



北京市高等教育精品教材立项项目

统计学

TONG JI
XUE

刘扬 毛炳寰 主编

 中国统计出版社
China Statistics Press



北京市高等教育精品教材立项项目

统计学

TONGJI

XUE

刘扬 毛炳寰 主编



中国统计出版社
China Statistics Press

(京)新登字 041 号

图书在版编目(CIP)数据

统计学/刘扬,毛炳寰著. —北京:中国统计出版社,2010.7

ISBN 978-7-5037-5964-2

I. ①统… II. ①刘… ②毛… 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(2010)第 115465 号

统计学

作者/刘扬 毛炳寰

责任编辑/吕军

装帧设计/杨超

出版发行/中国统计出版社

通信地址/北京市西城区月坛南街 57 号 邮政编码/100826

办公地址/北京市丰台区西三环南路甲 6 号

网 址/www.stats.gov.cn/tjshujia

电 话/邮购(010)63376907 书店(010)68783172

印 刷/河北天普润印刷厂

经 销/新华书店

开 本/787×1092mm 1/16

字 数/370 千字

印 张/15.25

版 别/2010 年 7 月第 1 版

版 次/2010 年 7 月第 1 次印刷

书 号/ISBN 978-7-5037-5964-2/C·2338

定 价/28.00 元

中国统计版图书,版权所有。侵权必究。

中国统计版图书,如有印装错误,本社发行部负责调换。

前言

伴随着信息时代数据的爆发式增长,数据分析技能已经成为各类职业对从业人员的一项基本要求,也相应地成为当代大学生的必备技能。作为一门提供数据分析技能的课程,统计学的重要性也日益提高。2009年8月5日的《纽约时报》中一篇题为《对当今大学生,关键只有一个:统计学》的报道对统计学的广泛应用进行了评述。

本教材是为了满足非统计专业统计学的教学需要而编写的,主要针对的是学习了概率论与数理统计这门课程的经济、管理类学生。从内容上说本教材有以下几个方面的特色:

1. 由于学生已经具备了概率论和数理统计基础,本教材避免了对相关内容的重复介绍,对相应的知识点只做了简要的概括,但补充了一些新的知识点和软件操作的内容。

2. 随着统计软件的迅速发展和普及,一些原本需要编写复杂的计算机程序才能实现的统计方法已经变得“傻瓜化”,非统计学专业的学生完全有能力和必要掌握这些方法的基本原理并学会正确运用。根据经济、管理类专业的需要,本教材增加了非参数统计、主成分分析与因子分析、聚类分析与判别分析以及对应分析等内容。根据我们在中央财经大学的教学经验,在要求学生通过教学资料自学统计软件的情况下,本教材的内容完全可以在54课时的时间内完成。

3. 根据国内统计软件的应用情况,本教材对IBM SPSS Statistics软件(教材中简称为SPSS)的相关操作进行了详细的讲解,对软件的输出结果进行了必要的解释。这些内容为学运用相应的统计方法处理实际问题提供了保证,使学生能够学以致用。为了使读者对SPSS的输出结果有更加直观的了解,教材中对部分SPSS输出的表格未作修改(请注意这些表格的格式并不完全符合统计规范)。

4. 本教材主要着眼于培养学生运用统计方法解决实际问题的能力,尽量避免了统计方法涉及的数学推导。这一方面可以增加学生的学习兴趣,减少对统计学的畏难情绪,另一方面也节约了教材的篇幅。

本教材由中央财经大学统计学院的教师共同编写,刘扬、毛炳寰担任主编。各章分工如下:第一章、第十章刘扬;第二章王健;第三章王亚菲;第四章、第五章毛炳寰;第六章吕光明;第七章赵楠;第八章胡永宏;第九章满向昱;第十一章张宝军;第十二章马景义。编写过程中由主编提出教材编写大纲,各章的负责教师提供初稿,由主编对内容进行增删和修改。其中,为了尽可能保持风格的一致,毛炳寰对第六章、第十二章进行了



较大幅度的改写。

为了方便教学和学生学习,本书的配套资料,包括数据集、PowerPoint 课件、统计软件指导、实验操作题、习题答案等,均在中央财经大学统计学精品课程网站(<http://stat.cufe.edu.cn/statistics/>)上公布,欢迎读者访问。

