的身材也较高。
- 每两天服一片阿司匹林,会减少心脏病第二次发作的几率。
- 身体超重 30%,会使寿命减少 1300 天。
- 每天摄取 500 毫克维生素 C,生命可延长 6 年。
- 上课坐在前面的学生,其平均考试分数比坐在后面的学生高。

看懂这些结论并不困难,但这些结论是怎么得出来的?你相信这些结论吗?学一些统计知识将有助于正确理解它们。本章将从介绍统计学的一些基本问题开始,包括统计学的含义、统计数据及其分类、常用的一些概念、常用统计分析软件的基本操作等。

第一节 统计及其应用领域

一、什么是统计学

要了解什么是统计学,首先要了解什么是统计。在当今信息时代,经常可以从各种媒体上听到或看到各种各样的数据,例如某国家或地区某年国内生产总值(GDP)多少亿元,增长速度百分之几,某地区某年人口出生率千分之几、人均 GDP 多少元等。这些数据意味着什么?为什么要得到这些数据?这都是统计所涉及的问题。

那么什么是统计?根据理解的角度不同,“统计”一词可以有三种含义:统计活动、统计数据和统计学。统计活动是对各种统计数据进行收集、整理并做出相应的推断、分析的活动,通常可以划分为统计调查、统计整理和统计分析三个阶段;统计数据是通过统计活动获

得的,用以表现研究现象特征的各种形式的信息;统计学则是指导统计活动的理论和方法,是收集、处理、分析、解释数据,并从数据中得出结论的一门通用的方法论科学。显然,统计的三种含义,是以统计数据为核心而紧密联系在一起,统计数据与统计活动是统计结果与过程的关系,统计活动与统计学则是统计实践与理论的关系。英文单词“Statistics”有两个含义:用作单数时,表示“统计学”;用作复数时,表示“统计数据”或“统计资料”。可见统计学与统计数据之间具有不可分的关系。

统计学经过 300 多年的发展,已经形成了完整的学科体系。统计学可以分为理论统计学和应用统计学,前者研究一般的收集、整理和分析数据的方法,后者则以各个不同领域现象的具体数量特征为研究重心。从方法论的角度,通常可以把统计学方法概括为描述统计方法和推断统计方法两大类。

(1)描述统计(descriptive statistics)。描述统计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的数量特征。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等。

(2)推断统计(inferential statistics)。推断统计研究如何根据样本数据去推断总体的数量特征,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做出以概率形式表述的推断。比如,要了解一个地区的人口特征,不可能对每个人的特征一一进行测量;再如,对产品的质量进行检验,往往是破坏性的,也不可能对每个产品进行测量,这时需要抽取部分个体即样本进行测量,然后根据获得的样本数据推断所研究总体的特征,这就是推断统计要解决的问题。这里所提到的个体和样本两个概念,我们将在本章的第二节中再进行详细介绍。

二、统计的应用领域

统计是适用于各个学科领域的通用数据分析方法。在信息社会,几乎所有的工作领域都离不开统计方法的应用。具体到经济与管理领域,统计学的应用更是十分广泛。

(1)公共管理领域。决策科学化使统计学在公共管理部门的应用不断增强,对于公务人员来说,如果不能掌握一些搜集和分析数据的技能,将很难适应政府管理信息化和科学化的需要。

(2)经济研究领域。经济研究正在走向定量化和数理化,在进行经济分析和预测时,必然要用到各种统计数据和统计方法。

(3)产品质量管理。质量是企业的生命,质量管理与控制已经成为统计学在生产领域中的一项重要应用,各种统计方法已经被广泛应用于生产过程的监测活动中。

(4)市场调研领域。企业要及时应对市场的变化,才能在竞争中取得优势,而了解市场必然要涉及市场信息的搜集和分析过程,统计方法在这方面的应用也极为广泛。

(5)人力资源管理。管理人员可借助统计方法全面掌握本单位的人力资源状况,设计更为合理的绩效考评机制等。

(6)金融保险领域。随着金融保险行业的快速发展,金融与保险数据的规模和复杂程度都急剧增大,随之而来的相关数据处理工作,使金融保险行业对统计人才的需求与日俱增,保险精算人才更是供不应求。

