

· STATISTICAL ·
ANALYSIS

统计分析方法与应用研究

綦 路 著

TONGJI FENXI FANGFA YU YINGYONG YANJIU

IC 吉林科学技术出版社

统计分

统计分析方法与应用研究

綦路 著

 吉林科学技术出版社

图书在版编目(CIP)数据

统计分析方法与应用研究 / 綦路著. —长春 :
吉林科学技术出版社, 2019. 10
ISBN 978-7-5578-6112-4

I. ①统… II. ①綦… III. ①统计分析—分析方法 ②
统计分析—应用—研究 IV. ①C81

中国版本图书馆 CIP 数据核字(2019)第 233227 号

TONGJI FENXI FANGFA YU YINGYONG YANJIU 统计分析方法与应用研究

著 綦 路

出版人 李 梁

责任编辑 李思言

封面设计 崔 蕾

制 版 北京亚吉飞数码科技有限公司

开 本 710mm×1000mm 1/16

字 数 242 千字

印 张 13.5

印 数 1—5 000 册

版 次 2020 年 3 月第 1 版

印 次 2020 年 3 月第 1 次印刷

出 版 吉林科学技术出版社

发 行 吉林科学技术出版社

地 址 长春市人民大街 4646 号

邮 编 130021

发行部传真/电话 0431—85635176 85651759 85635177

85651628 85652585

储运部电话 0431—86059116

编辑部电话 0431—85635186

网 址 www.jlscbs.net

印 刷 三河市铭浩彩色印装有限公司

书 号 ISBN 978-7-5578-6112-4

定 价 80.00 元

如有印装质量问题 可寄出版社调换

版权所有 翻印必究 举报电话:0431—85635186

前 言

21 世纪是信息经济时代,它的显著特点就是信息爆炸、泛滥。随着信息化技术的迅猛发展,数据已经成为炙手可热的高频词;在“大数据”概念火热的当今,数据处理和运用已经成为全社会的热门话题。要想掌握新经济时代的制胜先机,就需要从这些杂乱无章的信息中看出端倪。从本质上讲,信息经济所依赖的不只是信息处理手段的先进,更重要的是信息收集、整理、分析的准确性和可靠性,而统计学学科的发展与准确、可靠的信息收集、整理和分析有着密不可分的关系。统计学就是在这种情形下脱颖而出的。

统计学的研究对象是客观现象的总体数量特征和数量关系,从数据的获取到数据的运用都是统计学的研究范畴,统计学侧重揭示客观事物在数量上的客观规律。统计学是能够帮助人们把握复杂信息现象的一种有力工具。正是因为如此,实用的统计概念与方法成为社会各行各业的自然需求。在一定程度上而言,统计学对信息经济的发展具有重要的支撑作用。

正是出于这样的需要,作者经过长时间的探索和研究,结合自身多年从事概率论与数理统计学、计量经济学、应用统计学工作的经验,面向应用,从实际需求出发,又不失系统性和创新性,写作了这本《统计分析方法与应用研究》。本书共七章,主要内容包括统计数据的搜集、整理与汇总,概率与概率分布,参数估计和假设检验,相关分析与回归分析,时间序列分析,统计比率与统计指数。全书运用大量的图表来阐述概念,通过实例进一步理解较复杂的数学结构,以达到举一反三的效果。本书注重应用,将统计指标及统计分析方法通过范例演练,更方便统计学学习,能供人们较易体会统计学的应用精髓。

在本书的写作过程中,作者结合自身研究及教学实践经验,遵循以下几个原则:

(1)不失一般系统性原则,内容、结构与体系安排按照分层递进、由浅入深的思路,章节段落力求条理清晰,文字表达力求简洁明快,反映统计学原理的基本内容。

(2)做到基本概念准确,叙述简洁,反映思想方法和过程。

(3)注重方法技术,把数学理论推导通俗化。在理论与方法的阐述上,既有严谨规范的专业术语,又有通俗易懂的直观描述,借以帮助不同层次、

不同专业背景的读者提高学习效率。

(4)注重实际运用,书中配备了一些富有启发性的例题,有利于培养统计思维方式、综合应用能力以及开拓创新能力,把应用统计学的运用落到实处。

本书深入研究了国内外统计学教育教学改革的成果,借鉴了众多优秀著作的内容结构和处理方法。本书的写作得到了学院同事及出版社的支持和帮助。同时,作者参阅和借鉴了国内外不少专家学者和同行的著述和文献。在此,一并致以真诚谢意。由于作者水平所限,加之许多主客观的因素,书中难免有疏漏甚或错误之处,敬请读者不吝赐教。以臻完善。