在教材的编写过程中我们参考了许多的教材和网络资料,在此向作者们表示感谢。但由于参考的网络资料甚多,为了节约篇幅在参考文献中未一一列出。

本教材的编写得到了北京市教委“北京高等教育精品教材建设项目”立项资助以及中央财经大学教务处的资助。硕士研究生余玥、祝新宇细致的校对工作和修改建议使本书避免了许多错误。统计出版社的吕军编辑在统计教材出版方面的丰富经验和细心的工作是本书得以顺利出版的保证。

因受水平所限,书中错漏之处在所难免,欢迎读者和专家批评指正。联系 email 地址为:maobinghuan@gmail.com。

编者

2010年3月于中央财经大学

目 录

第一章 绪 论	(1)
第一节 统计与数据	(1)
一、什么是统计学	(1)
二、什么是数据	(2)
第二节 统计学中的基本概念	(3)
一、确定性和随机性	(3)
二、总体和样本	(3)
三、参数和统计量	(4)
四、变量和指标	(4)
第三节 常用统计软件简介	(6)
一、几种常用的综合统计软件	(6)
二、关于学习和使用统计软件的几点说明	(7)
小结	(8)
思考与练习	(8)
第二章 数据的搜集	(9)
第一节 数据的来源	(10)
一、一手数据和二手数据	(10)
二、常用的统计调查方式	(11)
第二节 抽样调查	(13)
一、概率抽样	(13)
二、非概率抽样	(16)
三、抽样误差与非抽样误差	(18)
第三节 调查方案设计	(21)
一、调查方案的基本结构	(21)
二、数据的收集方法	(22)
三、调查问卷设计	(24)
小结	(31)
思考与练习	(31)



第三章 数据的描述	(33)
第一节 统计图与统计表	(33)
一、频数分布与统计分组	(33)
二、列联表	(36)
三、常用统计图	(37)
四、绘制统计图应注意的问题	(41)
五、统计表	(42)
第二节 数据描述的数值方法	(42)
一、集中趋势	(42)
二、离散程度	(48)
三、数据分布形态的描述	(50)
第三节 用 SPSS 进行描述统计分析	(52)
一、用 SPSS 绘制统计图表	(52)
二、用 SPSS 计算描述统计量	(55)
小结	(58)
思考与练习	(59)
第四章 参数估计与假设检验	(61)
第一节 参数估计	(61)
一、参数估计的基本概念	(61)
二、抽样分布	(62)
三、总体均值和比例的区间估计和软件实现	(66)
四、样本容量的计算	(68)
第二节 假设检验	(70)
一、假设检验的基本原理	(70)
二、假设检验中的 p 值	(72)
三、几种常用假设检验方法的软件实现和结果分析	(72)
小结	(77)
思考与练习	(78)
第五章 方差分析	(79)
第一节 方差分析中的基本概念和假设	(79)
一、方差分析中的基本概念	(79)
二、方差分析中的基本假设与检验	(80)
第二节 单因素方差分析	(81)
一、单因素方差分析的数据结构和模型	(81)
二、方差分析的基本原理	(82)
三、方差分析的步骤	(83)

四、方差分析中的多重比较	(86)
第三节 双因素方差分析	(87)
一、无交互作用的双因素方差分析模型	(88)
二、有交互作用的双因素方差分析模型	(89)
三、双因素方差分析的步骤	(90)
第四节 SPSS 方差分析的相关操作	(93)
一、对数据格式的要求	(93)
二、用 SPSS 检验数据分布的正态性	(93)
三、用 SPSS 进行单因素方差分析和多重比较	(95)
四、用 SPSS 进行双因素方差分析	(95)
小结	(97)
思考与练习	(98)
第六章 非参数检验	(101)
第一节 非参数检验概述	(101)
第二节 单样本的非参数检验	(102)
一、 χ^2 拟合优度检验	(102)
二、单样本 K-S 检验	(104)
三、单样本中位数的符号检验	(106)
第三节 两个样本和多个样本的非参数检验	(108)
一、匹配样本的非参数检验	(108)
二、两个独立样本的 Wilcoxon 秩和检验	(111)
三、多个独立样本的 Kruskal-Wallis 检验	(113)
小结	(115)
思考与练习	(115)
第七章 相关与回归分析	(117)
第一节 相关分析	(117)
一、函数关系与相关关系	(117)
二、相关关系的类型	(118)
三、相关关系的测度	(118)
第二节 一元线性回归分析	(123)
一、总体回归函数与样本回归函数	(123)
二、一元线性回归模型的估计	(124)
三、一元线性回归模型的评价和检验	(126)
第三节 多元线性回归分析	(132)
一、多元线性回归模型	(132)
二、多元线性回归方程的参数估计	(132)
三、回归方程的拟合优度	(133)



