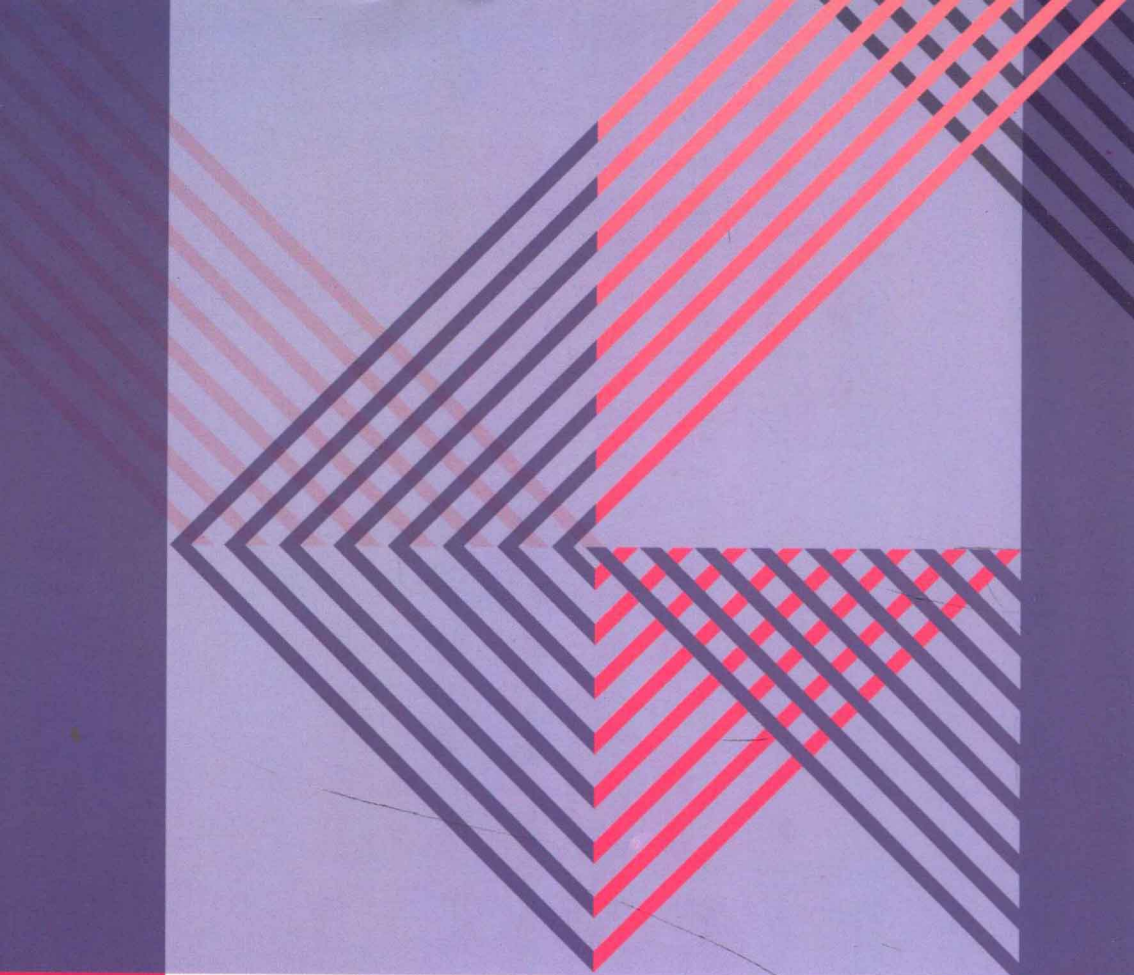




全国成人高等教育规划教材

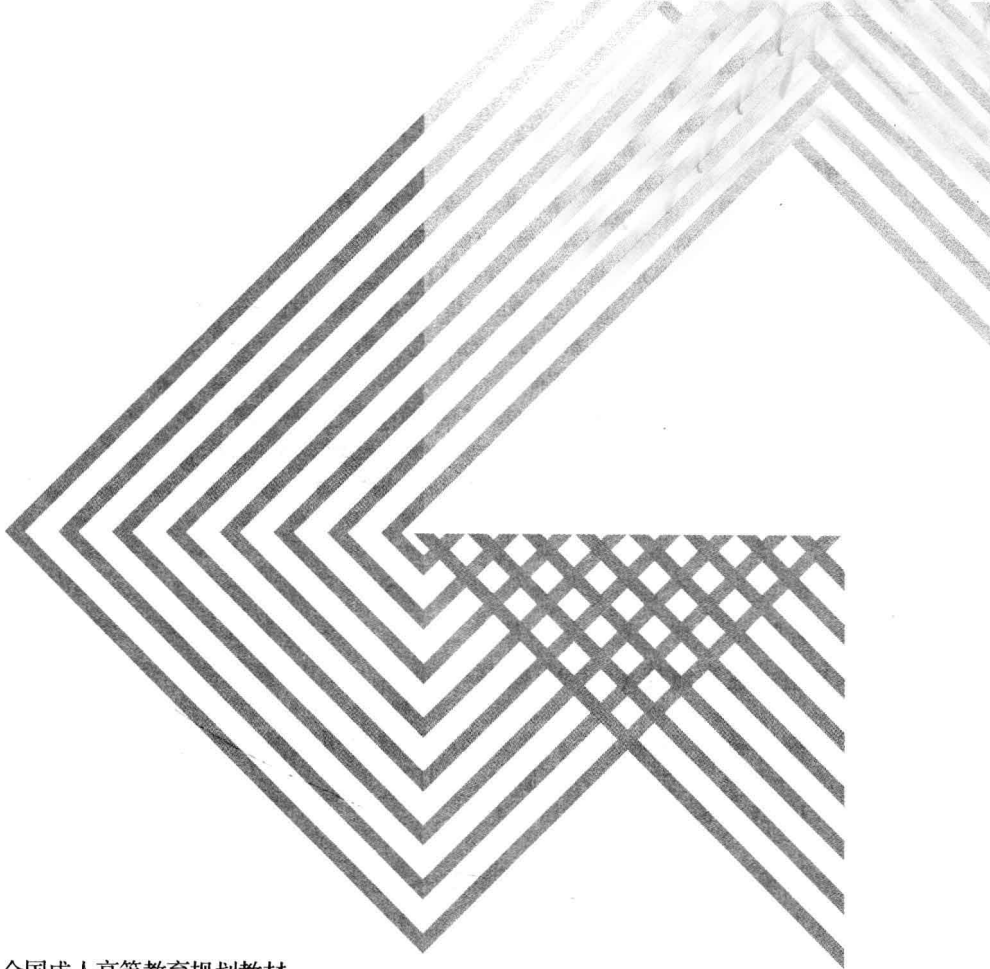
统计学

(第二版)

教育部高等教育司 组编

邱 东 主编

 高等教育出版社
HIGHER EDUCATION PRESS



全国成人高等教育规划教材

统计学 (第二版)

T o n g j i x u e

教育部高等教育司 组编

邱 东 主编

 高等教育出版社·北京
HIGHER EDUCATION PRESS BEIJING

内容简介

本书是教育部高教司组编的全国成人高等教育规划教材的修订版,侧重于经济管理领域统计学知识的介绍,有着鲜明的经济与社会统计学特色,是一本理论与实践充分结合的统计学教材。

本书的知识体系包括:总论、数据的收集、数据的整理与显示、参数估计与抽样推断、假设检验、相关与回归分析、时间序列分析、统计指数、经济统计概要、社会统计概要等。

本书经再版修订后,不仅可作为成人高等学校经济管理类相关专业的统计学课程教材,也可作为高等学校普通本科经济管理类相关专业的统计学课程教材,还可作为经济管理领域从业人员的统计学参考用书。

