



北京高等教育精品教材
BEIJING GAODENG JIAOYU JINGPIN JIAOCAI



经济与管理专业研究生及高年级本科生通选教材

JINGJI YU GUANLI ZHUANYE
YANJIUSHENG JI GAONIANJI BENKESHENG
TONGXUAN JIAOCAI

应用统计学

YINGYONG TONGJIXUE

马立平 刘 娟 / 编 著

(第二版)



首都经济贸易大学出版社
Capital University of Economics and Business Press



北京高等教育精品教材
BEIJING GAODENG JIAOYU JINGPIN JIAOCAI



经济与管理专业研究生及高年级本科生通选教材

JINGJI YU GUANLI ZHUANYE
YANJIUSHENG JI GAONIANJI BENKESHENG
TONGXUAN JIAOCAI

应用统计学

YINGYONG TONGJIXUE

马立平 刘 娟 / 编 著

(第二版)



首都经济贸易大学出版社
Capital University of Economics and Business Press

· 北 京 ·

图书在版编目(CIP)数据

应用统计学/马立平,刘娟编著.—2版.—北京:首都经济贸易大学出版社,2015.6

(经济与管理专业研究生及高年级本科生通选教材)

ISBN 978-7-5638-2364-2

I. ①应… II. ①马… ②刘… 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(2015)第 103696 号

应用统计学(第二版)

马立平 刘娟 编著

出版发行 首都经济贸易大学出版社

地 址 北京市朝阳区红庙(邮编 100026)

电 话 (010)65976483 65065761 65071505(传真)

网 址 <http://www.sjmcb.com>

E-mail publish@cueb.edu.cn

经 销 全国新华书店

照 排 首都经济贸易大学出版社激光照排服务部

印 刷 北京地泰德印刷有限责任公司

开 本 710 毫米×1000 毫米 1/16

字 数 382 千字

印 张 21.5

版 次 2011 年 10 月第 1 版 2015 年 6 月第 2 版

2015 年 6 月总第 3 次印刷

印 数 3 001~6 000

书 号 ISBN 978-7-5638-2364-2/C·121

定 价 35.00 元

图书印装若有质量问题,本社负责调换

版权所有 侵权必究

前言

统计学是收集、分析、表述和解释数据的应用性很强的学科。通过数据资料对现象进行数量方面的分析,不仅能够更客观地认识研究对象,也能使对问题的认识与研究更加深入。统计方法的应用在某种程度上左右、影响着实际工作与理论研究的水平。统计方法既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,它在科学研究与社会经济管理中能够发挥重要的作用。相信读者在学习了统计学并将其在实际中应用之后,能够更深刻地感受到这一点。

本书是作者根据多年的本科及研究生统计学教学工作的体会编写出来的,目的是让读者掌握统计学的基本理论与方法,并能够使用统计学的基本方法进行初步的统计分析与研究。撰写时我们特别注意突出以下几方面的特点:

第一,本书主要是针对高年级本科生和非统计学专业研究生,为使初学统计学的学生尽快了解统计研究的思维方式,掌握统计分析的基本方法并加以应用,编写时注重用通俗易懂、深入浅出的语言阐述统计学的基本思想与方法,同时力求强调统计思想,略掉了一些不必要的数学证明和公式推导。书中在对各种具体方法作必要的阐述之后,都配备有具体的案例或例子说明其基本思想与方法的应用,并用图表等比较直观的形式进行解释。

第二,本书的内容完整,整体框架清晰,逻辑性较强,突出了实际问题分析与统计研究中的工作顺序与知识的逻辑关系。从内容上看,不仅包括描述统计,也包括推断统计;从分析方法来看,不仅包括传统意义上的统计学原理的基本内容,也包括实际分析研究中常用的多元统计分析的方法与应用,力求使读者较为系统、完整地掌握统计方法论的总体框架,从而能够在学习统计学的基本理论、思想与方法的同时,把握何时、用何种方法、对何种问题进行统计分析与研究。

第三,统计学是通过对大量数据的分析与研究发现问题、得出结论的方法论的科学,具有很强的应用价值。为了避免读者学完之后只掌握统计学基本思想与方法而无法实际操作,本书的编写力求具有较强的实用性和可操作性。在介绍方法的基础上,结合统计软件,

