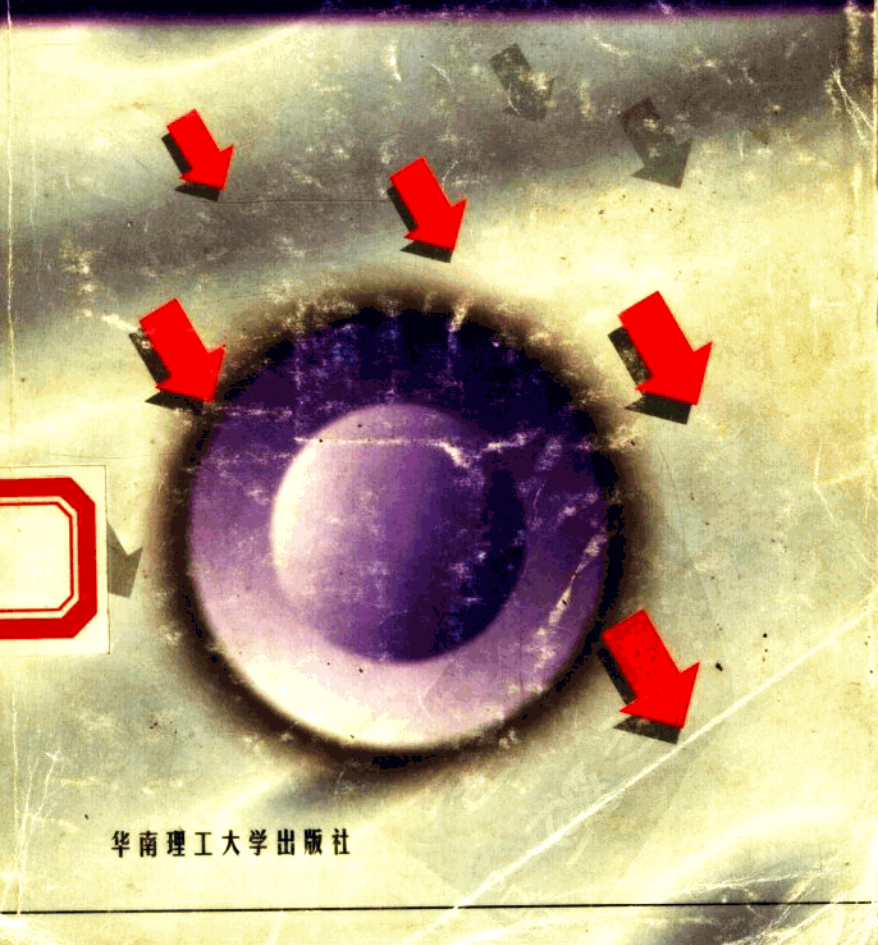


应用统计学教程

陈光潮 曾庆宾 编著

周兆麟 审稿



华南理工大学出版社

内 容 简 介

本书从应用的角度出发,论述统计学的理论和方法,按照描述统计、推断统计体系的基本内容逐步展开。表述清晰,简明扼要,信息量大,较好地处理了描述统计与推断统计的衔接,以及传统社会经济统计学原理与现代统计学的衔接。

本书为广东省自学考试开考课程使用教材。

前 言

统计学是一门既古老又现代的学科，说其古老，是因为她的发展历史与数学一样悠久；说其现代，是因为现代统计方法的发展日新月异，其应用层次不断深入、应用领域不断扩大。现在，几乎能在所有的领域找到统计学的应用痕迹。由于统计学研究对象的普遍性和统计方法的有用性，当她根植于某些领域，便会极大地推动这些领域的繁荣与发展，其中，有些已经逐步形成相对独立的统计学分支，如国民经济统计学、生物统计学等；由于统计学的渗透性和融合性，当她和某些学科相结合，便会给原有研究对象和内容赋予新的意义，其中，有些已经发展成为边缘性学科，如经济计量学、数量遗传学等。所以，统计学奠基人 R.A. 费雪曾经说过：“给 20 世纪带来人类进步独特方面的是统计学。”