图书在版编目(CIP)数据

统计学 / 邱东主编;教育部高等教育司组编. --2
版. --北京:高等教育出版社,2012.12
ISBN 978-7-04-035630-4

I. ①统… II. ①邱… ②教… 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(2012)第249223号

策划编辑 施春花	责任编辑 施春花	封面设计 张楠	版式设计 马敬茹
插图绘制 郝林	责任校对 金辉	责任印制 毛斯璐	

出版发行	高等教育出版社	网 址	http://www.hep.edu.cn
社 址	北京市西城区德外大街4号		http://www.hep.com.cn
邮政编码	100120	网上订购	http://www.landaco.com
印 刷	国防工业出版社印刷厂		http://www.landaco.com.cn
开 本	787mm×960mm 1/16	版 次	1999年6月第1版
印 张	21.5		2012年12月第2版
字 数	390千字	印 次	2012年12月第1次印刷
购书热线	010-58581118	定 价	31.40元
咨询电话	400-810-0598		

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 35630-00

第二版前言

统计学是为适应国家管理和社会经济需要而产生和发展起来的。进入 21 世纪以来,我国统计学得到了快速的发展,社会各界对统计学知识的需求也不断增强。

1999 年,在教育部高等教育司制订的全国成人高等教育主要课程教材建设规划的指导下,我们与高等教育出版社合作,编写了《统计学》教材,该教材发行后,受到很多读者的欢迎。总结起来,该教材内容简洁、应用导向明确的特点是值得肯定的。十几年过去了,我国《统计学》教材建设进入了一个新的阶段,经过沟通,我们决定再次与高等教育出版社合作,推出新版的《统计学》教材。在这一版的教材中,我们保持了原有《统计学》教材的特色,同时又结合新形势下的统计学教学特点进行了一些优化,以适应当前统计学的教学需要。

本书重点加强了以下几方面的工作:一是侧重经济管理领域统计学知识的介绍。统计学知识在经济管理领域应用广阔,本书在引例设置、内容选取、例题分析等方面引入了大量经济与管理统计学素材,重点为经济管理相关专业统计学初学者服务。二是强化了经济与社会统计学的知识体系。经济与社会统计学是统计学的重要组成部分,当前亟需把有关经济与社会统计学的进展引入统计学教材建设中来,为此,本书专门设置了经济统计概要和社会统计概要两章,以进一步完善统计学的知识体系。三是坚持理论与实践充分结合的思路。本书在介绍统计学基本理论的同时,重点加强了有关统计方法的应用介绍,使读者既能知其然,知其所以然,又能知其为何所用,如何使用。

邱东教授负责本书的框架设计和总纂工作。参与本书编写工作的人员有:邱东教授(第一章)、刘丹丹博士(第二章)、杨烨军博士(第三章)、石刚副教授(第四章、第五章)、吕光明副教授(第六章、第七章,张军丽同学对第六章编写亦有贡献)、陈梦根教授(第八章)、王亚菲副教授(第九章)、席玮博士(第十章)。

我们在本书编写过程中参阅了国内外诸多的教材与著作,在此向有关作者致以诚挚的谢意。同时,要特别感谢高等教育出版社的施春花编辑为本书的出版所做的大量工作。

由于编者学识水平所限,书中难免会有一些不当之处,我们衷心希望得到有关专家和同仁的指教,以便在今后加以改进。

编者

2012年10月

第一版前言

《统计学》是由教育部高等教育司组织编写的,是一部适用于成人专科水平的统计学专业教材。

在本书的编写中,编者注意到成人教材的特点,在吸收国内外有关教材优点的基础上,我们突出了基本理论、基本知识和基本技能的传授,在概念表述上力求准确,在内容安排上力求适度。每章都有内容提示、本章小结、思考与练习等,在内容论述中也安排了大量的例解,以尽可能地适应成人学习的需要。