四、显著性检验	(134)
第四节 回归分析的其他问题	(136)
一、非线性回归分析	(136)
二、逐步回归	(139)
三、违背回归模型统计假设的后果和补救办法	(140)
小结	(141)
思考与练习	(142)
第八章 时间序列分析	(146)
第一节 时间序列的分解	(147)
一、时间序列的构成成分	(147)
二、时间序列分解模型	(148)
三、时间序列长期趋势分析	(148)
四、时间序列季节变动分析	(150)
五、时间序列循环变动分析	(152)
六、时间序列分解预测法	(152)
第二节 指数平滑	(154)
一、单参数(一次)指数平滑	(154)
二、双参数指数平滑	(155)
三、三参数指数平滑	(156)
第三节 ARIMA 模型	(157)
一、平稳时间序列模型	(158)
二、ARIMA 模型	(161)
第四节 时间序列分析的软件实现	(163)
一、建立时间序列趋势模型	(163)
二、时间序列季节变动分析	(165)
三、指数平滑预测	(165)
四、建立 ARIMA 模型	(166)
小结	(166)
思考与练习	(167)
第九章 统计指数	(168)
第一节 指数编制的基本原理	(168)
一、统计指数的概念	(168)
二、统计指数的种类	(168)
三、简单指数的编制方法	(169)
四、加权综合指数的编制方法	(170)
五、加权平均指数	(172)
第二节 常用的统计指数及其应用	(175)

一、居民消费价格指数	(175)
二、生产指数	(179)
三、股票价格指数	(181)
第三节 指数体系与因素分析	(183)
一、指数体系	(183)
二、总值变动的因素分析	(184)
小结	(185)
思考与练习	(185)
第十章 主成分分析与因子分析	(188)
第一节 主成分分析	(188)
一、主成分分析的基本思想	(188)
二、主成分分析的步骤和结果分析	(189)
第二节 因子分析	(193)
一、因子分析的基本思想和数学模型	(193)
二、因子分析的步骤和结果分析	(194)
第三节 主成分与因子分析的软件实现	(198)
小结	(201)
思考与练习	(201)
第十一章 聚类分析与判别分析	(203)
第一节 聚类分析	(203)
一、聚类分析的基本思想	(203)
二、符号说明	(203)
三、相似性度量	(204)
四、系统聚类法	(207)
五、K-均值聚类分析	(208)
第二节 判别分析	(210)
一、判别分析的基本思想	(210)
二、距离判别法	(210)
三、Fisher 判别法	(212)
第三节 聚类分析与判别分析的软件实现	(214)
一、聚类分析的 SPSS 实现	(214)
二、判别分析的 SPSS 实现	(218)
小结	(221)
思考与练习	(222)
第十二章 列联表和对应分析	(224)
第一节 列联表与独立性检验	(224)



第二节 对应分析	(226)
一、对应分析的基本思想	(226)
二、对应分析的软件操作和结果解释	(227)
小结	(229)
思考与练习	(229)



绪 论

若想了解上帝在想什么,我们就必须学统计,因为统计学就是在测量他的旨意。

——弗洛伦斯·南丁格尔

第一节 统计与数据

一、什么是统计学

关于统计学的比较权威的定义来自《不列颠百科全书》,用一句话可概括为:统计学是关于收集和分析数据的科学和艺术。

统计学具有如下显著性质:

1. 统计学是研究数据的科学。统计学所研究的数据不是抽象的数和形,而是客观世界与现实生活中实际发生的数据,具体是指那些会经常发生变化的数据,它们往往是和偶然现象联系在一起的,或称随机而发生变化的数据。如经济波动、股市行情、降水概率、彩票中奖号码等。

2. 统计学是以归纳推理为研究方法的科学。统计学是从样本证据出发,利用归纳推理的思想得出更一般的结论,从随机现象中寻求事物的本质,从个别得出一般的结论。

3. 统计学的研究结论带有不确定性。经典决定论给世界描绘出一张精美的具有因果关系的蓝图,19世纪,人们能够看到,世界可能是有规则的,但不服从自然的普通定律,随着统计学的出现和普及,哲人和世人都发现传统的决定论受到了侵蚀,一些新的定律往往是根据概率来表述的,所以统计学的研究成果往往带有不确定性,往往通过统计平均来描述客观世界。

4. 统计学的研究往往伴随新思想的产生。传统的演绎推理结论蕴含在前提之中,而以归纳推理为主的统计学结论具有发散性,但我们可以给出每一种结论对应的概率。当一个新的现象发生时,传统学说只能根据已有的理论外壳解释这种新的现象不是什么,而统计学可以说明它可能是什么,所以统计学是科学研究的原动力。

5. 统计学与大部分学科都有着密切联系,具有广阔的应用领域。随着统计学的发展和科学研究的需求,现在已经很难找到一个不应用统计学的学科领域。归纳目前世界各国统计学的应用,其领域可大略地概括为:精算,农业,动物学,人类学,考古学,审计学,晶体学,人口学,牙医学,生态学,经济计量学,教育学,选举预测和策划,工程,流行病学,金融,水产



渔业研究,遗传学,地理学,地质学,历史研究,人类遗传学,水文学,工业,法律,语言学,文学,劳动力计划,管理科学,市场营销学,医学诊断,气象学,军事科学,核材料安全管理,眼科学,制药学,物理学,政治学,心理学,心理物理学,质量控制,宗教研究,社会学,调查抽样,分类学,气象改善,博彩,……。

二、什么是数据

统计学的研究对象是数据。世界上各种现象无不可用数据来描述。以前我们往往把定性的东西作为理论分析和描述的对象,殊不知,它们也可用数据来表述,如气温、民族、性别、政治观点、满意度、爱情等。随着科学的进步,这些定性描述的现象一方面可以用定性数据来刻画和分析,另一方面迟早会表现为定量数据,如气温最早表现为人的感受:很热、热、……、很冷,现在我们可用温度计来准确的刻画人们的主观感受。我们这个世界是可以数据描述的世界,统计学是描述这个世界的科学。

根据客观世界的表现和人们的认知程度,数据又可分为以下不同类型:

1. 观测数据和试验数据

观测数据是指没有对观测对象进行任何人为因素控制条件下得到的客观数据,如GDP、交通流量、收入、年龄、性别、学历、天体测量、降水概率等。

试验数据是在对影响试验对象的相关因素进行人为控制条件下得到的数据,如减肥前后的体重、临床医学试验观测的人体体征、不同地质状况的农作物产量、产品销售试验等。

2. 定性数据和定量数据

定性数据是指用文字、符号、语言等描述的信息,又可具体分为定类数据和定序数据。定类数据是指对事物按照某种特征进行分类,如性别分为男和女;籍贯分为北京、上海、天津等。定序数据是指在对事物进行顺序上的定性描述,如满意程度描述为很不满意、不满意、……、很满意等;学历描述为高学历、中学历、低学历等。

定量数据是指事物某种特征的具体数量描述,又可具体分为定距数据和定比数据,如GDP、固定资产投资、收入、年龄等。其中定距数据的特点是,两个数值之间可以计算差值,但它们的比值没有实际意义,常见的例子有摄氏温度、海拔高度、年份等。在定距数据中零点的选定只是为了方便或惯例,而不是通常所指的自然数或固定的零点。也就是说,零只是坐标上的一个点,并不表示该现象没有或不存在。例如,当某天的气温为 0°C 时并不是说这天就没有温度。定比数据则有绝对零点,并且两个数字间的比值具有实际意义,常见的例子有身高、体重等等。在实际应用中定距和定比数据常常不做区分而作为一个类别(定量数据)来使用。