任何事物、任何现象都是质与量的统一，人们要准确地把握事物，透彻地认识现象，就必须通过量的研究。统计学以自然现象、社会现象、客观事物的数量方面作为自己的研究对象，研究如何有效地收集、整理、表达、解释、分析数据，达到认识客体对象数量特征，把握客体对象本质的目的。任何事物、任何现象都不可能孤立存在，它们都必须以一定的形式组成一定意义的群体。认识、推断客体对象群体的数量特征及其关系是统计学的主要目的。群体性按统计的术语讲就是总体性。

统计学的生命力在于统计学的有用性，也即统计学的成功应用。现实需求和成功应用是统计发展的根本动力，应用是统计学

发展的永恒主题。本书从应用的角度出发，论述统计学的理论和方法，按照描述统计、推断统计体系的基本内容逐步展开。第一章至第三章是描述统计及其应用的基本内容，论述如何通过资料收集、整理、表达等方法反映、解释客体对象的数量特征及相关关系。第四章至第八章是推断统计及其应用的基本内容，论述推断客体对象未知数量特征及相关关系的基本数量方法及其应用。统计推断大致有四种基本的类型：其一，由局部推断总体；其二，由过去、现在推断将来；其三，由此推彼；其四，因素关系密切程度推断。第九章是国民经济统计的基本内容。

概括而言，本书有如下几个显著特点：

1. 表述清晰，简明扼要，信息量大。
2. 采用从特殊到一般的论述方法，通过具体实例引发统计思想和统计方法，便于自学和教学。
3. 较好地处理了描述统计与推断统计的衔接，以及传统社会经济统计学原理与现代统计学的衔接。
4. 理清了传统统计学原理中许多混乱的概念、观点和关系。
5. 重在应用，对读者不需过高的数学要求。

本书第一章主体部分由曾庆宾编写，第二章至第九章以及第一章第二节的统计图部分由陈光潮编写。全书由陈光潮总纂定稿。由于水平有限，差错难免，敬请批评指正。

作者

1999.1

目 录

第一章 资料的收集与表达	(1)
第一节 资料的收集	(1)
一、资料收集的设计	(1)
二、资料收集的方法	(5)
第二节 资料表达	(6)
一、资料整理	(7)
二、资料表达	(7)
练习题	(18)
第二章 统计资料分布特征	(21)
第一节 集中趋势	(21)
一、数值平均数	(21)
二、位置平均数	(28)
第二节 离散趋势	(32)
一、标准差、方差	(33)
二、全距、四分位差	(35)
三、离散系数、四分位系数	(36)
练习题	(37)
第三章 统计指标及其应用	(40)
第一节 相关概念	(40)
一、总体与总体单位	(40)
二、指标与标志	(41)
第二节 绝对指标、相对指标	(43)
一、绝对指标及应用	(43)

二、相对指标及其应用.....	(44)
第三节 平均指标	(47)
一、静态平均指标.....	(48)
二、动态平均指标.....	(50)
第四节 指数及其作用	(61)
一、指数的概念.....	(61)
二、指数的编制与指数体系.....	(62)
三、指数的分类.....	(70)
四、指数的应用.....	(71)
练习题	(76)
第四章 概率分布	(82)
第一节 概率基础	(82)
一、试验和事件.....	(83)
二、事件的概率.....	(83)
三、几个重要结论.....	(85)
第二节 随机变量及其分布	(88)
一、随机变量的概念.....	(88)
二、随机变量的分布.....	(89)
三、常用离散型随机变量.....	(89)
四、常用连续型随机变量.....	(92)
五、随机变量的数字特征.....	(95)
练习题.....	(100)
第五章 抽样估计及其应用	(102)
第一节 抽样分布定理.....	(102)
一、总体与样本	(102)
二、抽样分布定理	(105)
第二节 简单随机抽样.....	(107)
一、重复随机抽样	(107)