参加本书编写的有:东北财经大学邱东(第一章、第十章),中国人民大学王琪延(第二章、第八章、第九章),中央广播电视大学熊应进(第三章、第四章、第五章),东北财经大学李培军(第六章、第七章)。由邱东任主编,负责全书的审订、修改、总纂和定稿工作,李培军协助主编做了大量工作。本书还请中国统计学会投资统计分会副会长、东北财经大学马秀岩教授主审。

我们非常感谢教育部高等教育司有关领导对本书编写的组织和领导,感谢高等教育出版社特别是本书的责任编辑孙乃彬对本书出版的大力支持,还要感谢马秀岩教授为提高本书质量所做的辛苦工作。在本书编写中,我们还参考了国内外有关教材的内容,在此一并感谢!

由于编者学识水平有限,书中会有一些不恰当乃至错误的地方,我们衷心希望得到专家和同仁的指教,以便在今后加以改进。

编者

1999年1月10日

目 录

第一章 总论	1
第一节 走近统计学	1
第二节 统计数据	5
第三节 统计学的几个基本概念	7
本章小结	8
思考与练习	9
第二章 数据的收集	10
第一节 数据的收集方法	11
第二节 调查问卷设计	25
第三节 量表	31
本章小结	37
思考与练习	38
第三章 数据的整理与显示	40
第一节 统计数据整理	41
第二节 统计图和统计表	53
第三节 统计分布的集中趋势测度	57
第四节 统计分布的离散程度与分布形态测度	71
本章小结	84
思考与练习	85
第四章 参数估计与抽样推断	87
第一节 参数估计与抽样推断中的基本概念	88
第二节 点估计	95
第三节 区间估计	100
第四节 几种基本的抽样方法	110
本章小结	119
思考与练习	119
第五章 假设检验	121
第一节 假设检验中的基本问题	122
第二节 总体均值的假设检验	127
第三节 总体比例的假设检验	139

第四节 总体方差的假设检验	143
本章小结	146
思考与练习	147
第六章 相关与回归分析	149
第一节 相关分析	150
第二节 一元线性回归分析	157
第三节 多元线性回归分析	167
第四节 非线性回归分析	173
本章小结	178
思考与练习	179
第七章 时间序列分析	185
第一节 时间序列的动态描述分析	186
第二节 时间序列的成分分析	196
第三节 时间序列的成分预测	211
本章小结	218
思考与练习	219
第八章 统计指数	223
第一节 统计指数的概念与分类	223
第二节 指数编制方法	226
第三节 指数体系和因素分析法	235
第四节 常见统计指数介绍	243
本章小结	251
思考与练习	252
第九章 经济统计概要	254
第一节 国内生产总值统计	254
第二节 经济过程统计	262
第三节 经济动态统计	276
本章小结	282
思考与练习	283
第十章 社会统计概要	284
第一节 人口与人力资源统计	285
第二节 居民生活水平统计	300
第三节 社会发展水平综合评价	312
本章小结	319
思考与练习	319
附录 常用统计表	321
附表1 标准正态分布表	321

附表 2 t 分布表	323
附表 3 χ^2 分布表	324
附表 4 F 分布表	326
参考文献	331

第一章 总 论

知 识 点

1. 统计学的定义、任务与分类
2. 数据及其分类
3. 总体和样本
4. 变量和指标

引 例

现代社会已经是一个海量信息社会,信息成了经济发动机的新“燃料”,“越来越多”的人用“越来越多”的时间来面对“越来越多”的信息,信息的作用“越来越重要”,这表明了信息对生活影响的广度和深度,统计信息与人们的利益更是息息相关。

如果单从统计的角度看,世界上无非只有两种人,一种是从事统计工作的人,一种是使用统计信息的人。当然,统计信息的使用者是多数,哪怕并不从事与信息相关的职业,不管我们本人是否意识到,我们也仍然可以是统计信息的使用者,统计信息在人们的生活中随处可见。