全面、系统地介绍了统计计算与分析的过程及其技术的实现。同时，针对具体问题进行分析，并对计算结果进行解读，以提高学习者的实际操作能力与水平。

本书由马立平、刘娟编著，姜月香参加了第三、四、六章的编写。本书的编写建立在我们多年本科及研究生统计学教学经验与体会的基础上，同时也参考和借鉴了许多统计学著作和教材，并得到了首都经济贸易大学出版社编辑兰士斌的大力支持和帮助，在此一并表示感谢。

尽管编写此书时我们投入了许多时间和精力，但书中难免存在一些不尽如人意之处，真诚欢迎广大读者提出宝贵意见，我们在此先行表示感谢。

第一章 统计学与统计数据	1
第一节 统计学的产生与发展	1
第二节 统计学的性质与特点	5
第三节 统计数据与统计规律	7
第四节 经济管理实践中的统计 ——统计应用的领域之一	12
第二章 统计数据收集的方法与数据质量	15
第一节 统计数据的来源	15
第二节 数据收集的方法	18
第三节 问卷的设计	21
第四节 统计数据的质量	29
第三章 数据的图表展示	35
第一节 定性数据的图表展示	35
第二节 定量数据的图表展示	42
第三节 统计图表应用中的几个问题	53
第四章 数据的统计量描述	60
第一节 数据集中趋势的测度	60
第二节 数据离散程度的测度	69
第三节 数据分布形状的度量	73
第四节 描述数据特征的统计量的计算与应用	75
第五章 概率抽样与抽样分布	80
第一节 随机变量的概率分布	80
第二节 概率抽样方法	91
第三节 抽样分布与中心极限定理	97
第四节 常用的抽样分布	99
第五节 几个重要的小样本抽样分布	102

第六章 参数估计 105

第一节 参数估计的基本问题	105
第二节 一个总体参数的区间估计	110
第三节 两个总体参数的区间估计	115
第四节 样本容量的确定	119

第七章 参数的假设检验 125

第一节 假设检验的基本问题	125
第二节 一个总体参数的假设检验	130
第三节 两个总体参数的假设检验	138

第八章 方差分析 150

第一节 方差分析的原理	150
第二节 单因子方差分析	153
第三节 多因子方差分析	157

第九章 非参数的假设检验 164

第一节 非参数检验的主要内容	164
第二节 单样本非参数检验	166
第三节 独立样本的非参数检验	172
第四节 相关样本的非参数检验	180

第十章 相关分析与回归分析 188

第一节 相关分析	189
第二节 线性回归分析	198
第三节 可线性化的非线性回归	209
第四节 回归分析应用中的几个问题	214

第十一章	主成分与因子分析	230
第一节	主成分分析	230
第二节	因子分析	234
第三节	综合案例——我国工业企业经济效益评价	244
第十二章	列联分析与对应分析	252
第一节	列联分析	252
第二节	对应分析	258
第三节	综合案例——失业原因与教育程度关系的 研究	264
第十三章	聚类分析和判别分析	273
第一节	聚类分析	273
第二节	判别分析	286
第十四章	时间数列分析与预测	295
第一节	时间数列的动态分析指标	295
第二节	时间数列的构成要素	302
第三节	时间数列的长期趋势分析	307
第四节	包含季节变动的时间数列预测	319
附录		325
参考文献		335

第一章 统计学与统计数据

第一节 统计学的产生与发展

统计活动源远流长,其历史可追溯到远古的原始社会。可以说,自从人类社会有了数的概念,有了计数活动,也就有了统计。但将统计实践上升到理论,并将其概括、总结成为一门科学的统计学,却是近代的事。虽然对于统计学产生于什么年代,人们的看法不尽一致,但多数人认为,统计学大概兴起于 17 世纪,距今只有 300 多年的历史。

