

现实世界中的 统计学

朱元兰 陈海梁 / 著



■ 贵州民族出版社

图书在版编目 (C I P) 数据

现实世界中的统计学/朱元兰,陈海梁著. —贵阳:
贵州民族出版社,2006.9
ISBN 7-5412-1403-5/C·95

I. 现... II. ①朱...②陈... III. 统计学
IV. C8

中国版本图书馆 CIP 数据核字(2006)第 116425 号

作 品 名	现实世界中的统计学
作 者	朱元兰 陈海梁 著
出版发行	贵州民族出版社
地 址	贵州省贵阳市中华北路 289 号
邮 编	550001
责任编辑	郭堂亮
封面设计	张 超
责任校对	曹永兰
印 刷	贵阳经纬印刷厂
开 本	850×1 168 毫米 1/32
字 数	230 千字
印 张	8.25
版 次	2006 年 9 月第 1 版
印 次	2006 年 9 月第 1 次印刷
印 数	1~1 000 册
定 价	22.00 元

责任编辑：郭堂亮

封面设计：张超

责任校对：曹永兰

内容简介：

统计是人们认识世界的一种方法。本书着重介绍统计方法在现实世界中的应用。内容包括统计资料的收集、数据合理性的判断、数据的显示以及数据关系描述、居民消费价格指数的计算及其所指的意义、概率与概率模型和统计推论等等。

前言

随着人类社会发展的不断信息化,如何测度经济活动、社会活动,已经成为社会发展共同了解、科学分析、活动规划、科学决策的重要基础。在对社会科学与自然科学广泛、深入的研究中,也越来越依赖科学、系统的信息测量,以及应用科学的量化分析方法,获得最前沿的科学成就,如社会学研究、政治科学研究、教育研究等。而这些都是需要统计学的支持。统计的概念对日常生活、公共政策和许多领域都产生了很大的影响。统计学已成为认识社会发展规律、经济规律以及认识各门科学的重要工具和手段。如医疗领域的专家必须有统计知识,才能了解医学研究报告的内容;经理人需要统计,因为大量的数据需要经过有效浓缩,才能看得出所以然;老百姓需要了解统计,才能知道居民消费价格指数到底是怎么回事。

统计为人们认识社会经济现象提供了一种独立且基本的思考方法。本书编写的宗旨偏重

于介绍常用的统计方法以及统计方法在现实世界中的应用,比如在什么情况下最适宜采用这个方法;怎样进行运算及算出的结果所说明的意义等等,而不是统计方法的数理基础,避免了复杂的数学推算,力求简单明了。适合于医疗专家、经理、老百姓等对统计学方法的掌握。

本书在介绍统计方法时,注重以下几个问题:(1)以统计新思想发展为纲,强调如何做统计思考;(2)本书的内容是多一些数据和观念,少一点推导过程;(3)它的内涵尽量符合素质教育的精神,有很多观念和很多思考。

本书力求简明、准确、实用,联系国内外最先进的理论和方法,并且与中国社会发展实际相结合,在社会统计分析当中,密切联系社会的变迁和社会的发展,全面了解国内外社会发展的各种趋势,把握国内外社会研究的理论前沿。如果通过对本书的学习,能够让您学会用手中的数据来分析事物,并能对重大问题做出比较明智的决策的话,那么编写它的目的就达到了。

本书在写作过程中受到了社会各界的广泛支持和鼓励,并参阅了国内外的有关书籍与资料,在这里表示感谢。这里要特别感谢贵州省统计局机关党委书记张北平同志、贵州民族报社总编辑农文成同志、贵州省工商银行花溪分校校长陈美鸿同志的大力支持和热情的帮助,感谢贵州财经学院数学与统计学院领导与陈森良教授的关心,感谢贵州财经学院统计学重点学科组的支持。本书的出版还得到了相关人员

前 言

的支持,为本书的修改、编辑、排版等做了大量工作,付出了辛勤的劳动,在此一并表示衷心的感谢。

应当指出,尽管为提高本书的质量作了很大的努力,但由于水平有限,加之时间仓促,书中难免有疏漏或错误之处,恳请同行和读者不吝赐教,以便今后有机会进一步修改与完善。

朱元兰

2006年7月于贵阳

目 录

前言	(1)
第一章 数据的收集	(1)
一、总体、个体与变量	(1)
(一) 总体	(1)
(二) 个体	(2)
(三) 变量	(2)
(四) 数据	(6)
二、观测研究	(8)
三、普查	(10)
四、抽样调查	(10)
(一) 总体与样本	(11)
(二) 样本的选择	(12)
(三) 简单随机样本	(13)
(四) 样本信息	(16)
(五) 现实世界中的抽样调查	(24)
五、实验	(31)
(一) 实验	(31)
(二) 现实世界中的实验	(33)
第二章 度量和数据合理性的判断	(40)
一、度量的基本原理	(40)