作者

2019年7月

目 录

第一章 绪论	1
第一节 统计的产生与发展	1
第二节 统计研究的对象、方法和作用	5
第三节 变量与数据的分类	9
第四节 统计学的应用领域	12
第二章 统计数据的搜集、整理与汇总	15
第一节 统计数据的搜集	15
第二节 统计数据的整理	18
第三节 频数图	22
第四节 统计数据分布特征的综合分析	25
第五节 典型应用实例	39
第三章 概率与概率分布	41
第一节 随机事件及概率	41
第二节 概率计算规则	43
第三节 随机变量的概率分布	58
第四节 随机变量的数据特征	68
第五节 典型应用实例	79
第四章 参数估计和假设检验	83
第一节 点估计与参数的区间估计	83
第二节 假设检验的一般方法	96
第三节 几种典型的假设检验	101
第四节 置信区间与假设检验之间的关系	108
第五节 非参数检验	109
第六节 典型应用实例	113

第五章 相关分析与回归分析	115
第一节 相关分析概述	115
第二节 线性相关分析与非线性相关分析	117
第三节 一元线性回归分析的评价和检验	120
第四节 多元线性回归分析的描述和推断	127
第五节 多元回归分析中的 F 检验	132
第六节 关于多元回归模型假定条件的讨论	133
第七节 多元回归和多重相关分析的综合运用	139
第八节 回归分析的典型应用实例	145
第六章 时间序列分析	147
第一节 时间序列及其分类	147
第二节 时间序列的分析指标	148
第三节 时间序列的构成分析	161
第四节 平稳时间序列的预测	165
第五节 复合型时间序列的预测	167
第六节 时间序列分析的软件实现	176
第七章 统计比率与统计指数	177
第一节 统计比率	177
第二节 统计指数	181
第三节 指数体系与因素分析	189
第四节 价格指数的编制与应用	195
参考文献	206

第一章 绪 论

统计学是通过搜索、整理、分析、描述数据等手段,以达到推断所测对象的本质,甚至预测对象未来的一门综合性学科。目前,统计学中大量的定量分析方法,被广泛地应用在各门学科之中,从自然学科、社会科学到人文科学,甚至被用在工商业及政府的情报决策之中。

第一节 统计的产生与发展

统计作为一种社会实践活动已有悠久的历史。在英语中“统计”一词与“国家”(state)一词来自同一词源。因此,可以说自从有了国家就有了统计实践活动。最初,统计只是一种为满足统治者管理国家的需要而搜集资料,弄清国家的人力、物力和财力,作为国家管理依据的计数活动。而在现代社会中,“统计”已不再是统治者的专利,从国家管理者到一般平民百姓,都已把“统计”以及由此而得到的各种结果——统计信息运用到国家管理、社会经济活动及日常生活中,成为参与上述各种活动的工具和依据。

今天,“统计”已被赋予多种含义。在不同场合,“统计”一词可以有不同的含义。它可以指统计数据的搜集活动,即统计工作;也可以指统计活动的结果,即统计数据或统计信息;还可以指分析统计数据的方法和技术,即统计学。

统计学(statistics)是随着人类社会的发展和社会管理的需要而发展起来的。随着社会管理的日趋复杂,仅用数字计量客观现象已不能满足社会的需要,人们试图对客观现象进行定量分析,与此相适应,研究数据计量和分析方法的统计学诞生了。目前,随着统计方法在社会科学、自然科学和工程技术等领域的广泛应用,统计学已发展为具有众多分支学科的大家族,因而人们也从不同的角度对统计学的概念给出许多大同小异的解释。在这里,综合各种解释,我们认为:统计学是一门关于客观现象数据搜集、整理、归纳、分析的方法论学科,其目的是探索数据的内在数量规律性,以达到对客观事物的科学认识。