3. 截面数据、时序数据和面板数据

截面数据是对多个事物在同一时期或时点上进行测量得到的数据,反映事物在相同时间或时点上的表现,如2008年年末我国各地区的GDP、某年全国各高等院校在校学生数等。

时序数据是对某个事物在不同时期或时点上进行测量得到的数据,反映该事物随时间



变化的情况,如改革开放以来我国历年的 GDP、恢复高考以来全国高校历年的大学生在校大学生数等。

面板数据是对多个事物在不同时期或时点上进行测量得到的数据,反映多个事物随时间变化的情况,如 1990—2008 年全国 30 个省份的 GDP、1990—2008 年全国各高等院校在校大学生数等。对于面板数据,如果只考虑某一时期或时点的情况,就是截面数据;如果只考虑某一事物的情况,就是时序数据。

第二节 统计学中的基本概念

在统计应用中,统计方法可以分为描述统计方法和推断统计方法两大类。描述统计指的是用表格、图形和数字来概括、显示数据特征的统计方法。推断统计指的是从总体中抽取样本,并利用样本数据来推断总体特征的统计方法。推断统计的任务一般来说就是用样本统计量来推断总体参数。推断统计的方法非常丰富,一般以概率论与数理统计为理论基础,是现代统计学的主体内容。这一节我们先来介绍一下描述统计和推断统计中涉及的基本概念。

一、确定性和随机性

从中学起,我们就知道自然科学的许多定律,例如物理中的牛顿三定律,物质不灭定律以及化学中的各种定律等等。但是在许多领域,很难用如此确定的公式或论述来描述一些现象。比如,人的寿命是很难预先确定的。一个吸烟、喝酒、不锻炼的人可能比一个很少得病、生活习惯良好的人活得长。

因此,可以说,活得长短是有一定随机性的。这种随机性可能和人的经历、基因、习惯等无数说不清的因素都有关系。但是从总体来说,我国公民的平均年龄却是非常稳定的。而且女性的平均年龄也稳定地比男性高几年。这就是规律性。一个人可能活过这个平均年龄,也可能活不到这个平均年龄,这是随机的。但是总体来说,平均年龄的稳定性,又说明了随机之中有规律性。这种规律就是统计规律。

对于有随机性的事件,我们常常用概率来描述其发生的机会。例如在天气预报中会提到降水概率。大家都明白,如果降水概率是百分之九十,那就很可能下雨;但如果是百分之十,就不大可能下雨。显然,这种概率不可能超过百分之百,也不可能少于百分之零。换言之,概率是在 0 和 1 之间的一个数,说明某事件发生的机会有多大。

二、总体和样本

总体是我们所研究的全部同类对象(事物)组成的集合。总体中的一个对象称为个体。比如,我们要研究某大学学生的体能状况,那么,该大学所有学生构成研究的总体,每个学生都是一个个体;我们想了解一批产品的质量,这批产品就是研究的总体,每一件产品都是一个个体;我们想了解某地区居民家庭消费状况,该地区的所有居民户构成研究的总体,每一个居民户都是研究的个体。这里所说的同类事物是指所研究的具有一定同质性的多个事物,比如上面说到的大学生、居民户等,我们不能把产品与居民户放到一起作为一个总体来研究。

根据总体中包含个体的数目是否有限,可将总体分为有限总体和无限总体两类。有限

总体是指总体范围可以明确界定,个体数目有限的总体。比如,上面提到的三个总体都是有限总体。无限总体是指个体数目无限的总体。比如,在某一项科学试验中,所有进行的试验构成研究的总体,每一次试验可以看作一个个体,由于试验可以一直进行下去,试验次数是无限的,该总体是一个无限总体;我们想了解连续生产的某电子器件的质量,所有电子器件构成研究的总体,每一个电子器件是研究的个体,由于生产可以一直进行,个体数量是无限的,该总体也是一个无限总体。将总体区分为有限总体和无限总体,其意义在于不同类型的总体可以采取不同的数据获取和数据分析方法。现实中的大多数总体都是有限总体。