二、有效量度和无效量度·····	(41)
三、准确和不准确量度·····	(43)
四、数据合理性的判断·····	(45)
第三章 数据的显示 ·····	(49)
一、数据表·····	(49)
二、数据图·····	(54)
(一)饼状图·····	(55)
(二)柱状图·····	(56)
(三)线图·····	(58)
(四)直方图·····	(60)
(五)茎叶图·····	(70)
第四章 简单数据关系描述 ·····	(73)
一、基本技术·····	(73)
(一)次数分布·····	(73)
(二)比重(相对频数)·····	(74)
(三)比率·····	(75)
(四)对比值(比例)·····	(76)
(五)累加(计)次数和累加(计)百分率·····	(76)
二、集中趋势·····	(78)
(一)众数(M_o)·····	(78)
(二)中位数(M_e)或者中位值·····	(80)
(三)四分位数·····	(83)
(四)平均数(均值)·····	(85)
三、离散趋势·····	(88)
(一)极差·····	(88)
(二)四分位差·····	(89)

(三)标准差(s)	(92)
(四)变异系数	(96)
第五章 正态分布	(98)
一、密度曲线	(100)
二、密度曲线的中心和离散程度	(102)
三、正态分布	(103)
四、68—95—99.7 规则	(105)
五、标准计分	(108)
第六章 两个变量关系描述	(112)
一、散布图	(113)
二、相关系数	(117)
三、交互分类	(122)
四、消减误差	(126)
第七章 相关关系描述	(129)
一、回归直线	(130)
(一)回归直线	(130)
(二)回归方程式	(131)
(三)预测的意义	(135)
二、相关系数、回归及因果问题	(138)
(一)相关系数和回归	(138)
(二)因果问题	(140)
三、其他相关测量方法	(143)
(一)Lambda 相关测量法	(143)
(二)tau-y 相关测量法	(147)

第八章 居民消费价格指数	(151)
一、指数	(152)
二、居民消费价格指数	(153)
(一)消费品分类和代表规格品的选择.....	(153)
(二)基本分类商品价格指数的计算.....	(153)
(三)计算中类指数.....	(154)
(四)计算大类指数.....	(155)
(五)计算总指数.....	(155)
三、固定权数物价指数	(156)
四、居民消费价格指数的现实意义	(157)
(一)如何应用居民消费价格指数.....	(157)
(二)了解居民消费价格指数的意义.....	(161)
第九章 概率与概率模型	(164)
一、概率	(164)
(一)概率.....	(165)
(二)随机与概率.....	(167)
(三)意外的机遇.....	(170)
(四)平均数定律.....	(171)
(五)个人概率.....	(172)
二、概率模型	(174)
(一)概率规则.....	(174)
(二)抽样的概率模型.....	(178)
(三)模拟.....	(182)
三、期望值	(191)
(一)期望值.....	(193)
(二)大数法则.....	(194)

第十章 置信区间	(198)
一、95%置信区间	(199)
二、了解置信区间	(202)
三、总体比率的置信区间	(204)
第十一章 显著性检验	(208)
一、统计检验的理论基础	(208)
(一)断言	(209)
(二)抽样分布	(209)
(三)数据	(209)
(四)概率	(210)
二、假设及 p 值	(211)
(一)原假设 H_0	(211)
(二)p 值	(212)
三、统计显著性	(215)
第十二章 统计推论的使用	(218)
一、明智地做推论	(218)
(一)产生数据的设计很重要	(218)
(二)了解置信区间的运作	(219)
(三)了解统计显著性的意义	(219)
(四)了解你用的方法必须满足的条件	(220)
二、显著性检验面临的困难	(220)
三、置信区间的优点	(224)
第十三章 双向表及卡方检验	(227)
一、双向表	(228)
(一)双向表	(228)

(二)双向表的推论·····	(230)
二、卡方检验·····	(233)
(一)卡方检验·····	(233)
(二)卡方检验的应用·····	(237)
第十四章 总体平均数的推断 ·····	(241)
一、样本平均数的抽样分布·····	(242)
二、总体平均数的置信区间·····	(243)
三、总体平均数的检验·····	(245)

第一章 数据的收集

统计数据是我们利用统计方法进行分析的基础,离开了统计数据,统计方法就成了无米之炊,失去了用武之地,那么从哪里取得所需要的统计数据呢?或者,我们经常从报纸、电视上看到或听到很多的统计数据,比如说某月的居民消费价格指数是 103.5%,或今年的粮食产量与去年比增加了 2 成,或低收入的家庭得到了政府良好的照顾收入增加了,生活水平提高了等等,这些数据是从哪里来的呢?我们该不该相信这些数据?统计是数据的科学,在取得统计数据时,除了要有好的判断外,还应该有良好的数学,这样才能做好统计。好的判断中有一部分决定于你要度量什么,这样产生的数据才是你所关心的问题。好的数据是我们对社会进行分析得出正确结论的前提,草率的数据就像粗糙的产品一样,令人沮丧。所以这一章就首先介绍一些基本概念,它们是用来描述与组成数据的原始材料,然后介绍数据的收集及好坏数据的分辨方法。