统计学的产生与统计实践活动是密不可分的,统计作为一种社会实践

活动,已有四五千年的历史。据历史记载,在远古时代有“结绳记事”“刻木记数”之说,其实这可以说是最原始的统计。随着人们在这数千年的统计活动中的不断探索,对统计规律的认识也逐渐加深,这就产生了统计学。一般认为,统计学产生于 17 世纪中期,至今已有 300 多年的历史。统计学的产生与形成是从几个不同领域开始的,主要源自于英国的政治算术、德国的国势学和概率论。并在此基础上产生和形成了“政治算术学派”“国势学派”“数理统计学派”和“社会统计学派”等。

政治算术产生于 17 世纪中期的英国,其代表人物有约翰·格朗特(John Graunt)和威廉·配第(William Petty)。约翰·格朗特于 1662 年出版了《对死亡表的自然观察和政治观察》一书。书中他通过大量观察,研究并发现了一系列人口统计的规律,例如男性多于女性,在伦敦其比例为 14:13,在地方其比例为 15:14;地方的出生率大于死亡率;一般疾病和事故的死亡率比较稳定,而传染病的死亡率波动较大等。与格朗特同时代的英国经济学家威廉·配第的《政治算术》(1676)和其对国民收入估算的方法,对经济学、统计学的发展具有重大意义。《政治算术》中用于探索社会和经济现象的数量规律方法,最早地体现了统计学的基本思想。

从 17 世纪中期起,在德国兴起的国势学派用记述的方法来描述和研究一国的人口、财政、军事地理、政治和法律制度等。到 18 世纪,哥廷根大学教授阿亨瓦尔(Gottfried Achenwall)首次提出了“统计学”[Statistik,(英)Statistics]这一名词,并将其定义为关于“国家显著事项”的学问,通过这门科学可以清楚地了解国家的运行情况,这与当今的《官方统计》的内容基本一致。国势学派最初的“统计”只用文字记载,在后来发展中使用了数字和图表。

统计学的另一个重要源头是以概率论的研究为开端,并以其为基础形成和发展起来的以随机现象为主要研究对象的数理统计学。

概率论的产生起源于对赌博中掷骰子输赢问题的研究,其目的是找出一套办法,合理算出取胜的机会,早在但丁(Alighieri Dante)的《神曲》中就记载了投掷三个骰子所可能出现的结果,后来到 16 世纪中期,意大利数学家卡达罗(Kirolama Cardano)写出《论赌博》一书,为运用数学理论研究概率开辟了道路。几乎在同一时期,意大利天文学家伽利略(Galilei Galileo)也进行了相同的研究,提出了概率论的基本原理。但概率论的真正历史始于 17 世纪中期,法国数学家帕斯卡尔(Blaise Pascal)和费尔马特(Pierre de Fermat)通过通信方式,将赌博中出现的具体问题归纳为一般的概率原理,为以后的概率论和统计学的发展奠定了基础。直到 18 世纪初,瑞士数学家贝努里(Jakob Bernoulli)的《推算法》问世,概率论作为一门理论才初步形

成。在此之后,概率的乘法运算法则、正态分布、先验概率等理论相继问世,使概率论得以快速发展起来。

概率论的发展和在社会现象研究中的应用促进了具有科学推论特色的近代统计学的形成。到 19 世纪,用概率论研究社会各种现象的人日益增多。在这方面对近代统计学的形成做出重大贡献的有法国数学家拉普拉斯(Pierre Simon Laplace)、德国数学家高斯(Johann Karl Friedrich Gauss)和比利时统计学家凯特勒(Adolphe Quetelet)。特别是凯特勒,他以大量观察和综合平均的方法对社会现象进行研究,从而将概率论(正态分布等)、国势学和政治算术中用于观察群体现象并进行数量分析的方法融合为统计学。

20 世纪初到现在的数理统计学时期称为现代统计学时期。在这一时期,数理统计在随机抽样基础上建立起推断统计学。现代数理统计学分为数理统计学理论部分和应用部分。前者包括抽样理论、估计理论、假设检验理论、实验设计、决策理论、非参数统计、序贯分析、博弈理论等。后者包括计量经济学、生物统计、统计力学、质量管理、政府统计、遗传统计、天文统计等。

20 世纪 60 年代以后,数理统计学的发展有三个明显的趋势:

(1)随着数学的发展,数理统计学越来越广泛地应用数学方法。R. A. Fisher 对小样本理论做了进一步研究,创立了实验设计理论,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。尼曼(Neyman)和皮尔逊(Pearson)共同对假设检验理论做了系统的研究,创立了“尼曼—皮尔逊”理论,同时尼曼又创立了区间估计理论。G. W. Snedcor 将 Z 分布转换成 F 分布,A. Wald 提出决策理论和序贯抽样法。J. V. Neumann 和 O. Morgenstern 提出博弈论,使决策理论更加系统化。N. Wiener 的控制论和 C. E. Shannon 的信息论使得推断统计学的理论更加健全。

(2)数理统计学的新分支或以数理统计学为基础的边缘学科不断形成。作为统计学分支的新的研究领域层出不穷,如多元统计分析、随机过程、非参数统计等。随着社会经济和金融交易的多样化,国际市场间资本流动的迅猛增长,以及电子商务、云计算的出现,统计学及其应用将会继续发展扩大。

(3)数理统计学的应用日益广泛深入。尤其是借助电子计算机后,数理统计学发挥的作用日益增强。诸如 Excel、SAS、SPSS、Stata、Eviews 等统计软件在统计分析中正在发挥其强大的统计功能。

①Excel。Microsoft Excel 是美国微软公司开发的 Windows 环境下的电子表格系统,它是目前应用最为广泛的办公表格处理软件之一。Excel

具有强有力的数据库管理功能、丰富的宏命令和函数,是强有力的决策支持工具。

②SPSS(Statistical Package for Social Science)是当今世界上最优秀的统计软件之一,其具有统计方法先进成熟、操作简便、与其他软件交互性好的特点,被广泛应用于经济管理、医疗卫生、自然科学等各个方面。SPSS 历经多次改版,目前最新版本为 SPSS 22.0。

SPSS 具有强大的数据管理功能,共计 28 种,包括定义变量属性(Define Variable Properties)、设置未知测量级别(Set Measurement Level for Unknown)、复制数据属性(Copy Data Properties)、新建定制属性(New Custom Attribute)、定义日期(Define Dates)、定义多响应集(Define Multiple Response Sets)、验证(Validation)[具体包括加载预定义规则(Load Predefined Rules)、定义规则(Define Rules)和验证数据(Validate Data)]、标识重复个案(Identify Duplicate Cases)、标识异常个案(Identify Unusual Cases)、比较数据集(Compare Datasets)、排序个案(Sort Cases)、排序变量(Sort Variables)、变换(Transpose)、合并文件(Merge Files)[包括添加个案(Add Cases)、添加变量(Add Variables)]、重组(Restructure)、搜索权重(Rake Weights)、倾向得分匹配(Propensity Score Matching)、个案控制匹配(Case Control Matching)、汇总数据(Aggregate Data)、拆分为文件(Split Into Files)、正交设计(Orthogonal Design)、复制数据集(Copy Dataset)、拆分文件(Split File)、选择个案(Select Cases)及加权个案(Weight Cases)等。

SPSS 共提供 19 种数据变换功能,包括计算变量(Compute Variable)、可编程性变换(Programmability Transformation)、对个案内的值计数(Count Occurrences of Values within Cases)、转换值(Shift Values)、重新编码为相同变量(Recode into Same Variables)、重新编码为不同变量(Recode into Different Variables)、自动重新编码(Automatic Recode)、创建虚拟变量(Create Dummy Variables)、可视分箱化(Visual Binning)、最优分箱化(Optimal Binning)、准备建模数据(Prepare Data for Modeling)[包括交互式数据准备(Interactive Data Preparation)、自动数据准备(Automatic Data Preparation)及逆转换得分(Backtransform Scores)]、个案等级排序(Rank Cases)、日期和时间向导(Date and Time Wizard)、创建时间序列(Create Time Series)、替换缺失值(Replace Missing Values)、随机数字生成器(Random Number Generators)及运行挂起的转换(Run Pending Transforms)。

SPSS 统计分析方法共有 24 大类,120 个小类。

SPSS 的绘图功能非常强大,在统计分析过程中可选择多种图形,也可

直接由绘图菜单中产生,并加以修饰、编辑。

③SAS。SAS(Statistical Analysis System)是全球最大的软件公司之一,是由美国 North Carolian 州立大学 1966 年开发的统计分析软件。SAS 系统基本上可以分为四大部分:SAS 数据库部分,SAS 分析核心,SAS 开发呈现工具,SAS 对分布处理模式的支持及其数据仓库设计。