对于总体而言,我们总是希望了解其某些方面的特征。但是总体中个体的数量往往是巨大的,对每一个个体都进行考察和了解显然不现实,对无限总体来说这也是不可能的。实际中采取的方法大多是从总体中抽取一部分个体进行研究,据此对总体特征进行推断。这就需要引入有关样本的概念。

样本是我们所研究的全部同类对象中部分对象组成的集合,或者说,样本是部分个体组成的集合,是总体的一部分。样本中所包含个体的数目称为样本容量。比如,从某地区的所有居民户构成的总体中随机抽出 100 户居民,这 100 个居民户构成一个样本,样本容量为 100。可由此 100 户居民家庭消费状况推断该地区居民家庭的消费状况。

三、参数和统计量

参数与统计量是分别对应于总体和样本的两个概念。

参数是描述总体数量特征的一种概括性数字度量,是研究者想要了解的总体的某种特征。比如,某大学学生的平均身高与身高的标准差,某批产品的合格率,某地区某月所有居民户的平均消费额,某地区某月居民户的消费额与收入的相关系数等。由于总体数据通常是不知道的,所以这些反映总体某方面数量特征的参数就是一个未知的常数。正因为如此,我们才需要从总体中抽取样本,对总体参数进行估计和推断。

统计量是根据样本数据计算出来的描述样本某方面数量特征的一种概括性数字度量,它是样本的函数。比如,从某大学随机抽取的 100 名学生,得到一个样本,由其身高数据可以计算出样本平均数和样本标准差;从一批产品中随机抽出 50 件产品进行质量检测,就可以得到样本合格率;从某地区抽取 100 个居民户,由其消费额与收入数据就可以计算消费额与收入的样本相关系数。由于样本是从总体中抽取出来的,所以统计量总是知道的,这样我们就可以根据样本统计量去估计总体参数。

为了区别和叙述方便,总体参数通常用希腊字母表示,比如,总体均值 μ , 总体标准差 σ , 总体比例 π ; 样本统计量通常用英文字母表示,比如,样本均值 \bar{x} , 样本标准差 s , 样本比例 p 等。

除了上面提到的用于估计总体参数的统计量外,还有一类常用的统计量,这类统计量是为了进行统计检验而构造出来的,而且作为样本的函数形式上往往较为复杂,比如 t 统计量, F 统计量, χ^2 统计量等。

四、变量和指标

1. 变量

变量是统计学中一个非常重要和常用的概念。前面提到,总体是所研究的全部同类



(同质性)对象(事物)组成的集合。变量就是说明这些同类事物某方面特征的名称。由于同类事物某方面特征总会表现出变化和差异,因此也就把这种特征的名称称为变量。变量的具体取值称为变量值。比如,表1-1为某班学生情况统计表,这里的“姓名”、“学号”、“性别”、“思想表现”和“成绩”都是说明该班学生某方面特征的名称,因此,也就是变量。表中的内容,比如“男”、“优秀”、“380”等就是变量值。

表 1-1 某班学生情况统计表

姓 名	学 号	性 别	思想表现	成 绩
张三	01	男	良好	380
李四	02	男	优秀	340
王五	03	女	优秀	385
...

事实上,变量值是从不同方面对研究对象进行测量的结果,也就是统计数据。前面提到,按照测量的尺度与数据层次的不同,可将统计数据分为定类数据、定序数据和数值数据。相应的,我们也可以根据变量取值(数据)的类型把变量分为定类变量、定序变量和数值变量。说明事物类别的名称,也就是取值为定类数据的变量,称为定类变量;说明事物有序类别的名称,也就是取值为定序数据的变量,称为定序变量;说明事物数字特征的名称,也就是取值为数值数据的变量,称为数值变量。表1-1中的“性别”、“思想表现”和“成绩”分别是定类变量、定序变量和数值变量。在这里“姓名”和“学号”只是研究对象的标记,但一般不参与统计分析,所以没有必要区分其类型。

现实中遇到的变量,多数是数值变量,大多数的统计方法也是处理和分析数值变量。根据数值变量的取值情况不同可以将数值变量分为离散数值变量和连续数值变量,简称为离散变量和连续变量。

