



普通高等教育“十五”国家级规划教材

应用统计

主 编 朱洪文 副主编 宋 力
主 审 王维国



高等教育出版社
HIGHER EDUCATION PRESS

普通高等教育“十五”国家级规划教材

应用统计

主 编 朱 洪 文

副主编 宋 力

主 审 王 维 国

高等教育出版社

内容提要

本书是普通高等教育“十五”国家级规划教材,是根据“统计学”核心课程的基本要求,并结合工科、经济管理类学校的特点编写而成。

本书内容共十章:绪论、统计资料的收集与整理、统计资料的描述、推断理论基础、参数估计、假设检验、 χ^2 检验和方差分析、相关与回归分析、时间数列分析、指数。

本书的最大特色:一是应用性强,基于问题提出统计方法的使用,便于使用者真正掌握统计方法及其应用;二是通俗易懂,全书在不失严谨的前提下尽量避免数学公式的推导,主要通过案例介绍统计方法的使用,使学生掌握统计方法应用的背景要求和前提条件;三是本书在统计计算上强调运用统计软件来完成,为统计方法在实践中能够广泛使用奠定基础。

本书在体系上参阅了大量的国内外优秀教材,是笔者多年教学经验的总结,可作为经济管理类专业本科生、研究生的教材,也可供从事工商管理分析和经济分析的各类人员参考。

图书在版编目(CIP)数据

应用统计/朱洪文主编. —北京:高等教育出版社,
2004.7 (2005 重印)
ISBN 7-04-014386-0

I. 应… II. 朱… III. 应用统计学—高等学校—
教材 IV. C8

中国版本图书馆 CIP 数据核字(2004)第 054102 号

策划编辑 马 丽 责任编辑 蒋 青 封面设计 王凌波 责任绘图 尹文军
版式设计 王 莹 责任校对 杨雪莲 责任印制 孔 源

出版发行 高等教育出版社
社 址 北京市西城区德外大街 4 号
邮政编码 100011
总 机 010-58581000
经 销 北京蓝色畅想图书发行有限公司
印 刷 化学工业出版社印刷厂

购书热线 010-58581118
免费咨询 800-810-0598
网 址 <http://www.hep.edu.cn>
<http://www.hep.com.cn>
网上订购 <http://www.landaco.com>
<http://www.landaco.com.cn>

开 本 787×960 1/16
印 张 22.25
字 数 410 000

版 次 2004 年 7 月第 1 版
印 次 2005 年 2 月第 2 次印刷
定 价 25.60 元

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换。

版权所有 侵权必究

物料号 14386-00

前 言

本书是根据“统计学”核心课程的基本要求,并结合工科、经济管理类院校的特点编写的。随着我国政府和管理水平的日益提高,人们对数量分析提出了更高的要求。统计学作为一门定量分析的方法论科学得到了前所未有的重视。统计方法已成为理、工、农、医、人文、管理及军事等领域进行科学研究和高效管理的基本工具。与此相适应,高等院校的统计教育也出现了可喜的变化。尤其是在教材建设方面,出现了许多优秀的创新教材及国外优秀教材,这些教材为提高我国统计教育水平发挥了重要作用。但是,有些教材在强调体系完整的同时,所包含的内容对于有限的学时来说却显过多,这在某种条件上造成了教材普及的困难。在这种背景下,编写一本体系基本完整、内容适合实际需要的统计教材是十分必要的。

本书编写的指导思想是:一要通俗易懂,全书在不失严谨的前提下尽量避免数学公式的推导,主要使读者掌握统计方法应用的背景要求和前提条件;二是要以计算机为主要计算工具,本书选择 Excel 统计软件辅助进行复杂的计算;三是方法为主,书中主要侧重于介绍统计方法在经济、管理中的应用条件和统计思想,期望给读者如何使用统计方法的一个清晰概念。本书体系是,作者参阅了大量的国内外优秀教材的研究成果,并结合多年教学经验而确定的。全书以数据为线索,从数据的收集到整理及分析,将描述统计与推断统计有机的结合。每章都附有学习要点和本章小结,这有利于读者在本章学习之前对要点有所侧重,在学习之后能够进一步加深理解。这也是本书的一个特点。本书适合作为经济、管理类本科生、研究生的教材,也可供从事工商行政管理和经济分析的各类人员参考。