此外,在医学领域,人们利用统计方法研究疾病的原因或影响因素、判断药物和治疗方法的有效性;在考古领域,人们利用统计方法推断特定文物的历史年代;在心理学领域,人们利用统计方法分析特定刺激的心理效应;在气象学领域,人们利用统计方法预测未来的天气情况;在物理学领域,人们利用统计方法对由大量粒子组成的宏观物体的物理性质及宏观规律做出微观解释;在生物学领域,人们利用统计方法研究基因定律,等等。

不过需要注意的是,统计也不是万能的,它并不能解决人们面临的所有问题。统计方法可以帮助人们通过分析数据得出某种结论,但对统计结论的进一步解释,还是需要相应的专业知识。比如,吸烟会使患肺癌的几率增大,这只是一个统计结论,但要解释吸烟为什么可能导致罹患肺癌,还是需要更多的医学知识才行。

三、统计研究过程

尽管在实际的统计研究活动中,由于研究的对象不同,具体的统计内容会千差万别,但作为从数量方面认识事物的一门方法论学科,一项完整的统计研究过程,一般包括以下四个基本环节:统计设计、数据收集、数据整理以及数据分析与解释。

(一)统计设计

统计设计是对整个统计研究过程所涉及的各个环节进行的通盘考虑和事先安排。统计设计应在相关学科的理论指导下,根据研究问题的性质、目的和任务,科学地确定统计研究的对象,明确所要收集数据的种类,确定相应的统计指标及其体系并给出统一的定义和标准,确定统计数据收集、整理、推断和分析的基本方法,规定研究工作的进度安排和质量要求,拟定研究工作的资源配置和组织实施方式等。统计设计对于统计数据分析工作的研究质量至关重要,要求设计者不仅应掌握系统的统计学理论和方法,而且还应具有所研究领域的相关知识和理论素养。

(二)数据收集

数据收集是按照统计设计的要求,通过统计观测或实验的方式、方法,收集统计研究所需要的原始数据以及其他已经存在的各种相关数据的过程,也称为统计调查。数据收集是否准确、及时、完整,将直接影响到统计分析的质量。

(三)数据整理

数据整理是对通过统计观测或实验所获得的原始数据,进行必要的系统化处理,使之条理化、综合化,成为能够反映总体特征的统计数据的过程,也称为统计整理。数据整理环节也包括对已有数据的再加工和深加工。通过统计分组、汇总和计算等数据整理的手段,可以得到能够一定程度上反映总体特征的统计图、统计表或统计指标等。

(四)数据分析与解释

数据分析是在数据整理结果的基础上,围绕统计设计所确定的研究任务,运用各种统计方法对数据进行分析,得出某些有用的定量结论的过程,也称为统计分析。数据分析实质上是对数据进行深加工的整理过程,是整个统计研究工作的核心,也是统计研究的最终目的。在此环节,既要用到描述统计方法,也可能用到推断统计方法。数据解释则是对数据分析结果的进一步说明,即说明为什么会得出这些数据分析结果,可以进一步得出哪些规律性的结论,以及需要进一步探讨的问题等,数据解释是对数据分析过程的深化。

第二节 统计学中的基本概念

为便于后续在各章中对统计学基础原理及方法进行系统的讲解,有必要先来熟悉一下统计学中的若干常用术语。

一、总体与样本

(一) 总体

统计方法的有效运用需要建立在研究对象的大量性的基础之上。总体(population)是指包含所研究对象的全部个体(数据)的集合,通常由所研究的一些个体组成,如由多个企业构成的集合,多个居民户构成的集合,多个人构成的集合,等等。组成总体的每一个元素称为个体(unit)或总体单位(population unit),在由多个企业构成的总体中,每一个企业就是一个个体;由多个居民户构成的总体中,每一个居民户就是一个个体;由多个人构成的总体中,每一个人就是一个个体。