而同样的信息,对不同知识背景的人而言,所能利用到的信息量是截然不同的。从技术和方法论的角度看,要使用,就有一个会不会使用的问题。如果能对这些信息的“生产过程”有一个基本的了解,知道它是如何统计出来的,知道如何来识别和使用这些信息,知道使用中还应注意哪些问题,就会给我们带来更多的益处,提高我们生活和工作的质量。

第一节 走近统计学

统计是为适应国家管理需要和社会政治经济的发展而产生并发展起来的。

一、统计的产生与发展

统计的起源很早,在古代奴隶制社会,由于国家在赋税、徭役、征兵等方面的需要,就开始了人口、土地等基本国情的登记和计算工作。

据历史记载,我国夏禹时代(公元前 22 世纪)分为九州,有人口 1 355 万,可见人口统计的久远。《书经·禹贡篇》记述了九州的基本土地状况,被西方经济学家推崇为“统计学最早的萌芽”。西周建立了较为系统的统计报告制度,统计作为治理国家的重要手段已经被人们所认识。春秋时管子曾说过:“不明于计数而欲举大事,犹无舟楫而欲经于水险也。”秦时《商君书》中提出“强国知十三数”,其中包括粮食储备、各类人口数、农业生产资料及自然资源等。在中国封建社会,户籍统计和田亩统计等都有很大的发展,其制度、方法和组织都居于当时世界先进水平。

在地中海沿岸,统计活动也有悠久的历史。公元前 27 世纪,埃及为了建造金字塔和大型农业灌溉系统,曾进行全国人口和财产调查。公元前 15 世纪,犹太人为了战争的需要进行了男丁的调查。公元前约 6 世纪,罗马帝国规定每 5 年进行一次人口、土地、牲畜和家奴的调查,并以财产总额作为划分贫富等级和征丁课税的依据。到了 15 至 18 世纪,欧洲出现了许多以报道国情为内容的统计著作。

到了近代,统计得到了广泛发展。除了在人口、税收、土地等传统领域外,商业、航运、外贸和工业等领域统计数字的记录和传播也空前活跃,其范围逐渐从国家管理扩展到社会经济活动的领域。从 18 世纪起,许多国家先后设立专门的统计机关,定期或不定期地举行人口、工业、农业、贸易和交通等调查,收集统计资料,出版统计刊物。19 世纪 30 年代,还出现了所谓“统计狂热时代”。在一些统计学家的积极倡导下,召开了国际统计会议,并建立了国际统计组织。

现代社会中,统计更是无所不在。统计在医学、心理、生物、物理、化学、地质,甚至考古、军事、文学等领域都有着广泛的应用。可以说,统计已经成为一种工具,只要有数据的地方就要用到统计学。

二、什么是统计学

统计一词源于中世纪拉丁语,意思是各种现象的状态和状况。在中国古代,统计一词仅仅有数据总计的意思。现代意义的统计,大约是在 20 世纪初传到中国的。

一般认为,统计学是处理数据的一门科学。具体说来,统计学是研究如何收集、整理、分析、核算数据,并从数据中得出结论的科学。

统计学的主要任务包括:

(1) 数据收集。数据收集是统计工作的基础,没有数据也就没有统计学。如何获得需要的数据、如何从海量数据中识别有用的数据,需要有专门的方法。

(2) 数据整理。数据收集来以后,还要进行必要的整理。比如将数据以图表的方式展现出来,使数据更加直观、清晰和易读。

(3) 数据分析。数据分析是指运用统计方法对有关数据进行处理,并从数据中提取有效信息的过程。如何正确选择和运用统计方法对于数据分析至关重要。恰当的数据分析方法可以使人们准确地描述和推断数据所表达的信息,并从数据中得到想要的结论。

(4) 数据核算。统计学还要在有关学科理论的指导下,通过制定有关定义、标准、方案和体系,为数据的收集、整理和分析提供核算框架。尤其是在社会经济统计领域,为规范数据的开发与分析,核算方法有着较多的应用。