一、统计学的产生

17 世纪中叶,西方社会首次出现了有意识地用数字语言说明问题,从数量的角度探索客观事物变化规律的研究活动。当时,最具有代表性的主要有政治算术学派和国势学派。

政治算术学派的代表人物是英国的学者威廉·配第(W. Petty)。17 世纪,威廉·配第在他的著作《政治算术》一书中,对当时英国、荷兰、法国之间的“国富和力量”进行了数量上的比较,首次从数量方面明确地用大量的数据资料分析问题,揭示了一些经济学的科学原理,研究了许多经济范畴的经济关系。也正是在这个意义上,马克思称威廉·配第是政治经济学之父,是统计学的创始人。在统计史的研究中,一般把以威廉·配第为代表的关于社会经济现象“算术”式的研究,称为“政治算术”。

除威廉·配第外,政治算术学派还有一位重要的人物——约翰·格朗特(J. Graunt)。他对伦敦市 50 多年的人口出生和死亡的情况进行了计算与研究,1662 年写出了其代表性的著作——《关于死亡表的自然观察与政治观察》,该书通过对人口变动数据的分析,揭示了一系列人口变化的规律。

自此,人们说明国家重要事项除了用记述的方法外,开始逐渐使用量的方法,这为统计学作为一种从数量方面认识事物的方法开辟了广阔的研究与应用前景。政治算术学派第一次有意识地运用可度量的方法,把自己的论证建立在具体的数字基础上,在统计学发展史上具有重要的地位。但是,由于其毕竟还处于统计核算的初始阶段,它只能用最简单

的算术方法对社会经济现象进行计量和比较。

与政治算术产生的时期差不多,在17世纪的德国,康令(H. Conring)、阿肯瓦尔(G. Achenwall)等人在大学中开设了一门叫做“国势学”的课程。“国势学”主要研究有关国家兴衰强弱的重大问题以及治理国家必备的方法。作为“国势学”的奠基人,康令对“国势学”的研究目的、研究对象、研究方法等基本问题作了深入细致的阐述,其方法主要是文字记述和逻辑比较。

经过100多年的发展,到18世纪中期,“国势学”的发展达到了顶峰。

统计学的另一个重要起源是概率论。概率论起源于赌博游戏,但真正意义上的概率论是从17世纪开始的。在早期从事概率论研究的众多学者中,拉普拉斯是古典概率论的集大成者。他给出了概率数学的古典解释,建立了严密的概率数学体系。“政治算术”研究的是简单的、确定的数量关系,而概率统计研究的则是复杂的、随机性的现象。概率论的出现,极大地充实和深化了数量问题研究的内容。以概率论为基础,统计学进入了一个新的发展时期。

二、统计学的发展

凯特勒(A. Quetelet)是统计学发展史上承前启后的重要人物。凯特勒于18世纪在统计学上作出的突出贡献是把概率论全面引进“政治算术”“国势学”以及其他社会问题的研究,大大推动了概率论和数学方法在社会科学领域的应用,促进了数量研究由“算术”阶段向“数理”阶段的迅速转化。

19世纪后半期,统计学在生物遗传学、农业田间试验等领域取得了创新性的成果。如生物统计学的主创者高尔顿(F. Galton)利用正态法则研究优生学、遗传学等问题,先后提出了“百分位数”“中位数”“四分位数差”“相关与回归”等概念及计算方法。此后,皮尔逊(K. Pearson)系统地发展了高尔顿的相关与回归理论,研究了复相关和偏相关、极大似然估计方法,导出了卡方分布等。以皮尔逊为代表的统计学家在大量观察和正态分布的基础上,进行了总体分布曲线的研究,确立了大样本统计理论,从而构建了描述统计学的框架体系。

进入20世纪,新的统计思想和统计方法大量涌现,带有归纳性质的统计推断逐渐占据了主流地位。自苏歇米尔斯(J. Sussemilch)提出大数法则到20世纪初的一段时期里,大量观察法一直是统计思想的核心,直到1908年戈塞特(W. Gosset)导出了重要的 t 分布,统计学逐渐实现了由描述统计阶段向推断统计阶段、由大样本统计向小样本统计理论的转

