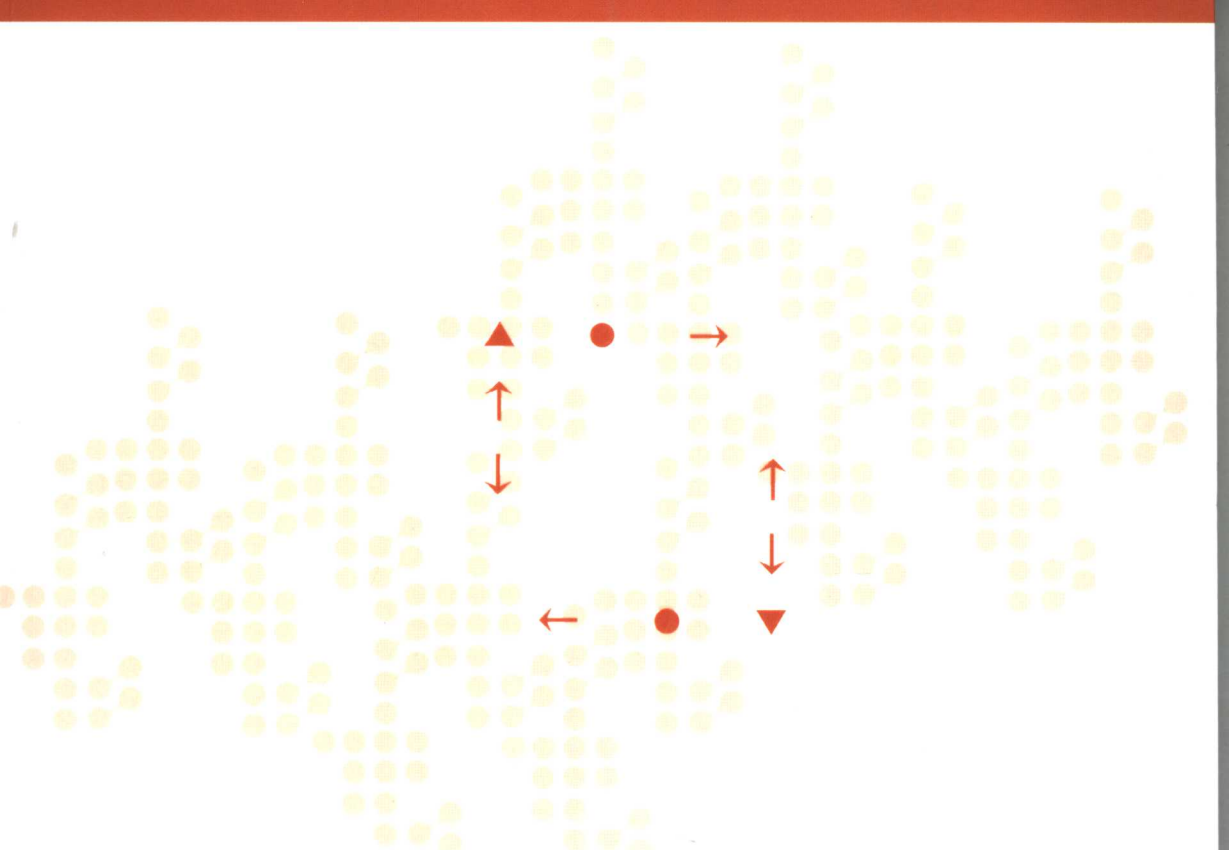




统计学

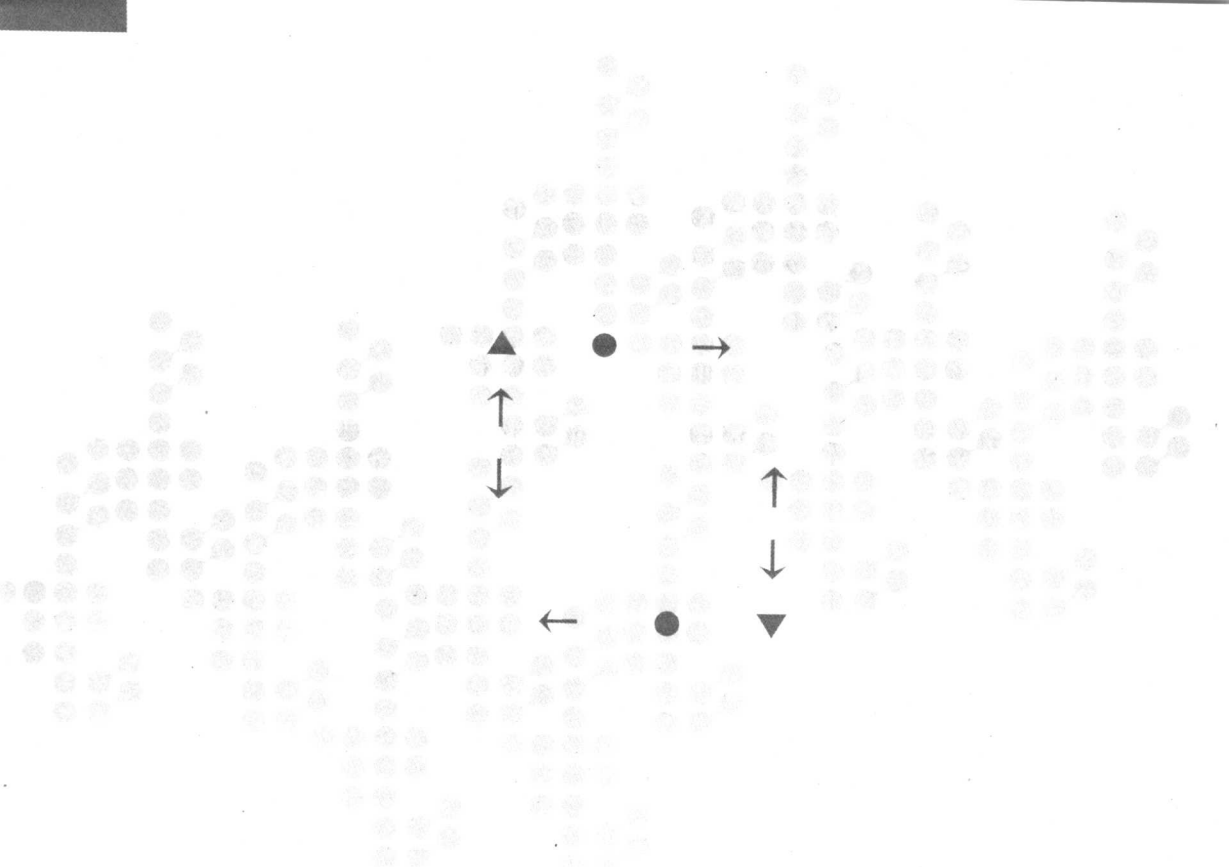
社会统计学与数理统计学的统一

王见定◎著 **statistics**

SHE HUI TONG JI XUE
YU SHU LI TONG JI XUE
DE TONG YI



中国经济出版社
CHINA ECONOMIC PUBLISHING HOUSE



统计学

社会统计学与数理统计学的统一

王见定◎著 **statistics**

SHE HUI TONG JI XUE
YU SHU LI TONG JI XUE
DE TONG YI



中国经济出版社
CHINA ECONOMIC PUBLISHING HOUSE

北 京

前 言

统计学如同计算机一样,广泛地应用于各个行业与专业。长期以来,社会统计学与数理统计学的严重分离,使得本来描述同样社会与自然现象的科学人为地隔裂开业(国际上也是这样)。