三、统计学的分类

通常可以把统计学分成描述统计学和推断统计学两类。

(一) 描述统计学

描述统计学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、整理和显示,进而通过综合、概括与分析得出反映事物的数量特征和数量关系的统计方法。

(二) 推断统计学

推断统计学是研究如何利用样本数据来推断总体特征的统计方法。它是在对收集的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。其中总体是指所要研究事物的全体,样本是指从总体中抽取的一部分单位的集合。

比如,要了解消费者对某一种产品的偏好,不可能对每个消费者一一进行测试,而只能选取一部分消费者,通过这部分消费者的情况对全体消费者的情况进行推断。这就用到了推断统计学。

描述统计学与推断统计学的划分,反映了数据分析的不同层次。如果所处理的数据是一个总体的数据,则可以通过描述统计学认识该总体的数量规律性;而如果所处理的数据只是总体的一部分,那么就必须通过推断统计学对总体的数量规律性进行科学的推断。

描述统计学与推断统计学都是统计学的重要组成部分。其中描述统计学是统计学的基础;推断统计学是统计学的主干。随着数据量的不断增多和统计任务的不断复杂化,统计工作的重心已经从对过去的回顾转移到对现状的分析和

对未来的预测上。在当前情况下,大量的统计工作所收集到的数据是样本数据,因此,推断统计学在统计学中的地位和作用日益增强。当然,这并不意味着描述统计学已经不重要了,事实上,描述统计学在日常工作和生活中一直有大量的应用,而且,如果没有描述统计提供有效的样本信息,推断统计也难以进行。

四、统计使人豁达

哲人培根曾经说过:“历史使人明智,诗歌使人巧慧,数学使人精细,博物使人深沉,伦理使人庄重,逻辑与修辞使人善辩。”按照这个思路,我们想加上一句,这就是:统计使人豁达。

那么具体说来,统计学对人们有什么帮助呢?

(一) 良好的统计训练使人宏观意识比较强

总体和总量是统计学中最基本的概念,虽然统计调查和处理都要从个体入手,但其最终目的是为了通过它对现象的总体做出评价。统计总是把着眼点放在宏观大势的把握上,对个体数据如何处理,也取决于怎样科学地得到对总体的认识。久而久之,统计的训练使人宏观意识得到强化,遇到问题习惯将其放在时空的大背景中去考虑。比方说,判断学校某个学科的地位,不仅要将其与校内其他学科相比较,更要将其与国内其他院校的相同学科进行比较。在制定长期发展目标时,还要与发达国家该学科水平相比较,只有在大系统中去把握,所得结论才会是可靠的。先人云:会当凌绝顶,一览众山小。统计训练能给人的就是这种感觉。

(二) 良好的统计训练使人在看待和处理问题时不那么偏激

统计中的平均分析和平衡分析都在教导人们对事物持中庸的态度。统计人的豁达也源于这一点。知道了平均水平,便把握了事物的一般趋势,心中有数;同时也知道事物围绕着平均水平还会有变异,当好的极端出现时,能受宠不惊,而当差的极端出现时,又可视之泰然,胸襟于是开阔通畅。

(三) 良好的统计训练使人注重研究事物的数量规律

统计注意研究事物量变引起质变的度,注意研究事物的数量规律。统计是用数据说话的,不能捕风捉影,得出的认识通常有方法依据和信息支持,比起单纯的理论说教,能给人以更实在的感受,分寸感强。

总而言之,现代社会对统计是很认可的。进入 21 世纪以来,我国统计学发展也进入了快车道,统计的专业优势也开始得到社会的广泛认可。

第二节 统计数据

如前所述,统计学是处理数据的科学。统计离不开数据,让我们先对数据有所认识。

有人觉得数据就是一堆枯燥的数字,其实事实并非如此,数据是立体的、多元的,有着丰富的内涵和表现形式。用统计学的语言来说,数据是人们对现象进行调查研究所收集、整理、分析和解释的事实和数字,是对客观现象进行观测、计量的结果。