变。此后,费希尔(R. Fisher)开辟了方差分析、试验设计等统计分支,论证了相关系数的抽样分布,提出了 t 检验、 F 检验、相关系数检验等,这些研究成果在统计学发展史上有着很高的地位。此后,内曼(J. Neyman)和皮尔逊共同完善了现代统计学的核心内容——区间估计和假设检验的理论,瓦尔德(A. Wald)则提出了统计决策理论和质量检验的“序贯分析”,进一步推动了统计学研究和应用的范围。20世纪五六十年代以后,稳健统计、时间序列、抽样理论、统计诊断、探索性分析、贝叶斯统计等也取得了重要的进展。

总体上说,20世纪以来,统计学的发展有三个明显的趋势:首先,随着数学的发展,统计学依赖和吸收数学方法的程度越来越深;其次,统计学方法向其他学科领域广泛渗透,以统计学为基础的边缘学科不断形成;最后,随着统计学应用的日益广泛和深入,特别是随着计算机的发展和广泛使用,统计学正发挥着越来越大的作用。

三、统计学科体系

目前,统计学已经形成了由理论统计学、应用统计学、统计学史等若干个分支组成的完整的学科体系。这里着重介绍理论统计学和应用统计学两个分支。

(一)理论统计学

理论统计学侧重于从数学学科中吸取营养,是研究统计学的方法和基础理论,以解决统计学自身发展过程中的重大问题为目标的统计学分支。它主要包括以下几方面内容。

1. 统计估计。统计估计是统计学的核心内容之一,它包括两个方面的内容:一是在总体分布已知时,对总体未知参数或参数组合的函数进行估计;二是在总体分布未知时,对有关分布的特征数字及分布密度进行估计。其研究的重点内容包括估计量的确定和对估计量的评价等。

2. 假设检验。假设检验是指在总体分布已知的情况下,根据样本资料,对总体参数的某种假设命题进行检验和判断。其研究的重点在于检验统计量的确定、假设检验的原理和检验效率等问题。

3. 抽样调查。抽样调查是收集统计资料的基本手段与方法之一,其研究的重点是抽样方案的设计、样本的抽样方法、抽样分布、抽样效果与抽样误差等问题。

4. 试验设计。试验设计主要是研究如何安排试验方案,如何分析试验结果等问题。

5. 非参数统计。非参数统计主要研究总体分布未知或不依赖于总

体分布的各种统计问题。

6. 时间序列。时间序列是指按时间顺序排列的一组数据,时间序列方面的主要研究内容包括时间序列的基本结构、时间序列的分解、自回归过程与参数估计、非线性系统模型和空间序列分析等。

7. 统计决策。统计分析与研究的目的是认识客观规律、进行科学决策。统计决策的主要内容包括风险函数、损失函数、决策标准和决策函数等。

8. 序贯分析。序贯分析是指在得出分析结论之前,视具体的观察结果来决定决策方案的选择,具体内容包括抽样方案、序列检验统计量、风险判别等。

9. 多元统计。多元统计又称为多元统计分析,是关于多维随机变量的统计研究方法,主要包括多元分布、判别分析、典型相关分析、主成分分析、因子分析、聚类分析等。

10. 统计诊断。统计诊断主要研究观察数据、统计模型、统计推断方法的合理性问题,并对诊断中发现的缺陷进行治理和改进,是近二三十年来发展起来的一个新的统计学领域。

11. 稳健统计。稳健统计主要研究当理论分布与实际分布不一致时,如何确定不敏感的统计方法等内容。

12. 探索性分析。探索性分析主要通过观察数据进行详细考察,挖掘数据本身的结构和特征,然后在此基础上建立分析模型。

(二)应用统计学

应用统计学就是运用统计学的思想和方法,发现、处理实践中的问题。其内容大体上可以分为以下几类。

1. 统计计算方法。统计计算方法把统计方法、数学计算方法和计算机应用结合起来,重点解决数据处理过程中遇到的各类计算问题。

2. 应用统计学的理论基础。应用统计学是以理论统计学为理论基础,从应用的角度进行的统计学研究。同理论统计学相比,应用统计学一般带有较强的应用背景,它用简明扼要的语言,深入浅出地阐述统计学的思想与方法。