参加本书编写的人员有哈尔滨工业大学朱洪文(第一、四、五、六、七、八章)、王淑杰(第九章)、吴冲(第十章)、沈阳工业大学宋力(第二、三章)。全书由朱洪文任主编,负责大纲设计、书稿组织和最后编纂。宋力任副主编,负责本书的审订及出版工作。

在本书编写过程中,李一军教授为本书的编写提出了重要的指导性意见。东北财经大学的王维国教授认真审阅了全书,并提出了许多宝贵的修改意见。哈尔滨工业大学和高等教育出版社对于本书的编写也给予了大力支持。在本书付印之际,向他们以及所有帮助和支持本书编写及出版的朋友们表示衷心的感谢。

由于编者水平有限,书中难免有疏漏或错误之处,恳请专家和读者不吝赐教,以便今后修改和完善。

编 者
2004 年 3 月

目 录

第一章 绪论	1
第一节 统计学概述	1
第二节 基本概念	4
第二章 统计资料的收集与整理	14
第一节 统计资料的收集与审核	14
第二节 统计调查	17
第三节 调查方案的设计	25
第四节 资料整理的方法	32
第三章 统计资料的描述	46
第一节 总量描述与相对比较	46
第二节 集中趋势的描述	55
第三节 离散程度的描述	68
第四节 统计资料的图形描述	74
第四章 推断理论基础	89
第一节 概率	89
第二节 随机变量与概率分布	98
第三节 抽样分布	113
第五章 参数估计	123
第一节 参数估计的原理	123
第二节 总体参数的区间估计	128
第三节 样本容量的确定	138
第六章 假设检验	145
第一节 假设检验原理	145
第二节 一个总体均值的假设检验	150
第三节 两个总体均值之差的假设检验	155
第四节 总体成数与方差的假设检验	160
第七章 χ^2 检验和方差分析	168
第一节 拟合优度的 χ^2 检验	168
第二节 独立性和一致性的 χ^2 检验	176
第三节 单因素方差分析	182
第四节 两因素方差分析	188
第八章 相关与回归分析	200

第一节	相关与回归分析的基本概念	200
第二节	一元线性回归分析	203
第三节	一元线性回归方程的检验	213
第四节	多元线性回归分析	222
第五节	非线性回归	230
第九章	时间数列分析	244
第一节	时间数列分析的一般问题	244
第二节	时间数列的描述性分析指标	247
第三节	长期趋势分析	261
第四节	季节变动分析	273
第五节	循环变动和不规则变动	279
第六节	时间数列因素分析应用	280
第十章	指数	289
第一节	指数概述	289
第二节	综合指数	291
第三节	平均指数及其应用	295
第四节	指数体系与因素分析	298
第五节	指数的应用	306
附表 1	随机数表	314
附表 2	二项分布表	316
附表 3	泊松分布表	326
附表 4	标准正态分布表	332
附表 5	χ^2 分布表	334
附表 6	t 分布表	336
附表 7	F 分布表	338
参考文献		346

第一章 绪 论

学习要点

1. 统计学定义
2. 区别统计学的各种分类
3. 对基本概念能够准确理解
4. 对 Excel 软件的了解

第一节 统计学概述

一、什么是统计

学习统计学(statistics),首先就要理解什么是“统计”?有人认为“统计”就是政府的一项工作,也有人认为“统计”就意味着一些枯燥无味的数据,还有些人认为“统计”就是有着大量公式和计算的一门课程等等.也许你还可以列举许多种对于“统计”一词的理解.但不论多少种,归纳起来,“统计”一词主要包括三个方面的含义:统计活动;统计资料;统计学.