数据有很多种类,不同种类的数据对应着不同的统计处理方法。下面,我们来认识各种各样的统计数据。

一、分类数据、顺序数据、间隔数据、比率数据

按照测量尺度的不同,数据可分为四类:分类数据、顺序数据、间隔数据、比率数据。

(一) 分类数据

分类数据(categorical data)是按照研究对象的某种属性将其划分为若干类别的数据。例如,按照性别将人口分为男性和女性,按照民族将人口分为汉族、回族、满族等,按照基金投资对象的不同,将基金分为股票型基金、债券型基金、混合型基金、货币型基金、指数型基金等。为便于统计处理,往往将不同类别的事物用不同的数字标识,如“1”表示男性,“0”表示女性,但这些数字仅仅是不同类别的代码,不代表真正的值,不能够比较大小或进行运算。

分类数据只是对事物类别的区分,各类别之间是平等的并列关系,没有优劣、高低、大小之分,不同类别之间的顺序是可以改变的。

(二) 顺序数据

顺序数据(rank data)是测度事物之间等级差别或顺序差别的一类数据,例如高校教师职称从低到高分别为助教、讲师、副教授、教授,消费者对某种商品的评价为很差、差、一般、好、很好等。

顺序数据的不同类别间存在顺序性差异,因此能进行比较和排序,但这种差异无法准确计量,即只能说明这一类比那一类高,却说明不了具体高多少。

(三) 间隔数据

间隔数据(interval data)是对事物绝对差异进行测度的一类数据。间隔数据不仅能区分事物的类别、进行比较和排序,还能精确地计量出两个数字之间的

差距。间隔数据之间的每一间隔是相等的,可以进行加减运算,但由于没有真实反映绝对零点而不能进行乘除运算。温度是典型的间隔数据,我们可以计算出 20°C 比 10°C 高 10°C ,但不能说 20°C 是 10°C 的2倍。

(四) 比率数据

比率数据(ratio data)是测度事物绝对和相对差距的一类数据。大多数定量数据都属于比率数据,如年龄、身高、收入、GDP总额、通货膨胀率、商品的销量和价格等。

比率数据具有绝对零点。除了对事物分类、排序、比较大小、计算差距之外,还可以计算两个数值的比率。例如,商品A的价格是100元,商品B的价格是50元,我们既可以计算出A比B高50元,也可以计算出A的价格是B的2倍。

按照这四类数据的特点,又可将分类数据和顺序数据合称为定性数据,因为这两类数据反映的是事物的品质特征,一般不能用数值表示;将间隔数据和比率数据合称为定量数据,因为这两类数据反映的是事物的数量特征,一般表现为数值形式。

二、原始数据和二手数据

从统计数据本身看,它们最初都是来源于直接的调查或实验。但从使用者的角度看,数据主要来源于两种渠道,一是来源于直接的调查和实验,二是来源于别人调查或实验的数据。据此,可将其分为原始数据(primary data)和二手数据(secondary data)。原始数据是为了解决特定问题而专门收集的数据资料;二手数据是已经收集到的、不一定与当前问题相关的数据资料。

原始数据需要调查人员亲自进行实地调查,耗时长、成本高。二手数据的收集则只涉及寻找、摘录、整理等过程,相对较为简单,具有成本低、速度快的优点。

三、时间序列数据、截面数据和面板数据

按照被描述的现象与时间的关系,可将数据分为时间序列数据(time series data)、截面数据(cross-sectional data)和面板数据(panel data)。

时间序列数据是指同一空间不同时间上的数据,如2000—2011年北京市的国内生产总值(GDP)数据。截面数据是指同一时间不同空间上的数据,如2011年全国各省区直辖市的GDP数据。面板数据是时间数据和截面数据综合起来的一种数据类型,具有时间和空间两个维度,当这类数据按两个维度排列时,是排在一个平面上,与只有一个维度的数据排在一条线上有着明显的不同,整个表格像是一个面板,所以称为“面板数据”。如2000—2011年全国各省区直辖市的GDP数据就是面板数据。