本书是一本将社会统计学与数理统计学有机地结合在一起的著作,这是一种大胆的尝试。笔者采用了极其通俗、极其简明的方法,将统计学的基本原理、基本方法,其中包括笔者多年来在统计学方面的工作介绍给大家,希望对各行业、各专业的学者均有所受益。尤其对各级主管经济的同志,本书更是必读之物。

在本书的撰写过程中,笔者参阅了许多著作,尤其对袁卫等编著的《统计学》的部分章节进行了整改而加以采用,在此表示感谢。对张玲玲编辑付出的劳动表示感谢。

最后,希望该书的出版能对我国经济的高速发展作出一定的贡献。

著者

2006年9月于北京

目 录

第一章 数据的搜集与整理	1
第一节 数据的计量与类型	1
第二节 数据的搜集	4
第三节 数据的整理	9
第二章 数据的描述	14
第一节 绝对数与相对数	14
第二节 平均水平的描述	16
第三节 分散程度的描述	21
第三章 时间数列分析	24
第一节 时间数列的基础分析	24
第二节 长期趋势的测定	27
第三节 季节变动的测定	32
第四节 预测方程的建立	33
第四章 指数分析	35
第一节 指数及其分类	35
第二节 加权指数	36
第三节 指数体系	40
第四节 几种重要的指数	45

第五章 统计量的分布	52
第一节 总体与样本	52
第二节 统计量的分布	53
第六章 参数估计	55
第一节 点估计	55
第二节 估计量好坏的标准	58
第三节 区间估计	61
第七章 假设检验	68
第一节 假设检验问题的提出	68
第二节 假设检验的程序	69
第三节 一个正态总体的假设检验	71
第四节 两个正态总体的假设检验	76
第五节 概率的假设检验	79
第六节 分布函数的假设检验	80
第七节 假设检验的两类错误	85
第八章 非参数统计推断	86
第一节 次序统计量和秩统计量	86
第二节 估计问题	89
第三节 检验问题	93
第九章 回归分析	95
第一节 一元线性回归	95
第二节 多元线性回归	97
第十章 相关分析	100
第一节 样本相关系数	100
第二节 样本复相关系数	102

第十一章 方差分析	104
第一节 单因素方差分析	104
第二节 双因素方差分析	107
第十二章 判别分析	110
第十三章 国民经济核算	114
第一节 国民经济核算的对象	114
第二节 国民经济核算的分类	116
第三节 国民经济核算的内容框架	118
第十四章 国民经济总量统计	126
第一节 生产范围和生产总量统计	126
第二节 国内生产总值的核算方法	128
第三节 收入总量与财产总量	132
第四节 经济增长率	135
附录一	
标准正态分布函数表	137
χ^2 分布上分位数表	142
t 分布上侧分位数表	146
F 分布上分位数表($\alpha=0.01$)	148
F 分布上分位数表($\alpha=0.05$)	151
F 分布上分位数表($\alpha=0.1$)	154
附录二	
国民经济行业排列的有序化与经济矢量	157
附录三	
经济矢量的合成和资源的有效配置	160
参考文献	163

第一章 数据的搜集与整理

统计学是搜集、整理、分析数据的一门科学,可见数据是统计学的基础,离开了数据,统计方法就无从谈起。那么数据从何而来,在取得数据以后又该怎样整理呢?下面就此问题展开叙述。

第一节 数据的计量与类型

一、数据的计量尺度

按照我们对客观事物计量的精确水平,可将所采用的计量尺度由低级到高级、由粗略到精确分为四个层次:定类尺度、定序尺度、定距尺度和定比尺度。

1. 定类尺度

定类尺度,也称为列名尺度,它是最粗略、层次最低的计量尺度,它只能按照某种属性对客观事物进行平行的分类。例如:按



照性别将人口分为男、女两类;按所有制性质将企业分为国有、民营、外企等等。定类尺度只是计量了事物之间的类别差,各类别之间是平等的并列关系,无法区分优、劣、大、小。使用定类尺度区别事物类别必须符合穷尽和互斥的原则,即:在所区分的全部类别中,必须保证每一个个体都能归属某一类别,且只能属于一个类别。例如,一个人要么是男性,要么是女性,且只能属于其中一个类别。