所谓**统计活动**(或称统计工作)是指各国政府或其他机构为满足政治、经济、社会等方面的需要以及科学研究的需要而进行的收集、整理、分析、编制有关数据的一系列活动.这些活动一般包括统计调查、统计整理、统计描述、统计分析和数据显示等过程.**统计资料**是指由统计活动产生的原始的或已经加工、整理过的客观现象的数据资料,它是统计活动的成果.**统计学**是指研究客观现象数量关系及其变化规律的方法论科学,是一门关于统计资料的收集、显示、描述和分析方法的学科.

“统计”一词实际上就是上述三种含义的简称.在不同的场合和语言环境下,可以分别替代三种不同的含义.例如,“据统计,在全省搞统计的人中,有30%的人没学过统计”.在这句话中就分别使用了“统计”的三种含义.

本书中,在不会引起歧义的情况下,用“统计”分别代替“统计工作”、“统计资料”和“统计学”三种含义.

二、统计学的分类

从 17 世纪中叶开始,统计逐渐形成系统的理论和方法,成为一门独立科学。其应用领域也在不断扩大,由政府统计的应用开始,已经扩展到包括自然科学和社会科学在内的许多领域。统计学在长期的发展过程中,由于在不同领域、不同行业所应用的统计方法不同,或者研究对象的差别等原因,统计科学又衍生出了若干个分支学科和交叉学科,可以从不同的角度进行类别划分。下面的两种划分只是粗略的、本书所涉及到的分类:一是分为描述统计学与推断统计学;一是分为理论统计学与应用统计学。

1. 描述统计学与推断统计学

描述统计学主要研究准确、真实地反映某种客观现象的数量方面的特征的理论及方法。它包括数据的收集、整理、数据的分析和显示等内容。例如,当我们要研究某地区 1996—2000 年的经济发展状况时,首先要确定使用哪些指标来说明经济发展状况,决定怎样去收集这些指标数据,采取哪种计算分析方法等。其次,还要考虑通过什么形式来描述该地区的经济状况。在上述过程中所使用的统计方法,就属于描述统计研究的范畴。应该说描述统计学所研究的内容是统计学的基础。

推断统计学主要研究通过部分数据特征推断全体数据特征的理论和方法。它包括对部分数据的抽取、对数据的分析、根据部分数据对全体数据作出推断以及对推断有效性的判断等内容。例如,当需要根据一部分齿轮的质量对全部齿轮质量进行判断时,就会用到推断统计;当根据部分地区人口的增长速度对全国人口规模作出估计时,就需要使用推断统计的方法。可以说,推断统计学是现代统计学的核心内容。

2. 理论统计学与应用统计学

理论统计学主要是指对统计学的一般理论和统计方法的数学理论的研究,是由一系列的公理、定理以及严格的证明来组成。它包括概率论、抽样理论、样本分布理论、实验设计、估计理论、假设检验、决策理论、非参数统计、序贯分析、多元分析、时间数列以及博弈理论等。理论统计学着重阐明统计学的数学原理,为统计方法提供理论基础。

应用统计学是指在把统计方法应用于解决自然科学和社会科学领域中的实际问题时所产生的统计学。它不研究数学原理的推导,也不研究公式和定理的证明。它需要的就是基本的统计方法和所研究专业的知识,两者的结合就产生了各专业的统计学,如社会统计学、人口统计学、生物统计学、医疗卫生统计学等。

应用统计学与理论统计学有三点不同之处:

第一,从研究对象来看,应用统计学是以客观现象中具体的指标数值作为主要研究对象;第二,从研究方法来看,应用统计学是研究如何运用统计学的理论和方法来研究现实问题;第三,从研究的目的来看,应用统计学以了解掌握客观现象的数量特征、数量关系和统计规律为目的。