第三节 统计学的几个基本概念

统计学要用到许多概念,但有一些基本概念是经常用到的,这里先加以介绍。

一、总体

所谓总体(population),是根据一定目的确定的所要研究的事物的全体,它是由客观存在的、具有某种共同性质的许多个别事物构成的整体。组成总体的每个个别事物称为个体,也称为总体单位。总体中个体的数量被称为总体容量。

总体具有大量性、同质性和差异性三个特征。大量性是指总体中的个体数量要足够多;同质性是指总体中的各个个体必须具有某种共同属性或特征;差异性是指构成总体的各个个体在某些方面又必须存在一定的差异。

总体根据其包含的个体数量是否可数可以分为有限总体和无限总体。比如,某大学的学生数量是可数的,其学生总体就是有限总体。而对空气质量的研究,空气就是无限总体。

用来描述总体特征的某种特征值被称为总体参数(parameter),在统计学中,总体参数通常用希腊字母表示。

二、样本

统计学要研究总体的特征。但是,当总体中的个体数量很多,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,可以抽取总体中的一部分个体,作为总体的代表加以研究。这种由总体的部分个体组成的集合被称为样本(sample)。样本中个体的数量被称为样本容量。

用来描述样本特征的某种特征值被称为样本统计量(sample statistic),在统计学中,样本统计量一般用英文字母表示。

总体参数往往是很难获取的,而样本统计量是可知的,因此,用样本统计量去估计总体参数是统计学的一项重要工作。

三、变量

变量(variable)是统计学中的重要概念,统计研究往往从研究对象的某些特征开始。比如,要了解一支球队的情况,可以从队员的籍贯、年龄、身高、体重、进球数量等特征入手。显然,不同队员的上述特征是不尽相同的,一般可称之为

变量。

准确地说,变量是说明现象某一特征的概念。

变量的具体取值就是变量值。比如一个球员的身高是 180 cm,这里的 180 cm 就是“身高”这一变量的“变量值”。

根据变量值是否连续出现,变量可分为连续型变量和离散型变量。连续型变量是指变量的取值在数轴上连续不断,无法一一列举,比如上述“身高”、“体重”等变量。离散型变量是指变量的数值在数轴上是不连续的,只是一个个独立的点,比如上述的“进球数量”。

受随机性因素影响的变量称为随机变量。比如,球员每场进球数要受到竞技状态、裁判、天气、队友配合等多方面不确定因素的影响,是一个随机变量。再比如,公共汽车站的候车人数也是随机变量。

随机变量尽管受不确定性因素影响,但也有一定的规律性,通过科学方法找到随机变量变动的规律性并据此做出分析和预测,是统计学研究的重要内容。

四、指标

指标是反映总体数量特征的概念和数值。比如,球队的平均身高是 185 cm,其中“平均身高”是概念,“185 cm”是数值,既有概念又有数值的“平均身高 185 cm”是指标。

人们日常生活中经常可以见到各种各样的统计指标,比如“2011 年,我国国内生产总值 471 564 亿元,比上年增长 9.2%。其中,第一产业增加值 47 712 亿元,增长 4.5%;第二产业增加值 220 592 亿元,增长 10.6%;第三产业增加值 203 260 亿元,增长 8.9%。第一产业增加值占国内生产总值的比重为 10.1%,第二产业增加值比重为 46.8%,第三产业增加值比重为 43.1%”。

从结构上看,指标包括变量和变量值。但要注意到,指标更加强调对总体的描述,而变量和变量值往往是针对个体而言的。

客观现象是错综复杂的,为了更全面地反映总体各方面的情况,有时需要通过多个不同指标来研究总体。这些由反映同一总体多个方面数量特征的、一系列相互联系的指标所组成的体系,称为指标体系。

本章小结

1. 统计是为适应国家管理需要和社会政治经济的发展而产生并发展起来的。
2. 统计学是研究如何收集、整理、分析、核算数据,并从数据中得出结论的