离散变量是只能取有限个数值并且取值为整数的变量。比如,每分钟通过十字路口的车辆数,每隔1小时抽取的100件产品中不合格产品的数量,获得奖励的次数等,这些变量可能的取值是有限个整数,它们都是离散变量。

连续变量是可以在某区间中连续取值的变量。比如,年龄、体重、温度、销售额等都是连续变量。对于年龄这个变量,由于可以连续取值,因此理论上说是连续变量。但是实际中我们说某人的年龄习惯于用整数,因此也可以当作离散变量处理。

在一些统计方法中也将变量称为元,比如多元统计分析,多元回归分析等,其中的元就是变量。

2. 统计指标

在实际工作中常常会使用统计指标的概念。统计指标一般有两种理解和使用方法:(1)指反映现象数量特征的概念,例如年末人口数、商品销售额、劳动生产率等。(2)指反映现象数量特征的概念和具体数值,例如我国2008年的国内生产总值为300670亿元。

统计指标按其表现形式可分为总量指标、相对指标和平均指标。

总量指标也称为绝对数,指以绝对数形式表现现象规模和水平的统计指标。例如,



2008 全年入境旅游人数 13003 万人次;全年国内生产总值 300670 亿元;2008 年末全国参加城镇基本养老保险人数为 21891 万人等等。绝对数可以分为时点数和时期数。时点数是描述某种现象在某一个特定时刻(某一瞬间或某一时点)数量表现的数据,例如,2008 年年末全国总人口为 132802 万人。时期数是描述某种现象在某一个特定时间范围内所实现的成果的数据。例如,2008 年我国全年各种运输方式完成货物运输周转量 110301 亿吨公里。区分数据是时点数还是时期数的方法之一是看其加总后的结果是否有意义。若有意义则该指标必定是时期数。反之,则必定是时点数。

相对指标是采用两个有联系的绝对数进行对比而得到的比值。也称为相对数,如产业结构比例、性别比、人口密度等等。平均指标也称为平均数,反映现象在某一时间或空间上的平均数量水平。例如职工的平均工资,平均考试成绩,等等。

第三节 常用统计软件简介

20 世纪 40 年代以来,随着计算机技术的不断发展,专门用于数据收集和分析的统计软件也逐渐兴起和发展。功能强大的现代统计软件不仅可以替我们完成繁琐复杂的计算,还可以帮助我们全方位地分析与展示数据,充分挖掘出数据中蕴含的规律。

统计软件的发展使非专业人员应用复杂的统计方法成为可能,极大地促进了统计技术在各个领域的广泛应用,也反过来推动了统计科学的发展。可以说,现代统计学和统计软件已经密不可分。

一、几种常用的综合统计软件

常用的统计分析软件包括 SAS、SPSS、S-plus、R、Stata、Minitab 等等。这些软件都能完成常用的统计方法,如描述统计、参数估计、假设检验、方差分析、回归分析、多元分析等,但不同的软件在功能、易用性、扩展性等方面又各具特色,下面我们分别加以简要介绍。

1. SAS

SAS 过去是“Statistical Analysis System”的简称,由于其功能现已远远超出了统计分析的范围,“SAS”已经变成了一个单纯的商标。SAS 系统具有非常强大的数据分析能力,可以同时处理多个数据集,有很多模块,功能非常全面;在数据分析和统计分析领域被誉为国际上的标准软件和最权威的优秀统计软件包,广泛应用于政府行政管理、科研、教育、生产和金融等不同领域。

SAS 软件虽然也提供了许多菜单操作方式,但仍以编程为主,学习起来有一定困难,是最难掌握的统计软件之一。

2. SPSS

SPSS 最早是“Statistical Package for Social Sciences”的简称,它也是最早的统计软件之一。该软件从 17.0 版本开始更名为“SPSS Statistics”,2009 年 IBM 公司收购 SPSS 公司后,目前的最新版本开始称为 IBM SPSS Statistics。SPSS 是一个组合式软件包,集数据处理、数据分析、统计图表生成及统计编程等诸多功能于一身,涵盖了统计学的所有常用的统