三、统计学的应用

统计的应用非常普遍,几乎在社会生活的各个领域都能见到。它的影响也是巨大的,以至于无论怎样强调统计的重要性也不过分。

在政府中,统计资料被用来研究和影响政府在税收、公共设施建设(例如道路、桥梁等)、社会救济基金等领域方面政策的有效性;失业统计则左右着政府就业政策的制定;价格统计影响着银行利率的调整;抽样方法还用来评价各种军事装备的性能;经济学家经常帮助政府对未来的经济提供预测,他们在进行预测时要使用各种统计信息;统计决策被用于帮助人们作出科学的决定。

统计在工商企业以及农业中的应用同样重要,作用同样广泛。在工商企业中,统计方法被用于财务分析、制定生产计划、工序管理和质量控制;对销售量、雇员需求、生产率走势进行预测;还被用于市场研究、分析人事变动、出勤状况、员工评定;用于决定扩大营业;用于分析广告效果等等方面。在农业生产中,统计方法被用于田间试验;粮食产量的估计;化肥使用的最佳效果等方面。

统计方法在其他学科研究中也广泛应用。在医学和药学的研究上统计方法是必不可少的;统计方法还被用在地质学、生物学、心理学、社会学,用在那些必须在不完全的资料的基础上作出决定的任何领域;被用在教育试验和保险工程方面。在气象学、在天气预报科学中,统计方法也被使用着。

统计对于个人也是不可或缺的。可以说它融入于我们日常生活中的方方面面:在市场购买商品时,对价格和质量的判断;在就餐时帮助决定点哪一道菜;在玩扑克牌的游戏帮助确定出牌的策略;还可以帮助分析出家庭支出中的哪一个项目应该节俭……。

以上仅仅是统计应用的一部分,对此大家都有目共睹,很少有人怀疑。但对统计方法研究与应用的潜在的,更广泛的效果,可能没有被认识到,这也是我们学习统计的一个重要目的。

统计方法中有许多计算,这在以前成为广泛应用统计的障碍,而现在计算机可以帮助解决这一问题。现代统计学的特点之一就是统计研究与计算机应用相结合。现在已经开发出的统计分析软件,如 TSP、Minitab,还有专门用于统计研究的大型软件包 SPSS (Statistics Package for Social Sciences) 和 SAS (Statistics Analysis System) 等。另外,Microsoft Excel 目前的版本提供了多种统计分析的功能,是目前使用比较方便和有效的。

本书中,将适当地引入 Excel 的一些统计功能的应用,旨在展示其对统计方法的辅助作用。

第二节 基本概念

一、总体、样本及个体

1. 自然总体、测量总体

当我们需要收集数据资料研究某些问题时,必须要确定数据资料的来源,也即数据的所属范围。在统计学上将数据来源的范围定义为**总体**(population)。理解总体的概念,既可以避免对总体以外无用数据的收集,还可以使调查者对资料范围的特征认识清楚,防止总体数据的遗漏。

总体是一个简化的概念,它可以分为自然总体和测量总体。所谓**自然总体**就是由客观存在的具有相同性质的许多个别事物构成的整体。简单地说,总体就是所研究事物的全体。组成总体的那些个别事物就被称为**个体**(elements),个体通常都具有不同方面的属性。

例如,当我们要研究某个地区工业企业生产情况时,该地区的所有工业企业就成为我们研究的一个自然总体。该地区的每一家工业企业的数据,都是我们应该收集的数据,不能遗漏。而非该地区以及该地区非工业企业的数据属于非总体范围,不应包括在内。

总体所需要收集的数据资料全部来源于个体。在上例中,每一家工业企业就是一个个体,这些个体都具有“在本地区”及“工业企业”的相同性质,因此组成一个总体。除此以外,这些个体还具有其他一些不同方面的属性,如企业经济类型、职工人数、固定资产、企业增加值等,由此形成事物的多样性,也是进行统计的原因。

总体和个体并不是一个绝对的概念,它们是随着研究的范围不同而改变的。例如,当我们需要研究某市的一家工业企业的职工情况时,这家企业就是我们研究问题的自然总体,该企业的每一位职工就是个体;而当要了解全市的工业企业生产情况时,总体为该市所有的工业企业,而每一家工业企业将成为研究的个体。