2. 定序尺度

定序尺度也称为顺序尺度,它是对事物之间的顺序差别进行计量。例如:产品等级就是对产品质量好坏进行的一种次序计量,它将产品分为一等品、二等品、三等品、次品;考试成绩可以分为优、良、中、差,等等。很显然,定序尺度对事物的计量要比定类尺度精确一些,它给出了类别之间的顺序,但未给出类别之间的准确差值,不能进行加、减、乘、除等最基本的数学运算。

3. 定距尺度

定距尺度也称为间隔尺度,它不仅能将事物区分为不同类型,并进行排序,而且能准确地计量出它们之间的差距是多少。例如,考试分数用百分制计量、温度用摄氏度计量、长度用米计量等等。因此,只要给出一个计量单位,就可以准确地指出两个计数之间的差别。如:考试成绩 70 分与 90 分之间相差 20 分;一个地区的温度比另一个地区的温度高 10°C , 等等。也就是说,定距尺度的数据可以进行加、减运算。



4. 定比尺度

定比尺度也称为比率尺度,它除了具有上述尺度的全部特性外,还具有另一个特性,即可以计算两个计量值之间的比值。这就要求在定比尺度中必须有一个绝对固定的零点。这也是它与定距尺度的唯一区别。例如,甲月收入为6000元,乙为3000元,可以得出甲的收入是乙的两倍。可见,定比尺度的数据除了作加、减运算外,还能进行乘、除运算。(当然,定距尺度中也可以出现“0”,但这里的“0”只表示一个“0”水平,不表示不存在。)

二、数据的类型

从上述四种计量的结果来看,我们可以将数据分为定性数据和定量数据。定性数据也称为品质数据,它说明现象的品质特征,它是定类和定序尺度计量的结果。定量数据也称为数量数据,它说明了现象的数量特征,它是定距和定比尺度计量的结果。

在统计学中,说明现象某种特征的概念称为变量,变量的具体表现称为变量值。变量可分为品质变量和数字变量。

例如:“性别”是个品质变量,它表现为“男”或“女”;“产品等级”也是品质变量,它表现为“一等品”、“二等品”、“三等品”、“次品”,等等。而“产品产量”、“商品销售额”、“年龄”、“时间”等都是数字变量。

数字变量根据其取值的不同,又可分为离散变量与连续变量。离散变量的取值是离散的,可以一一列举;而连续变量的取

值是连续的,如“温度”、“体重”等都是连续变量。以后章节中提到的变量,都是数字变量。

第二节 数据的搜集

数据的计量我们清楚了,下面的问题就是数据从何处而来。数据最初都是来源于直接的调查和试验。但从使用者的角度看,数据来源于两种渠道:一是来源于自己的调查和试验;二是来源于别人的调查和试验。一般说来,来源于别人的调查和试验也是相当普遍的。

下面就从使用者的角度谈一谈数据的搜集方法。

一、调查方案

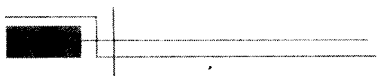
在搜集数据之前,首先要制定一个完整的调查方案,用来指导整个调查工作,使调查能够顺利进行。

一个完整的调查方案应包括以下内容:

1. 调查目的

在调查方案中首先应说明本次调查的目的、任务和意义,所回答的是“为什么调查”、“要解决什么问题”。明确这些问题之后我们才能确定向谁调查、调查什么以及用什么方法进行调查。