总体是统计研究的对象,是统计学中最基本的要素和概念。总体范围的确定依赖于研究目的,研究目的一经确定,相应的总体也就唯一地确定。比如,要检验一批灯泡的使用寿命,这一批灯泡构成的集合就是总体,每一个灯泡就是一个个体。

根据所包含的单位数目是否可数,可以把总体划分为有限总体和无限总体。有限总体指总体的范围能够明确界定,而且所包含个体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,一批待检验的灯泡也是有限总体。无限总体指总体中所包括的个体数目是无限的、不可数的。例如,在科学实验中,每一个实验数据可以看作是研究总体的一个个体,而实验可以无限地进行下去,因此由实验数据构成的总体就是一个无限总体。

把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,每次抽取一个个体并不影响下一次的抽样结果,因此每次抽取可以看作是独立的。对于有限总体,抽取一个个体后,总体元素就会减少一个,前一次的抽样结果会影响第二次的抽样结果,因此每次抽取不是独立的。抽取个体的过程是否独立,会影响到抽样推断的结果。相关的内容我们将会在第五章“参数估计”部分中再进行详细介绍。

(二) 样本

样本(sample)是从统计总体中抽取出来作为代表这一总体的、由部分个体组成的集合,构成样本的个体数目称为样本量(sample size)。统计学的研究目的是把握研究总体的数量特征,但当总体所包含的个体单位数量很多,甚至无限时,就没有必要或无法对构成总体的所有单位进行研究。这时,可以抽取总体中的部分个体组成样本,根据样本提供的信息推断总体的特征。比如,从一批灯泡中随机抽取 100 个,这 100 个灯泡就构成一个样本,然后根据这 100 个灯泡的平均使用寿命可以推断这批灯泡的平均使用寿命。本教材第五章“参数估计”部分,将主要介绍利用样本的数量特征来估计总体数量特征的问题。

二、参数与统计量

(一) 参数

参数(parameter)是用来描述总体特征的概括性数字度量,它是研究者想要了解的总体

的某种特征表现。研究者所关心的参数通常有总体平均数、总体标准差、总体比例等。在统计中,总体参数通常用希腊字母表示。比如,总体平均数用 μ 表示,总体标准差用 σ 表示,总体比例用 Π 或者 P 表示,相关系数用 ρ 表示,等等。

由于总体数据通常是未知的,所以参数通常是一个未知的常数,统计活动的目的就是为了解搜集或推断这一总体数值。比如,为了解某一地区所有人口的平均年龄、某个城市所有家庭的收入差距、某批产品的合格率等,通过抽样并根据样本的计算结果可以估计总体参数。

(二) 统计量

统计量(statistic)是用来描述样本特征的概括性数字度量,是根据样本数据计算得到的一个量,由于抽样是随机的,因此统计量是样本的函数。研究者所关心的统计量主要有样本平均数、样本标准差、样本比例等。样本统计量通常用英文字母来表示,比如,样本平均数用 \bar{x} 表示,样本标准差用 s 表示,样本比例用 p 表示,等等。

由于样本是实际抽取出来的,所以统计量总是可以计算。抽样的目的就是要根据样本统计量去估计总体参数。比如,用样本平均数(\bar{x})去估计总体平均数(μ),用样本标准差(s)去估计总体标准差(σ),用样本比例(p)去估计总体比例(P),等等。

除了样本均值、样本比例、样本方差这类统计量,还有一些是为了进行统计分析的需要而构造出来的统计量,比如用于统计检验的 z 统计量、 t 统计量、 χ^2 统计量、 F 统计量,等等。