一、总体、个体与变量

“数字是统计的语言”,“数据是统计的原料”,数据如何产生是我们所关心的问题。在介绍数据的产生以前,我们得先介绍一些名词,它们就是用来描述、组成数据的原始材料。

(一) 总体

统计的数据研究中所研究的事物的全体也就是所研究的对象

就是总体。它是由所有的个体组成。例如进行全国城镇居民家庭调查,全国城镇所有居民家庭就是总体。例如要研究某一个国家超过 25 岁已婚司机的情况,那么这个国家所有超过 25 岁的已婚司机就是一个总体。总体可以是我们所定义的任何事物,理解这一点很重要。例如过去一年中在某网站购物的所有消费者或者所有未拥有私人电话的个人也可以是总体。如果这个网站保存着有关的记录,那么就很容易找到顾客购买情况的全部数据,但是确认所有未拥有私人电话的个人也许不大可能。总体可以是现实世界中存在的任何事物的集合,如某国家的某一个球队,或者是潜在的但又未知的某个流程的输出,如某一装配线的汽车引擎,等等。

(二)个体

统计的数据研究是对现象总体中各单位普遍存在的事实进行大量观察和综合分析,进而得出反映现象总体的数量特征。而现象总体中的各单位就是总体单位,也称为个体。例如,进行城镇居民家庭调查,城镇所有居民家庭是总体,而每一个居民家庭就是个体,通过对每个居民家庭情况的了解可以知道这个城镇居民家庭的整体特征。研究某一个国家超过 25 岁已婚司机的情况,这个国家所有超过 25 岁的已婚司机就是一个总体,而这个国家中每一个超过 25 岁的已婚司机是个体,通过收集每个超过 25 岁的已婚司机的数据可以掌握所有超过 25 岁的已婚司机的情况。具体来说,个体就是数据所描绘的对象。个体可能是人、物、机构等实物单位,也可以是一种现象或活动过程等非实物单位,甚至可以是动物或其他东西。

(三)变量

1. 变量的概念

在统计中,一般把说明一个个体的某种特征名称称为变量。同一个变量对于不同的个体,可能有不同的值或者不同的表现。变量的不同的值或者不同的表现称为变量值。我们所说的数据就

是变量的具体表现即变量值。举例来说明,以下是某高校某一个学期某门课程结束后,某班学生成绩登记册资料,如表 1.1。

表 1.1

姓 名	成 绩
李文化	90
周 玲	85
赵历历	78
廖 建	63
张 宏	52
.....

在这个例子中,总体是某一个班的全体学生,个体就是这个班的每一个学生。在表中,除了学生的姓名之外,这里有一个变量,这个变量是成绩,100分为满分,表中的90、85、78、63、52等等都是变量的不同的表现即变量值。可以看出每个学生都有不同的成绩。

2. 变量的类别

(1) 品质变量与数量变量

如果一个变量由文字来记录就称为品质变量,如“性别”就是一个品质变量,它表现为“男”或“女”;“产品等级”也是一个品质变量,它表现为“一等品”、“二等品”、“三等品”、“次品”等。如果一个变量由数据来记录就称为数量变量或数字变量、数据变量,如“年龄”、“分数”、“产品数量”、“商品销售额”等都是数据变量,它们可以表现为不同的数值。比如表 1.1 中的成绩就是数量变量。现实世界中,多数情况下我们所说的变量是指数字变量,大多数统计方法所处理的也都是数字变量,因此有时把数字变量就简称为变量。

(2) 离散变量与连续变量

数字变量根据其取值的不同,可以分为离散变量和连续变量。离散变量只能取有限个值,而且其取值都以整位数断开,可以一一列举,如“人数”、“企业数”、“产品数量”等就是离散变量;连续变量可以取无穷多个值,其取值是连续不断的,不能一一列举,如“温度”、“年龄”、“零件尺寸”等都是连续变量。

(3)单变量与多变量

根据研究现象包含的变量的数目分为单变量和多变量。单变量是所研究的现象中只有一个变量,如果研究的现象中有两个或者两个以上的变量就是多变量。单变量的研究比较简单,只需要计算比率、集中趋势和离散趋势来反映其分布,对于多变量的研究较复杂。这些在后面都有介绍。

(4)定类层次、定序层次、定距层次与定比层次的变量

根据变量由低到高、由粗略到精确分为定类层次、定序层次、定距层次与定比层次。定类层次的变量是最低的,它是根据变量的某种属性进行分类。它的取值只有类别属性之分,而无大小、程度或者优劣之分。根据变量值,只能知道现象是相同或是不同。如性别、民族等,从数学运算的性质来看,定序层次的变量只具有等于或者不等于。

定序层次的变量高于定类层次,它的取值除了有类别属性之外,还有等级、次序的差别。其数学运算性质除了具有等于或者不等于外,还有大于或小于之分。常见的定序层次的变量如教育程度有文盲、小学、初中、高中、大学;社会经济地位有上等、中等、下等;重要性有非常重要、比较重要、一般重要、不太重要、不重要等;学生的成绩有优、良、中、及格与不及格等。从数学运算的性质来看,定序层次的变量的主要特征是大于或者小于。

定距层次的变量的层次又高于定序层次。定距层次的取值,除了类别、次序属性外,取值之间的距离还可以用标准化的距离去量度它。其数学运算性质除了等于或不等于、大于或小于外,还可