例如,1990年的第四次人口普查的目的是“为了准确地查清第三次全国人口普查以来我国人口在数量、地区分布、构成和素



质方面的变化,为科学地制定国民经济和社会发展战略,提高人民的物质和文化生活水平,检查人口政策执行情况,提供可靠的资料”。

2. 调查对象和调查单位

调查对象:根据调查目的确定的调查范围。

调查单位:构成调查对象的每一个单位。

在实际调查中,调查单位可以是调查对象的全部单位,也可以是部分单位。如普查中,调查对象中的每一个单位都是调查单位。如抽样调查中,调查单位只是调查对象的一部分。

3. 调查项目和调查表

调查项目所要回答的是“调查什么”和“采取什么形式调查”。

通常调查项目以表的形式出现,填调查表。调查表一般由表头、表体、表脚三部分组成。表头说明调查表的名称以及被调查单位的名称、性质、隶属关系等;表体包括调查的主体项目;表脚由填报人签名、填报日期等组成。

二、调查方法

常用的调查方法有普查、抽样调查、统计报表等。

1. 普查

普查是为某一特定目的而专门组织的一次性全面调查,如人

口普查、工业普查、农业普查等。世界各国都定期进行各种普查,以便掌握有关国情、国力的基本数据,为国家制定有关政策提供依据。普查的涉及面广、调查单位多,需要耗费大量的人力、物力、财力和时间。

普查具有以下特点:(1)普查通常具有周期性,例如,我国每逢“0”年进行人口普查,每逢“5”年进行工业普查;每逢“7”年进行农业普查;(2)普查需要规定统一的调查时间,以保证普查结果的准确性,避免调查数据的重复和遗漏;(3)普查的数据比较准确,因此可以为其他调查提供基本的参照依据;(4)普查一般只调查一些最基本的现象。

2. 抽样调查

抽样调查是一种应用最广的调查方法,它是从调查对象总体中随机抽取一部分单位作为样本进行调查,并根据抽样调查的结果来推断总体的一种非全面调查方法。

抽样调查有以下特点:(1)经济性。由于调查的样本通常是总体单位中的很小部分,调查的工作量小,所以节省了大量的人力、物力、财力。(2)时效性高。抽样调查可以迅速、及时地获得所需要的信息。由于调查的准备时间、调查时间、数据的整理时间都可以大大缩减,所以提高了数据的时效性。(3)适应面广。抽样调查可获得广泛的信息,它适用于对各个领域、各种问题的调查。从调查的项目或指标来看,可以大大多于全面调查。(4)准确性高。由于抽样调查的工作量小,可使各环节的工作更细

致,误差更小。由于抽样是随机的,它对总体应具有一定的代表性。

由于抽样调查具有以上的特点,因此被世界各国广泛采用。

3. 统计报表

统计报表是我国目前搜集数据的一种主要方式。它是按照国家的有关法规,自上而下的统一布置,又自下而上地逐级提供数据的一种调查方式。统计报表要以一定的原始记录为基础,按照统一的表式、统一的指标、统一的报送时间和程序进行填报。

统计报表类型多样。按调查范围的不同,可分为全面的和非全面的。全面的统计报表要求调查对象的每一个单位都填报;非全面的统计报表只要调查对象中的一部分单位填报。按报送时间的不同,统计报表可分为日报、旬报、月报、季报、半年报和年报等。

按报表内容和报送范围的不同,统计报表可分为国家的、部门的和地方的。国家统计报表从整个国民经济的角度出发,按国民经济的部门划分填报相应的指标,为整个国民经济的宏观管理服务;部门和地方统计报表是按部门和地区的要求填报,主要为部门和地方管理服务。

统计报表必须经过法定审批程序。凡按国家法律规定审批和制发的统计报表,有关地区、部门和单位都必须切实执行。

除了以上三种调查方法,有时也采用重点调查和典型调查。重点调查是从调查对象中选择少量重点单位进行调查。其中,重



点单位是指对调查指标占较大比重的单位。例如,通过对北京大学、清华大学、复旦大学、上海交大的教学质量的重点调查,就可以对我国重点高校的教学质量有一个大致的了解。典型调查是从调查对象的全部单位中选择一个或几个有代表性的单位进行全面的调查,目的是通过典型单位来描述所研究问题的规律。例如,要研究农村致富问题,我们可以深入对华西村的调查,从中找到农村致富的原因与经验。重点调查和典型调查都是非全面调查,具有一定的主观性,因此调查结果不能用于推断总体。