自然总体中的个体通常都具有多种属性,我们把个体所具有某种共同属性的数值的整体称为一个**测量总体**。在上例中,该地区工业企业有关经济类型数据的集合,就构成一个测量总体(虽然对于每一家企业来说经济类型是一种定性问题,不是用数字来表示,但汇总之后,所有企业不同的经济类型可以用数据——比率或实际数量来表示);同样,所有固定资产数据的集合也构成一个测

量总体。

为了区别自然总体和测量总体这两个概念,我们再举一个例子。如某生产厂家要调查消费者对一种新产品的购买意向,因此要确定调查数据的范围。在这次调查中,所有的消费者就构成一个自然总体,而消费者购买意向数据的集合就是一个测量总体。消费者的年龄数据、收入数据的集合同样也可以形成测量总体。

可以说,测量总体是依附于自然总体而存在的,一个自然总体至少有一个甚至许多个测量总体,这主要取决于研究内容的多少。

在描述统计中,总体通常是指自然总体。在推断统计中,主要针对数据进行研究,总体一般是指测量总体。

2. 有限总体、无限总体

总体还可以分为有限总体和无限总体。如果组成总体的个体数量是有限个,这个总体被称为**有限总体**;否则就称为**无限总体**。

例如,对一批汽车轮胎的质量进行检查。每一个轮胎就是一个个体,一批轮胎的数量是有限的,因此该次检查的轮胎是个有限总体。对一个小区居民家庭的调查,每一个家庭就是一个个体,小区居民家庭的数量是有限的,这也是有限总体。无限总体是指总体中个体的数量是无限的。在时间相对静止的情况下,无限总体并不实际存在,通常是在总体很大时(包含个体的数量很多),假定为无限总体。如对全世界家庭数量的调查。但如果数据允许随时间而相应延长,许多总体就会成为无限总体。例如,要收集某汽车制造厂生产的汽车质量的数据,如果截止到某一时刻,则已生产的汽车数量是一个有限的台数,汽车质量的数据是一个有限总体;如果对时间不加以限制,因为组装线上每分钟都有新生产的汽车下线,那么生产汽车的数量将随时间增加而增加,所要收集的数据就是无限的,某汽车制造厂生产的汽车质量的数据就是一个无限总体。

3. 样本

所谓**样本(sample)**,就是总体中的一部分个体所构成的整体。例如,由于时间限制,审计人员会从公司所有账簿中抽选出部分账簿进行审计,被抽选出的部分账簿就是所有账簿中的样本。再比如,由于对灯泡的质量检查具有破坏性,所以检查人员只能从一批灯泡中抽选出部分灯泡进行检测,而被抽选出的部分灯泡就是样本。

样本与总体一样,也包括自然样本和测量样本。**自然样本**是在自然总体中选取的部分个体所组成的集合,而自然样本中的所有个体的测量数据的整体就构成一个**测量样本**。在对灯泡质量检查的例子中,被抽选出的部分灯泡是自然样本,被抽选的灯泡寿命的数据、亮度的数据等则是测量样本。

如果将总体比作一个集合,个体就是集合中的元素,样本则相当于这个集合

的子集. 我们知道, 一个集合可以有多个子集, 如果 $\{1, 2, 3\}$ 是一个集合, 则 $\{1\}$ 、 $\{2\}$ 、 $\{2, 3\}$ … 都是子集. 这就是说, 总体是确定的, 但样本是不确定的. 样本与总体的关系, 如图 1-1 所示.