④Stata。Stata 是一种功能全面的统计软件包,它具有容易操作,运行速度快,功能强大的特点。Stata 不仅包括一整套预先编排好的分析与数据功能,同时还能允许软件使用者根据自己的需要来创建程序,从而添加更多的功能。作图功能、参数检验功能和回归分析功能是其强项,Stata16 是目前最新版本。

⑤Eviews。Eviews 是专门为大型机构开发的、用以处理时间序列数据的时间序列软件包的新版本。Eviews 的前身是 1981 年第 1 版的 Micro TSP。Eviews 相比于其他软件的优势在于 Windows 的可视化操作优点,可以使用鼠标对 Windows 菜单和对话框进行操作;Eviews 还拥有强大的命令功能和批处理语言功能,能提供基于 Windows 平台的复杂的数据分析;Eviews 相比于其他计量软件,可以更便利的从键盘、磁盘文件得到数据,对数据进行统计分析和相关分析,并根据这些统计关系进行预测。

第二节 统计研究的对象、方法和作用

一、统计学的研究对象

由统计学的概念可知,统计学的研究对象是具有某种相同属性的群体现象,以及探索群体现象数量表现的内在规律性及对群体现象进行计量描述和分析推断的方法。它包含了统计学的三个核心要点。

(一)研究对象是群体现象

总体、样本和个体是统计学中三个最基本的概念。总体是指研究对象全体的集合;样本是指来自总体中的部分对象的集合,个体是总体中的任何一个元素(称为总体单位)。统计学的研究对象是这些集合而不是组成这些集合的一个个元素。例如,如果要知道某一物体的重量,只要把它称一称即可得知,无须进行统计分析。然而,若要知道一批物体的总重量与平均重量,一群人的最高者、最低者和这群人的平均身高,就需要测量汇总,并采用

一定的统计方法进行平均。若期望通过其中的小部分,即样本的信息来了解总体,则需要抽样、估计检验、统计分析推论的一套过程,而这些正是统计学的内容。尽管总体或样本的信息都表现在一个个元素或个体上,研究总体离不开个体或部分个体,但统计学研究的不是个体现象,而是通过个体或部分个体所载有的信息来研究、认识和说明总体现象。

(二)探索群体现象数量表现的内在规律性

统计探索的是群体现象数量表现的内在规律性,这里有两个要点:数量性和规律性。

(1)数量性。统计学“Statistics”一词有两个含义:当其以单数名词出现时,表示作为一门学科的“统计学”;当其以复数名词出现时,则表示“统计数据”或“统计资料”。统计学的这一名词特性,直观地说明了统计学与统计数据之间密不可分的关系。统计学由一套搜集和处理统计数据的方法所组成,这些方法源自于对统计数据的研究,目的也在于对统计数据进行研究。所以,离开了统计数据,统计方法也就失去了用武之地,统计学也就失去了其存在的意义。而统计数据如果不用统计方法进行分析研究,也就仅仅是一堆数字而已,得不出任何有益的结论。

此外,统计数据在英文中是以复数形式出现的,这就表明统计数据不是单个的数字,而是由多个数据构成的数据集。统计学是用数字来说明群体现象特征的,作为特征,既可以是其数量方面的特征,如人的身高、年龄等;也可以是属性方面的特征,如人的性别、民族等。统计学所研究群体现象的特征,总是用数字来计量、说明的。例如,统计学所探索的某人口群体的性别特征,则表现为不同性别的人口在总人口中的数量、比例;民族特征则表现为各民族人口在总人口中的数量、比例等。

(2)规律性。统计学研究的是各个体之间在数量上存在差异的群体现象。如果一批物体中的每件重量都一样,一群人中的每个人身高都相同,也是用不着进行统计研究的。然而,现实中群体现象总是由多数量特征各异的个体所组成。而这些有差别的个体数量特征中却存在着某一数量规律性。统计学提供了探索数据内在规律性的一套方法。什么是统计数据的内在数量规律性?为什么统计方法能够通过数据的分析找出其内在的数量规律性?下面的例子可以说明这些问题。

(三)对群体现象数量特征进行计量描述和统计推断

统计学对群体现象数量特征的计量描述,是指通过对客观现象的数据搜集,并通过图表等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所研究群体现象的规律性数量特征。其内容包括统计数据的搜集方法、数据的加工处理(或整理)方法、数据的显示方法、数据分布特征的概括与分析方法。