3. 应用统计学。应用统计学从实际问题的背景出发,与具体学科紧密结合,重点研究如何使用统计方法。应用统计学按其应用的学科性质的不同,可分为理工科的应用统计学(如统计力学、生物统计学、医学统计学、气象统计学、地理统计学等)和社会科学类的应用统计学(如人口统计学、经济统计学、管理统计学、教育统计学、社会统计学等)。

4. 其他。统计学与其他应用数学学科结合,形成新的应用数学方法的基础学科,如博弈论、多目标决策、随机规划等。

第二节 统计学的性质与特点

一、什么是统计学

统计学是随着人类社会的发展和社会管理的需要而发展起来的。随着统计方法在各个领域的应用,统计学已发展成为具有多个分支学科的大家族。因此,对统计学的定义,人们有着不同的表述。其中,具有代表性的、被人们普遍接受的是《不列颠百科全书》的定义:“统计学是收集、分析、表述和解释数据的科学。”

统计学的定义告诉我们,统计学研究的是客观现象的数量方面,统计离不开数据。因此,进行统计应用研究,首先要收集能够反映或说明客观现象的数据资料,这是统计活动首要的、基本的环节。没有统计数据,统计方法就失去了用武之地;而如何取得可靠的、高质量的统计数据,则是统计学研究的内容之一。

分析数据是统计学的核心内容,它是对已有的数据资料,通过统计方法探索数据内在的规律、提取有关综合信息的过程,其目的是形成一个概括的、全面的数量描述。分析数据所用的方法可分为描述性统计方法和推断统计方法。

当然,分析数据之前,需要对统计数据进行处理和整理,即进行数据整理,目的是使统计数据系统化、条理化,以符合统计分析的要求。统计整理在统计活动中处于承上启下的位置:一方面,它是收集统计资料工作的继续;另一方面,它又是统计分析的前提,是统计工作的必要环节。数据整理的内容包括数据的甄别与筛选、数据的审核与修正、数据的分组与分类、编制频数分布表、用统计图及指标展示统计数据,目的是找出数据的初步特征,或者是方便他人看懂数据所要表达的问题。

分析数据之后,即形成了对数据的解释。数据的解释是对分析的结果进行的说明和进一步分析,阐明分析结果所隐含的事物的特征以及从数据中所得出的规律性的结论等。

二、统计学的研究方法

统计学的研究方法是指统计学研究和认识客观事物总体数量方面的各种方法。从研究主体的活动角度来说,与其他诸多学科需要经历资料的收集、整理和分析的过程一样,统计学的研究也要经历资料的收集、整理和分析的过程,这在统计学中称为统计调查、统计整理和统计分析。在这一过程的各个阶段,统计学运用了多种专门方法,从总体上看,其基

本方法有大量观察法、综合指标法和归纳推断法。

(一)大量观察法

大量观察法是指在客观事物的研究中,从总体出发,对其全部单位或足够多的单位进行观察和分析研究的方法。

大量观察法之所以是统计学研究的基本方法,是由统计学的研究对象及研究目的决定的。统计学研究的事物总体是由大量单位、大量数据构成的。研究对象的数量方面受到诸多因素的影响,这些因素可大致分为两类:一类是由研究对象的基本性质以及一般条件所形成的共同性因素,这类因素对所有个体单位都发生作用,是研究对象总体数量规律性存在的根据;另一类是源于研究对象的次要性质或偶然因素、随机因素,这类因素使得各个单位在数量表现上存在差异,各不相同,掩盖了研究对象的规律性。大量观察法的意义就在于通过把全部单位或足够多的单位联系在一起,使得个别性因素的作用在这种联系中相互抵消,突出共同性因素的作用,从而显示出总体相对稳定的数量特征和数量关系,即数量规律性。

(二)综合指标法

综合指标法是表明事物总体数量特征的数据,它是统计学研究事物总体数量方面的基本手段或工具。综合指标按其一般表现形式分为总量指标、相对指标和平均指标三大类。其中,总量指标的基本来源是对原始数据的整理汇总,以它为基础,可利用多种方法计算出各种派生的相对指标和平均指标。