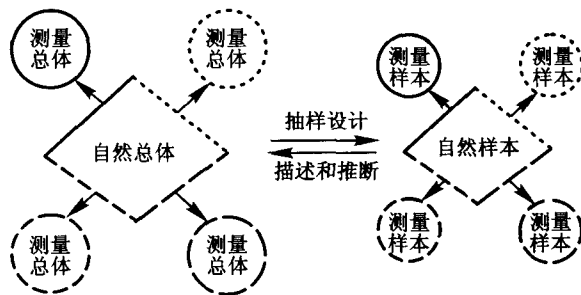


图 1-1 样本与总体关系示意图

二、标志、变量及指标

1. 标志

在自然总体中, 把反映个体所具有的某种属性或特征的概念或名称称为标志 (attribute). 标志可分为反映个体性质特征的**品质标志**和反映个体数量特征的**数量标志**.

如果企业是研究的总体, 每一个职工都是一个个体, 那么反映每一个职工属性特征的性别、民族、家庭住址等概念或名称就属于品质标志; 而反映每一个职工数量特征的年龄、身高、收入等概念就是数量标志. 如果总体中包括许多企业, 每一家企业都是一个个体, 则企业所属的行业、所有制性质等概念, 就是反映企业性质特征的品质标志; 而企业的资本金数额、固定资产原值、职工人数等就是反映企业数量特征的数量标志.

标志仅是指概念或名称, 而对标志的回答通常叫**标志表现**. 品质标志的表现一般用文字或语言描述, 数量标志的表现用数值表示.

例如, 学生性别是品质标志, 其标志具体表现为男或女, 学生的考试成绩是数量标志, 其标志表现为考试的分数.

虽然标志是用来说明个体特征的, 但对标志的数据进行综合却可以说明总体的数量特征. 因此可以说, 标志是对总体进行研究的基础.

2. 变量

按照标志表现是否相同, 标志可分为不变标志和可变标志. 标志表现相同的, 称为不变标志. 标志表现不同的, 称为可变标志, 简称为**变量 (variable)**.

例如, 在由 A 企业所有职工组成的总体中, “工作单位”这一标志对于每一

职工的回答来说都是相同的,即“A企业”,这时“工作单位”是不变标志。但“工龄”、“性别”等标志对于每一职工来说,他(她)们的回答都是不相同的,我们就称“工龄”、“性别”等标志是变量。

3. 指标

我们将反映个体特征的概念称为标志,而将反映总体或样本特征的概念称为**指标**(indicator)。所谓指标,就是指反映一些个体集合的属性和特征的概念。反映样本特征的指标叫**样本指标**,反映总体特征的指标叫**总体指标**。

指标的数据都是由标志数据计算得到的。例如,反映一家企业职工工资水平的“平均工资”就是属于指标,它是由个人的“工资”这个标志的标志值经汇总、计算而得。

由于总体、样本与个体是相对的概念,因此一个概念究竟称为指标还是称为标志,需要根据这个概念在本次研究时所反映的对象来决定。当这个概念是说明总体或者样本的特征时,它就被称为指标。当这个概念是说明个体的特征时,它就被称为标志。也就是说,同一个概念在一次研究时可能被称为指标,而在另一次研究时可能又被称为标志。

例如,在全省调查时,各地市是个体,全省是总体,此时反映各地市特征的概念是标志,反映全省特征的概念是指标;而在各地市独立调查时,各地市成为总体,其管辖的各区县成为个体,此时反映各地市特征的概念是指标,反映各区县特征的概念则变成了标志。

指标与标志只是在反映层面上有区别,除此以外,他们仅是一些概念的统称。

三、数据分类

1. 数据的四种测度水平

当变量的特征是用数据表现时,变量数据的取值称为**变量值**。变量值用来说明个体在数量方面的特征。变量能够被量化的程度是不同的,有些变量值是人为赋值,有些变量值是自然产生,虽然同为数据变量值,但代表含义却有很大不同。根据**测度水平**(measurement level)不同,变量可以分为以下四种。

(1) 定类水平的变量(nominal level)

当变量值的含义仅仅能够区别个体的不同类别,而不能说明个体的大小、程度等其他属性时,这种变量即是定类水平的变量,简称**定类变量**。这时变量的数值只表示个体的属性是相同或不同。