统计学对群体现象数量特征的统计推断是研究如何通过随机抽样的样本数据的特征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作出以概率形式表述的推断。

二、统计学研究方法

统计学是自成体系的学科,有自己的研究对象,相应地也有专门的研究方法。理解和掌握统计方法对于认识统计对象的数量规律性和完成统计工作有着重要的意义。本节从总体上介绍统计研究方法的特性,具体的方法将在以后的各章进行详细讨论。

(一)大数定律的方法论

统计研究现象总体的数量特征,其所用的基本方法都与数量的总体性有关,其数学依据是大数定律。

大数定律又称大数法则,它是说明大量的随机现象的平均结果具有稳定性质的法则。即说明如果被研究的总体数量特征是由大量的相互独立的随机变量形成的,每个变量对总体的影响都相对较小,那么对大量随机变量加以综合平均的结果,会使变量的个别影响相互抵消,而显示出它们共同作用的倾向,使总体数量特征具有稳定的性质。大数定律从数量方面论证了现象的偶然性和必然性之间的内在关系。也就是说,表面上看似偶然的现象,都会受内部隐蔽的规律所支配。因而,可以通过对大量随机现象的综合概括,消除偶然性的误差,发现必然性的趋势,认识规律的表现形式。

统计研究的对象,不论是自然现象还是社会现象,都具有随机性。这些现象的出现取决于多种多样的原因,既有必然的因素,也有偶然的因素。它们对于各具体单位所起的作用,在程度上,甚至在方向上都可能不同,这就导致同一现象在每个单位的数量表现具有随机性。统计研究运用大量观察法对总体中全部或足够多数的单位进行调查,并运用综合指标法对各单位变量加以综合,通过平均化的结果来抵消偶然因素的个别影响,描述现象的一般水平,进而推断总体的数量特征。这些都是在大量定律的作用下进行的。

从以上内容看出,大数定律的作用是帮助人们通过偶然性达到发现必然性,认识现象规律的表现形式,但它并不能说明现象的本质,须借助相关

性学科知识来解释现象的本质及其内在联系。

(二) 统计研究基本方法

统计研究过程的各个阶段,从统计资料搜集、数据的整理汇总,以至统计分析检验,都运用着各种专门的方法。在资料搜集阶段,对不能用科学实验取得资料的情况,如社会经济现象,则采用大量观察法;对可以用科学实验获取的资料,则可采用实验设计法。在数据整理汇总阶段,运用统计描述法,而统计分析与检验主要运用统计推断法,等等。

(1)大量观察法。大量观察法是对所研究对象总体的全部或足够多单位进行调查观察,并加以综合研究的方法。许多统计对象,特别是社会经济中的现象,不论是已经发生还是未发生的事件,都不可能进行重复实验去获取资料;其次,复杂的社会经济现象是在诸多因素错综作用下形成的,个别现象往往受各种偶然因素的影响,使各单位的特征和数量表现有较大差别,所以不能任意抽取个别或少数单位进行观察。必须在对所研究对象的定性分析的基础上,确定调查对象和总体范围,并对总体中的所有单位或足够多数单位的变量进行登记,然后把调查观察得来的各单位资料加以整理汇总,计算相应的综合指标来反映总体现象的数量特征。统计调查中的普查、抽样调查、统计报表调查、重点调查等,则是大量观察的组织形式。

(2)实验设计法。实验设计法是指通过合理的实验设计,经过反复实验获得符合统计分析要求的数据的方法,该方法主要运用于自然科学研究和工程技术领域的统计数据搜集。

(3)统计描述法。统计描述法又称为综合分析法,是指对调查观察或实验得到的数据进行登记、审核、整理、归类,计算出各种能反映总体数量特征的综合指标,并通过分析抽出有用的信息,用表格或图形将其显示出来的方法。统计描述是统计研究的基础,它为统计推断、统计咨询、统计决策提供必要的事实依据。统计描述也是对客观现象的认识过程,它通过对分散无序的原始资料的归纳整理,运用统计分组法、综合指标法得到客观现象的总体的数量特征,揭示其内在数量规律性。