三、数据的间接来源

数据如果能够通过直接的调查和研究而取得当然是最好的,但对大多数使用者来说,这往往是不可能的。这就需要通过间接的方法获得数据。例如,查阅书刊报纸、统计年鉴,也可以通过计算机网上查阅等。

有关国内数据我们可以查《中国统计年鉴》、《中国统计摘要》、《中国工业统计年鉴》、《中国人口年鉴》等等。

可以通过查阅《世界经济年鉴》、《国外经济统计年鉴》、《世界发展报告》以及联合国有关部门定期出版的世界各国统计资料获取国外数据。

在使用间接来源的数据时,应注意数据指标的含义、计算口径和计算方法,避免误用或滥用。



第三节 数据的整理

数据搜集上来以后,下一步就是对数据进行整理,使之系统化、条理化,以便进一步地分析,从而探索出事物的内在规律。

一、数据的审核

对于直接数据应从完整性和准确性两个方面去审核。完整性审核主要检查应调查的单位是否有遗漏,所有调查项目是否填写齐全。准确性审核主要分逻辑检查和计算检查。逻辑检查就是审查数据是否符合逻辑。例如,14 岁的学生填的是中共党员,显然不符合逻辑。计算检查是检查表中各项数据的计算办法、计算结果是否有误。例如,如果各项比例之和是 1.5 就肯定有误。

对于间接数据,除了对其完整性、准确性进行审核外,还要对它的适应性、时效性进行审核。适应性是指数据来源、数据口径是否符合我们研究的需要;时效性是指所得数据是否过于滞后,数据信息滞后将可能失去研究意义。

二、数据分组

数据分组是数据整理的初步工作,分组依照某种特征进行,这种特征称为分组标志。分组标志又可分为品质标志和数量标志两种。品质标志说明事物的性质;例如性别、产品等级、企业所

有制性质等等；数量标志说明事物的数量特征，例如工龄、工资、考核成绩等等。数量标志实际上就是前面所提到的定距尺度与定比尺度。

表 1.1 某城市人口按性别分组表

性别	人数	比例
男	583215	51.2
女	554321	48.8
合计	1137536	100.0

按数量标志分组，可以分为单变量分组与组距分组。其中组距分组较为普遍。

采用组距分组首先要确定分多少组，即组数。一般不应少于 5 组，也不应多于 15 组。组数太少，数据过于集中；组数太多，数据过于分散。过于集中或过于分散都不便于观察事物的规律。

组数确定后，就是确定组距。组距可由公式“组距 = (最大值 - 最小值) ÷ 组数”确定。组距分组时，一定要遵循“不重不漏”的原则。“不重”是指一个数据只能出现在一组里；“不漏”是指每个数据都能分在某一组。

为解决“不重”的问题，习惯上规定“上组限不在内”。即：相邻两组上下限重叠时，恰好等于某一组上限的变量值，不算在本组内，而算在下一组(如表 1.2 所示)。

在组距分组中，如果全部数据中的最大值和最小值与其他数据相差悬殊时，可采用开口组的方法，即：第一组采取××以下，最后一组采取××以上(如表 1.3 所示)。

表 1.2

某班数学成绩分组表

分数	人数	比例(%)
50~60	7	14
60~70	14	28
70~80	16	32
80~90	8	16
90~100	5	10
合计	50	100

表 1.3

某班数学成绩分组表

分数	人数	比例(%)
60 以下	7	14
60~70	14	28
70~80	16	32
80~90	8	16
90 以上	5	10
合计	50	100

在特殊情况下还可采取不等组的分组方法。

三、频数分布图

1. 直方图

为了更直观地看出数据的分布规律,可以采用直方图。