在统计学的研究中,综合指标有着重要的分析意义。它概括地表明了研究对象的规模、总量、联系程度和一般水平,在结构、比例和相互关系分析、发展变化分析、因素分析以及其他许多方面的分析中都起着重要的作用。可以说,统计分析就是对统计指标及指标关系的分析,也就是对数量特征、数量关系、数量界限及数量规律性的分析。

(三)归纳推断法

归纳推断法是指以一定的置信标准,根据样本数据来判断总体数量特征的归纳方法。它是现代统计的基本方法。在实际应用中,由于种种主客观方面的原因,当统计学所研究的对象的范围大于实际可能掌握的范围时,就需要应用统计推断法。这种情况十分普遍,尤其是当统计学对具有普遍意义的问题感兴趣的时候。研究的问题愈具有普遍性,统计学研究对象的范围也就愈大。

归纳推断的一个重要特点是它不能对问题作出绝对肯定的结论,只能在一定的可靠程度保证下,作出能满足研究精度的弹性结论。

三、统计学的性质与研究对象

由统计学的发展历史可知,统计学是从研究社会经济现象的数量特征开始的。随着统计方法的不断完善和应用领域的不断扩展,统计学得以不断发展。

总体上看,统计学的研究对象与性质主要表现在以下几个方面。

第一,统计学是一门关于数据的科学,统计研究的是客观事物的数量特征,包括数量状态、数量关系和数量变化规律,统计的基本语言是数据。无论是社会科学还是自然科学,只要有出现大量数据的地方,就需要统计。从认识论的角度看,任何事物都是数量和质量的统一体。如果数据资料准确、可靠,统计分析方法运用得当,那么,通过研究客观事物的数量方面,就可以正确地认识客观事物的特征与发展变化的规律。

第二,统计学对大量同类现象的数量方面进行综合反映与研究,也就是说,统计学的研究对象是客观现象总体的数量,单个数据的分析与研究不是统计学的研究对象。只有通过大量的现象或对某一现象进行多次重复的观察,才有可能找到统计关系和统计规律。当然,按照认识论由个体数量到总体数量这样的认识逻辑,统计研究的对象尽管是总体数量,也必须从个体数量的调查入手。

第三,统计学的研究对象是不确定的现象,即随机性的变量。不确定性现象的存在是由于受到偶然的、随机因素的作用,使得客观事物的实际数量表现存在一定程度的不可确知性。现实中,太多的现象都是不确定的现象,如人的寿命是一个随机变量,这种随机性可能和人的遗传基因、生活习惯、生活质量等很多因素都有关系。虽然一个人的寿命有一定的随机性,但从总体上看,我国公民的预期寿命非常稳定,且女性的预期寿命高于男性的预期寿命,这就是在随机性之中的规律性。

第四,统计的基本方法是归纳推断。统计对总体的认识有两种途径,一是通过掌握构成总体的全部事物的数据资料进行认识,这时运用算术方法和统计描述手段就可以达到认识总体的目的;二是从总体中抽取部分事物组成样本,然后依据样本的数据,对总体进行推断,其基本方法就是归纳推断,而不是逻辑推理。从应用的经济性、时效性、实用性和可行性等方面考虑,利用样本资料对总体进行归纳推断,其优势比较明显。

第三节 统计数据与统计规律

统计数据是对现象进行观测、计量的结果,如对经济活动总量的计量可以得到国内生产总值的数据,对人口性别的计量可以得到男女性别

比的数据,对城镇居民收入与消费的调查可以得到城镇居民的工资水平和消费支出的数据,等等。

一、数据的计量尺度

对不同的事物进行计量或测度,其精确程度是不同的。按照对事物计量的精确程度划分,可将所采用的计量尺度由粗略到精确分为以下几个不同的层次。

(一)定类尺度

定类尺度是最粗略、计量层次最低的计量尺度。它的主要特征是采用文字、数字代码和其他符号对事物进行简单的分类或分组。建立在对事物进行分类分组的基础上的计量尺度称为定类尺度。如对人口性别、民族、籍贯、婚姻状况的统计测量,对企业经济性质的划分等,这些用作归类统计的测量均属于定类尺度。定类尺度只是测度事物之间的类别差,各类别之间无法区分优劣或大小。