(4)统计推断法。在现实中常常会遇到这样的情况,我们只能获取部分单位的数据或有限个单位的数据,而我们所关心的是整个总体甚至是无限总体的数量特征。这时就产生了根据局部的资料对总体数量特征作估计判断及估计判断的置信度问题。例如,根据某城市 1 000 户居民的生活支出数据来估计判断该城市居民的消费水平;又如,某种新药对某种疾病有多大程度的治疗作用,等等。这类问题如果只靠部分数据的描述是无法了解总体特征的,但利用统计推断的方法则可解决此类问题。所谓统计推断,是指

在一定的置信标准要求下,根据部分数据(样本)来估计判断总体数量特征的归纳推理的方法。统计推断是逻辑归纳法在统计推理中的应用,故又称为归纳推断法。统计推断可以用于总体数量特征的估计,也可以用于对总体某些假设的检验,所以又有不同的推断方法。在实际工作中,最常用的统计推断方法有参数估计法、假设检验法、相关与回归分析法等,这些将在以后的章节中详细介绍。

三、统计研究的作用

随着社会主义市场经济体制的逐步建立和完善,统计已由单纯的统计信息搜集整理职能转变为信息、咨询、监督三大智能。统计部门已经成为社会经济信息的主体部门和国民经济核算的中心,成为国家重要的咨询和监督机构。

信息智能是指系统的搜集、整理、存储和提供大量的以数量描述为基本的社会经济信息资源。咨询智能是利用已掌握的丰富的信息资源,运用科学方法进行综合分析,为科学决策和管理提供情况和咨询建议。监督智能是利用统计信息对社会经济的运行状态进行定量检查、检测和预警,揭示社会经济运行中出现的偏差,提出矫正意见,预警可能出现的问题,提出对策,以促使社会经济持续、健康发展。

信息、咨询、监督三大智能是相互作用、相辅相成,共同构成了统计的立体功能。其中,信息功能是最基本的,咨询、监督功能是统计信息功能的延续。发挥统计整体功能是国家长期统计工作,特别是统计事件经验的总结,也是国家科学管理和宏观调控的客观需要。

统计的作用主要体现在信息、咨询、监督三大功能上。具体表现为:①为党和政府各级领导机构决策和宏观调控提供资料;②为企业事业单位经营管理提供依据;③为社会公众了解情况,参与社会经济活动提供资料;④为科学研究提供资料;⑤为国际交往提供资料。

第三节 变量与数据的分类

为了提高数据分析的效率,我们需要对统计变量和统计数据进行分类或分组处理。对不同类型的数据,需要采用不同的方法分析和研究。变量或数据分类的方式多种多样,根据不同的需要,基于不同的视角,可以对变量和数据进行不同的分类或分组。

一、变量与数据的基本分类

按照研究者所采用的计量尺度不同,统计变量可以分为分类变量、顺序变量与数值型变量三种,与上述变量相对应的取值分别称为分类数据、顺序数据与数值型数据。

(一)分类变量与分类数据

分类变量(Categorical Variable)是说明事物类别的名称或概念,其取值称为分类数据(Categorical Data)。分类数据是只能归于某一类别的非数字型数据,它是对事物进行分类的结果,数据表现为类别,一般用文字表述。

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、旅游业等,这些变量的取值属于分类数据。

(二)顺序变量与顺序数据

顺序变量(Rank Variable)是说明事物有序类别的名称或概念,其取值称为顺序数据(Rank Data)。顺序变量其本质属于分类变量,但其类别呈现出有序状态。

例如,“产品等级”就是一个顺序变量,其变量值可以为一等品、二等品、三等品、次品等;“受教育程度”也是一个顺序变量,其变量值可以为小学、初中、高中、大学等;一个人对某种事物的看法也是一个顺序变量,其变量值可以取同意、中立、反对等值。

如上所述,分类变量和顺序变量说明的是事物的品质特征,通常是用文字来表述的,因而也可将他们统称为定性变量或品质变量(Qualitative Variable)。相应地,这类变量的取值统称为定性数据或品质数据(Qualitative Data)。

(三)数值型变量与数值型数据

数值型变量(Metric Variable)是按照数字尺度说明事物数字特征的一个名称,其取值称为数值型数据(Metric Data)。数值型数据就是按数字尺度测量的观察值,现实中所处理的大多数的数据都是数值型数据。

例如,产品产量、商品销售额、零件尺寸、年龄等都是数值型变量,这些变量所取得的不同的数值就是数值型数据。

数值型数据也称为定量数据(Quantitative Data)。