例如,在市场调查中“职业”这个变量就属于定类变量,其变量表现为

- A. 教育工作者 B. 公务员 C. 工人
D. 律师 E. 医生 F. 军人 G. 其他

为了在计算机中进行统计分析,我们给“职业”是“教育工作者”赋值为1,“公务员”赋值为2,“工人”赋值为3,……,依此类推。1或者2仅用来为调查对象分类,数字1并不代表最高(或最低)的级别,仅是用来表示职业是教育工作者(1)而不是军人(6)。因此对于职业变量的赋值,只能用于比较“=”或“≠”,而不能用于其他数学运算。其他如性别、民族、地理区域及出生地等标志也属于定类变量。

(2) 定序水平的变量(ordinal level)

当变量的回答不仅能够区别个体的类别而且能够代表程度的不同,这样的变量就是定序水平的变量,简称为**定序变量**。

例如,“文化程度”这个变量就属于定序变量,其变量表现为

- A. 不识字或文盲 B. 小学 C. 初中
D. 高中及中专 E. 大专及以上

上述的五种回答不仅可以说明被调查者文化程度不同,而且还表明文化程度的高低。因此我们可以给“文化程度”排序,给“不识字或文盲”个体赋值为1,“小学”赋值为2,“初中”赋值为3,……,依此类推。1或者2不仅可以用来对文化程度分类,而且数字1还代表最低的文化程度,2代表的文化程度要高于1,5代表最高的文化程度。

当然,也可以用5代表“不识字或文盲”,4代表“小学”,1代表“大专及以上”,只要数字能够反映顺序即可。但不应用无序的数字代表有序的文化程度,否则,将失去顺序的这个信息。

定序变量的量化水平高于定类变量,它所包含的信息量大于定类变量,定序变量既可以进行“=”或“≠”的比较,还可以进行“<”或“>”的运算。

(3) 定距水平的变量(interval level)

定距水平的变量可以确定个体之间的数量差别和间隔的距离,具有统一的测量单位。这些测量单位产生具有不变的相等区间的标准,从而消除了定量测量中的主观性。使用定距水平的变量测量个体时,通过变量值之间的加减运算,可以得到两个个体间关于被测量特征的确切差别。定距变量的取值范围有任意的但不是绝对的零点。零点的选定是为了方便或出于惯例,并不表示该现象不存在。

例如,“智商”的变量值是用数字表示的,甲=130,乙=120,则甲智商比乙智商高10;“温度”的变量值也是用数字表示的,A地30℃,B地10℃,可以看出A地温度比B地高出20℃。这里的“智商”和“温度”的数值都不是以绝对零点为起点,也就是说,智商=0的人不是一点智商都没有;温度=0℃,也不是说没有温度了。因此,定距水平的变量可以进行“=”或“≠”的比较,也可以进行“<”和“>”的运算,还可以进行“+”,“-”运算,但不能进行比率

计算。

类似的例子还有“世纪”概念的使用等。

(4) 定比水平的变量(ratio level)

定比水平的变量是最高层次的变量,它的观测值大小可以反映出个体之间水平的绝对和相对差异。因为定比水平变量的数据具有绝对零点,即数值为零时代表个体不具有研究的特征,所以两个数值间的比率是有意义的。定比变量可以进行“=”或“ $\neq$ ”的比较,可以进行“ $<$ ”,“ $>$ ”的运算,可以进行“+”,“-”运算,还可以进行“ $\times$ ”,“ $\div$ ”运算。

例如,“收入”就是定比变量,研究者可以说1 500元的收入比500元多1 000元,或者说1 500元的收入是500元的3倍。

其他属于定比变量的例子还有生产周期、重量、长度、时间、速度等许多其他常用的尺度。

在描述统计中,经常将定类变量与定序变量称为**定性变量**,而将定距变量与定比变量称为**定量变量**。如图1-2所示。

2. 数据的分类

根据变量值是否连续,数据分为**离散型数据**(discrete data)和**连续型数据**(continuous data)两种。如果一个变量的测量值是间断的,可以一一列举,那么这种数据就是离散型数据。