由于定类尺度只能区分事物是同类或非同类,因而使用定类尺度计量的数据不能直接进行数学运算。当然,在进行数据处理时,为了便于计算机识别和信息的传输,对于定类尺度计量的数据,我们往往给每一个类别赋予一个数字代码,如在人口性别中用“1”表示男,用“0”表示女。然而需要注意的是,这些数字只是给不同类别赋予一个代码,并不意味着这些数字可以进行任何数学运算。

在使用定类尺度对事物进行分类时,必须符合穷尽和互斥的要求,即在进行分类时,必须保证所有个体都能归属于某一类别,不能有遗漏;而且只能归属于一个类别,不能在其他类别中重复出现。

(二)定序尺度

定序尺度又称顺序尺度,它是对事物之间的等级差或顺序差进行的一种测度。该尺度不仅可以将事物分成不同的类别,而且还可以确定这些类别的优劣和顺序。如调查了解城镇居民对公众服务的满意度,满意程度可分为很满意、比较满意、一般、不太满意和不满意,被调查者的学历可分为小学及以下、中学、高中、大专、大本、研究生及以上。再如产品按其品质划分,可分成合格品和不合格品,等等。这些都属于定序尺度。定序尺度的计量结果虽然也表现为类别,但这些类别之间可以比较并排序,其对事物的计量要比定类尺度精确一些。但是,它也只是测度了类别之间的顺序,不能测量出类别之间的准确差值。该尺度可以分类,也可以比较大小、好坏,但不能进行加、减、乘、除等数学运算。

(三) 定距尺度

定距尺度也称间隔尺度,它不仅具有定序尺度的所有特征,即能将事物区分为不同类型并进行排序,而且可以准确地指出类别之间的差距是多少。定距尺度是对事物类别或次序之间间距的度量,通常使用自然或度量衡单位作为计量尺度,其计量结果表现为数值。如考试成绩用百分制度量、温度用摄氏度或华氏度来度量等。由于这种尺度的每一间隔都是相等的,只要给出一个度量单位,就可以准确地指出两个计数之间的差值,因而其计量结果不仅可以进行分类和比较,而且可以进行加、减的数学运算。

(四) 定比尺度

定比尺度是最高级别的测量尺度,其计量结果和定距尺度一样也表示为数值。但是与定距尺度不同的是,定比尺度拥有一个绝对零点的测量原点,而定距尺度却没有这样的原点。因此,定距尺度只能进行加、减运算,而定比尺度不仅可以进行加、减运算,还可以进行乘、除运算。如工资是定比尺度,其绝对零点是零元,即没有。若一个人的月工资是6 000元,另一个人的月工资是3 000元,可以得出前者工资比后者工资多3 000元,是后者的两倍。而温度是定距尺度,其不存在绝对零点;即使温度为零,也不能说没有温度。我们可以说30度比15度高15度,但不能说30度比15度热一倍。

采用不同的计量尺度,可以得到不同类型的统计数据,而不同类型的统计数据又适用于不同的统计分析方法。定类、定序、定距、定比四种计量尺度对事物的测量层次是由低级到高级逐步递进的,高层次的计量尺度具有低层次的计量尺度的特征,我们可以将高层次的计量尺度的测量结果转化为低层次的计量尺度的测量结果,如将温度的定距尺度测量结果转化为很冷、较冷、舒适、较热、很热的定序尺度的测量结果。四种计量尺度的数学特征是:①定比尺度可以分类、排序、测量距离和计算比值;②定距尺度可以分类、排序、测量距离,不能计算比值;③定序尺度可以分类、排序,不能测量距离、计算比值;④定类尺度只能进行分类。

二、数据的类型

所谓数据,是对现象进行计量的结果,如对气温的测量、对经济活动总量的计量等。我们可以从不同的角度对统计数据进行分类。

(一) 分类数据、顺序数据和数值型数据

按照所采用的计量尺度划分,统计数据可以分为分类数据、顺序数据和数值型数据。

分类数据是对事物进行分类的结果,它描述事物的品质特征。例