上述所提到的这些统计量,我们也都将会在后续的各章节中展开详细介绍。

三、统计标志与统计指标

统计学的研究对象是总体的数量方面,但由于总体的数量特征是无法直接计量的,只有充分认识构成总体的个体,才能对总体进行特征描述。

(一) 统计标志

统计标志简称标志(characteristic),是指每个个体所共同具有的属性或特征,是说明个体的属性或特征的具体名称;对应于某个标志,各个个体的具体表现,即为标志值。由于各个个体就某些标志而言,具有相同的标志值,才构成了统计总体;也正是由于各个个体就另外一些标志而言,具有不相同的标志值,才有了统计的必要。例如,要研究某高校女生的身体素质,显然该高校所有女生构成总体,每名女生是一个个体。反映个体的标志可以有很多,比如所在学校、家庭住址、性别、民族、年龄、身高、体重,等等,对每名女生来说,她们的“所在学校”和“性别”所对应的标志值都是相同的,这是构成总体的条件,即同质性;而其他标志的标志值就不会完全相同,这是要调查研究的内容。显然,差异是需要重点关注的。统计研究总是从登记标志开始,并通过对标志的综合来反映总体的数量特征,因此标志是统计研究的起点。

标志按其性质不同,可以分为品质标志和数量标志。品质标志是表明个体属性方面的特征,品质标志的标志表现只能用文字说明,不能用数量表示,比如企业性质、所属行业等;数量标志表明的是个体的数量特征,其标志表现只能用数字表示,如职工人数、产值、销售收入、工资总额等。

标志按变异情况不同,可以分为不变标志和可变标志。不变标志是指某一标志对所有个体而言,具体表现都是相同的。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,它体现为总体的同质性。例如,以

某市所有的中小民营企业为研究总体,这里的各民营企业均具有所有制、企业规模这两个不变标志;如果某家企业并不具备这两者中的任何一个方面的特征,则该企业就不能成为该总体中的一个单位。可变标志,又称变异标志,是指某一标志在各个个体之间的具体表现不尽相同。一般来说,组成总体的各个总体单位具有多个可变标志。例如,把某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、固定资产总额、流动资产总额、工资总额、工业总产值、平均工资等,都是该总体中各个单位的可变标志。

有时会按照某一个品质标志,将总体划分为具有某一特征的个体的集合和不具有某一特征的个体的集合两类。如在全部产品中,分为合格品和不合格品;在全部人口中,分为男性和女性等。这种用“是”或“否”、“有”或“无”来表示的标志,叫作是非标志,也叫交替标志或0/1标志,其标志表现只有两种情况,即表现为每个总体单位要么具有该种属性,要么不具有该种属性。

（二）统计指标

统计标志描述的是个体的属性或特征,而统计指标(indicator)则是反映社会经济现象总体数量特征的概念及其具体数值,由各个个体的标志值汇总综合而成。统计指标包括指标名称和指标数值两部分,例如:2014年全国工业增加值为22.8万亿元,这一统计指标既包括概念“工业增加值”,又包括具体数值“22.8万亿元”,它还表明了在一定时间(2014年)和空间(全国)条件下的现象的数量,也就是说,统计指标不能离开时间和空间而存在。统计指标具有数量性、具体性和综合性的特点。

统计口径是对计算口径外延的明确定义,如果指标所包括的范围模糊,必然会造成所收集数据资料的不准确,进而可能导致得出错误的分析结论。比如,计算劳动生产率指标时,必须明确分母是指生产工人还是指全体职工,否则会将生产工人劳动生产率和全员劳动生产率相混淆。按照不同的分类依据,可以把统计指标划分成不同的类型。

按照表现形式的不同,可以把统计指标分为绝对指标、相对指标和平均指标。其中,绝对指标,又称作绝对数,反映的是现象总体的绝对数量特征,它以最直观、最基本的形式表现现象的外在数量特征,有明确的计量单位。例如某人的身高为176厘米、某地区国内生产总值1250亿元等。相对指标,又称作相对数,反映的是现象之间的相对数量特征,它通过两个具有一定联系的统计指标的对比,来体现现象内部或现象之间的关联关系,比如,某地区的人口密度为300人/平方公里。平均指标,又称作平均数,反映的是现象总体的平均数量特征,用于表明现象的一般数量水平。例如某高校某个班级学生的平均年龄为18.8岁,某年某地区的居民平均月消费支出2800元等。相对指标用于描述现象之间的相对程度,平均指标用于描述现象的集中趋势,关于相对指标和平均指标的具体分类及其计算方法和应用,我们将分别在第三章的第一节“相对程度的描述”和第二节“集中趋势的描述”中,再进行详细介绍。