最常见的离散型数据是通过计数得到的,例如,500个被调查的人中喜欢某一特定商品的人数:它的取值是0,1,2,...,499,500;某一会计账簿上有差错账的数量:它的

取值是0,1,2,...。如果变量的取值在某一个区间上是连续不断的,即可以取得无限多个数值,则该数据为连续型。例如,从一口井内每天抽出的石油数量;正常情况下一台电视机的使用寿命等。因为石油数量与使用寿命都是一个不间断的连续值。从理论上讲,它可以无限细分至任意小数位。

根据数据的来源不同,数据分为**横贯数据**(cross-sectional data)和**纵贯数据**(longitudinal data)两种。横贯数据指对同一时间、不同个体(或总体、样本)的观察数据,有时也称静态数据;纵贯数据指对同样的个体(或总体、样本)在不同时间上的多次观察所得到的数据,有时称为动态数据。

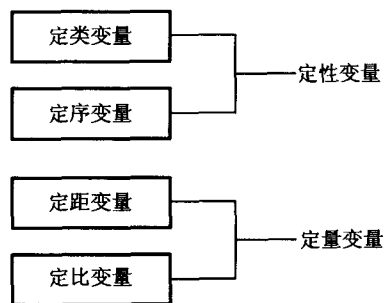


图 1-2 定性变量与定量变量示意图

本章小结

1. “统计”一词可以有三种解释,应根据不同的场合和语言环境来理解。

统计活动一般包括统计调查、统计整理、统计描述、统计分析和数据显示等程序。统计资料是指由统计活动产生的原始的或已加工、整理过的客观现象的数据资料。统计学是一门关于统计资料的收集、显示、描述和分析方法的学科。

2. 总体、样本是集合体的概念,个体是具体事物的概念,正确确定它们的范围,才能保证数据的准确性。标志、指标是关于特征的概念总称,在调查阶段和整理阶段,区别两者是有意义的。在实际工作中标志和指标有时不加以区别,都叫做指标,有时将指标及指标的数值统称为指标。

3. 数据本身并没有层次,其层次是通过变量代表的特征来决定的。定类变量与定序变量属于定性变量,也可以称为品质标志;同样,定距变量与定比变量属于定量变量,也可以成为数量标志。它们只是在推断统计与描述统计里的不同名称而已。

4. 统计的应用如此普遍,在现实中它几乎无处不在。但要使统计发挥真正的作用,就不能出现以下几种情况,否则,就会使看似精确的统计结果变成一堆废话:

- (1) 使用者事先已作出结论,再去挑选它们的数据。
- (2) 使用者故意扭曲、夸张、简化或刻意篡改数据,以便达到他们的目的。
- (3) 使用者还不能正确理解并恰当地使用统计方法,使别人误解。
- (4) 读者不能真正懂得公布的统计术语的含义,产生误解。

正如 H. T. 海斯莱特对这种人所作的评论那样:他使用统计就如同一个醉汉使用街灯一样,是用来支撑他,而不是用来照明。

思考与练习

1. 统计方法在科学研究中起什么作用?举例说明统计方法在实际生活中的应用。

2. 统计活动的环节之间有什么关系?统计活动的最终目的是什么?

3. 某大学拟从该校 20 000 名在校生中抽选 1 000 人进行调查,以了解大学生课外生活情况。调查项目主要包括:学生所在年级、课外时间的分配、课外活动的形式及占用时间、最喜欢的课外活动等。请写出这次调查的总体、样本及个体都是什么?

4. 根据题 3 写出调查项目中的数据属于哪一种测度水平?

5. 在下列问题中指出总体是有限总体还是无限总体?是自然总体还是测量总体?

- (1) 社会上还流传的绝版书籍。
- (2) 已使用过某品牌家用电器的家庭的人口数。
- (3) 新产品的未来消费者。