按照所反映问题的数量特征不同,可以把统计指标分为数量指标和质量指标。数量指标是反映研究总体的规模或水平的指标,是由个体标志汇总后直接得到的绝对数或总量,也称作总量指标。数量指标既包括标志总量,如总产值、总收入等,也包括总体单位总量,如企业总数、总人数等。数量指标是最基本的统计指标,是计算其他指标的基础,数量指标的外延与绝对指标的外延是完全一致的。质量指标是反映客观现象之间的相互联系、比例关系、发展速度、内部结构等的指标,一般用相对数或平均数表示。如劳动生产率、成本利润率、人均国民收入等。质量指标由两个数量指标对比得到,是数量指标的派生指标。

按照所反映现象的时间状态不同,可以把统计指标分为静态指标和动态指标。静态指标是反映现象总体在某一时刻或相对静止时间上数量特征的指标,包括一般的总量指标、静态相对指标和一般平均指标。动态指标是反映现象总体在不同时期或时点上发展变化情况的指标,包括增长量指标、动态相对指标和序时平均指标等。关于动态指标的计算方法及其应用我们将在第十章“时间序列分析”中具体介绍。

(三) 统计指标与标志的关系

统计指标与标志二者之间既有区别又有联系。区别主要表现在两个方面:一是说明的对象不同,指标说明的是总体的特征,而标志则反映的是个体的特征;二是表现形式不同,指标只能用数值来表现,而标志则既有用文字来表现的标志,也有用数值来表现的数量标志。二者之间的联系主要表现为,统计指标的数值是根据说明个体特征的标志表现汇总得到的,既可以是对总体各单位标志值的综合,也可能是总体单位数的汇总。

四、变量及其测量尺度

(一) 变量及其基本分类

与最初的研究目的相关联,我们总是对总体及其所包含的个体的某一个或某几个方面的属性感兴趣,这些属性在统计中又被称作变量(variable)。总体中每一个个体的属性是可以无限列举的,但构成统计中变量的属性只是那些我们感兴趣的,比如灯泡的使用寿命、质量等级,在校学生的性别、身高、体重,等等,都可以成为统计研究中的变量。

变量值的变动受诸多因素的影响。如果数值是随机生成的,有多种可能性,事先无法确定,这样的变量称为随机变量,如中奖号码等;如果受某些确定性因素影响,现象的变量会沿着某一方向持续变化,这样的变量就是确定性变量,如中奖人数等。变量按其取值是否连续,又可以分为离散变量和连续变量。变量是离散型还是连续型,以变量值是否可以无限分割为判断标准。如果变量的取值可以一一列举,在相邻的两个变量值之间不可能分割出新的变量值,即变量值只能表现为整数,这样的变量称为离散变量,如在校学生人数、全国规模以上工业企业数等,离散变量通常以计数的方法取得;如果变量的取值不能一一列举,在相邻的两个变量值之间可以继续分割,取得新的变量值,即变量值理论上可以取任何小数值,这样的变量称为连续变量,比如劳动生产率、人的身高和体重等,连续变量通常需要用计算或测量的方法取得变量值。

(二) 变量的测量尺度

我们在提到变量及其变量值的时候,总是与其测量尺度分不开的。测量本身是一个物理学中的概念,是指用仪器确定空间、时间、温度、速度、功能等的有关数值。统计测量则是指运用某种方法(分类或标识)使自然或社会经济现象数量化,测量结果即形成变量。统计测量尺度可以区分为四种类型,即定类尺度、定序尺度、定距尺度和定比尺度。

1. 定类尺度

定类尺度(nominal scale)是按照事物的某种属性对其进行的平行分类,是对现象的性质差异进行的辨别与区分,测量结果即形成定类变量或定类指标。定类尺度是最简单、也是层次最低的一种测量尺度,定类变量确切的值是以文字表述的,各类别间是平等的并列关系,不存在顺序上的差异,没有高低、大小、优劣之分。例如,人口按照性别分为男、女两类,企业按行业属性分为医药企业、家电企业、纺织品企业等,这些均属于定类变量。在实际工