二、不重复随机抽样	(110)
三、确定必要的样本容量 n	(112)
四、几个重要的结论	(115)
第三节 其他种类随机抽样	(116)
一、等距抽样	(116)
二、类型抽样	(117)
三、整群抽样	(120)
第四节 总体比率抽样估计	(120)
一、基本概念	(120)
二、主要结论	(122)
练习题	(125)
第六章 统计检验及其应用	(126)
第一节 参数检验步骤及类型	(126)
一、参数检验的基本思想	(126)
二、参数检验的步骤	(127)
三、参数检验的基本类型	(128)
第二节 总体平均数 μ 的检验	(129)
一、 U 检验	(129)
二、 t 检验	(133)
第三节 总体方差 σ^2 的检验	(134)
一、双侧检验	(135)
二、单侧检验	(136)
第四节 两类错误	(137)
一、两类错误的基本含义	(137)
二、两类错误的应用背景	(139)
第五节 非参数检验	(140)
一、非参数检验的基本含义	(140)
二、符号检验	(140)

练习题·····	(143)
第七章 回归分析 ·····	(145)
第一节 回归分析的基本问题·····	(145)
一、回归分析的基本含义 ·····	(145)
二、回归分析的主要内容 ·····	(146)
第二节 一元回归分析·····	(147)
一、一元回归分析应用实例 ·····	(147)
二、非线性模型线性化 ·····	(153)
练习题·····	(154)
第八章 时间序列分析 ·····	(155)
第一节 时间序列因素分析·····	(155)
一、趋势变动 ·····	(155)
二、周期变动 ·····	(156)
三、随机变动 ·····	(156)
四、因素类型 ·····	(156)
第二节 趋势预测分析·····	(158)
一、移动平均法 ·····	(158)
二、指数平滑法 ·····	(162)
三、曲线趋势模型预测法 ·····	(164)
第三节 季节变动分析法·····	(165)
一、季节变动分析的意义 ·····	(165)
二、季节指数分析法 ·····	(165)
练习题·····	(168)
第九章 国民经济统计 ·····	(169)
第一节 人口与劳动力统计·····	(169)
一、人口数量与构成统计 ·····	(169)
二、人口变动统计 ·····	(170)
三、劳动力统计 ·····	(171)

第二节 国民经济统计	(172)
一、社会产品统计	(172)
二、进出口价值量统计	(175)
三、国际收支统计	(176)
四、财政统计	(177)
五、金融统计	(178)
练习题	(181)
附表 1 二项分布	(183)
附表 2 泊松(Poisson)分布	(186)
附表 3 正态分布	(192)
附表 4 t -分布的双侧临界值表	(194)
附表 5 χ^2 -分布的上侧临界值表	(197)
附表 6 符号检验临界值表	(200)
练习题提示与答案	(202)

第一章 资料的收集与表达

第一节 资料的收集

统计资料是人们从事管理和对某些事项决策判断的基础。在我国,宏观、微观经济管理,市场研究等方面所需的统计资料主要来自两个方面。一方面来自国家和地方各级统计及管理部门搜集公布的数据资料,另一方面来自为特定问题的研究所做的专门统计调查。

一、资料收集的设计

统计调查是一项系统工程,是一项繁重复杂、高度统一和严格的科学工作,应该有计划、有组织地进行。因此,在调查工作正式开始之前,必须做好各种准备,事先设计一个切实可行的统计调查方案,以便统一思想、统一内容、统一方法、统一步调,使调查工作有组织、有计划地进行,从而达到预期目的。

一个完善的统计调查方案,一般包括以下基本内容:

1. 调查目的

调查目的要明确、具体,使参加调查的所有人员都清楚所进行调查的意义和要求。例如第三次工业调查的目的是“查清工业资产底数;查清主要工业产品产量,生产能力分布和能力利用状况,资源配置、市场容量和供需平衡状况;查清主要生产设备拥有量,工艺产品生产成本、价格及盈利状况,为建立现代企业制度,实行科学管理创造条件”。

2. 调查对象和调查单位

调查对象是应搜集其资料的单位的总和(即调查范围)。例如第三次工业普查的对象确定为国有企业,乡镇、村工业,三资企业。

确定调查对象时,还必须确定两种单位,即调查单位和报告单位。调查单位是构成调查对象的每一个单位,是调查内容的承担者。如第三次工业普查中的每个国有企业。报告单位也叫填报单位,它是提交调查资料的单位。

调查单位和报告单位,根据统计调查的目的任务不同,有时两者是一致的,如第三次工业普查中,调查单位和报告单位均是国有工业,乡镇、村工业,三资企业。但有时又不一致,如在工业设备调查中,调查单位是工业企业的每台设备,而报告单位则是每个工业企业。

3. 调查项目和调查表

调查项目即调查内容,是说明调查单位的特征。调查项目的有关信息通常归集在一张表上。该表被称为调查表。如全国第三产业个体户抽样调查表中,诸如单位名称、详细地址、行业类别、城乡类别、开业年份等为该表的调查项目(见表 1-1)。调查表一般有单一表和一览表两种形式。单一表是每个调查单位填写一份,可以容纳较多的项目。一览表是把许多调查单位填列在一张表上。为了帮助填表者正确地填写调查表,还必须编制填表说明,以说明注意事项、解释调查项目的内容和说明有关数字的计算方法。

4. 调查时间

对调查项目所反映的时间和调查工作进行的时间都要有明确的规定。假定某企业 1996 年经济活动成果年报呈报时间定在 1997 年的一月份,则调查项目所反映的时间为 1996 年一年,调查工作进行的时间为一个月。

除此之外,在调查方案中还应明确调查的具体方式方法,如直接观察法、采访法、报告法等;规定调查工作的组织领导,如人员培

养、调查经费和文件等；布置有关调查的准备工作，如宣传教育、干部培训、文件印刷、调查经费的预算和开支办法等。

表 1-1 全国第三产业个体户抽样调查表

199 年

制表机关：

表号：6101-3

电话号码：

国家统计局

户主姓名：

国务院全国第三产业普查协调小组

邮政编码□□□□□□

号：国三普字(1993) 3 号

C1. 单位名称	C2. 详细地址	C3. 行业类别	C4. 城乡类别	C5. 开业年份	C6. 计算机校验平衡项
	—县(市、旗、区) —乡(镇、街道) —村(号)	1. 交通运输业 2. 商业 3. 饮食业 4. 服务业 5. 修理业 6. 其他行业 ☐	1. 城市 2. 农村 ☐	1. 90 年及 90 年前 2. 91 年 3. 92 年	C3 + C4 + C5 ☐ ☐

下面通过一例粗略地说明上述的有关概念和统计工作的基本内容。

例 1.1 假定对某大学某班 50 名学生期末考试成绩进行调查，得到“应用统计”科成绩资料如下：

50	65	72	71	86	51	66	73	78	86
54	67	74	79	87	58	68	74	80	88
59	69	74	81	90	61	69	75	82	91
61	70	75	83	91	62	71	75	84	95
63	72	75	84	97	64	72	76	85	99

从上面资料看,这一组原始数据是一种分散、杂乱无章的数字资料,它只能说明各个单位的具体情况,无法从总体上说明学生成绩的分布规律。因此,我们就有必要对上述资料进行必要的分组。

假定上述调查的目的是要知道“应用统计”考试的试卷深浅度是否合理?该科考试人数中各类占的比例有多大?为完成此项目标,就必须选择考试分数作为分组标志。

对上述资料进行按组距为 10 分进行分组,得表 1-2。

表 1-2 某班学生考试成绩分组统计

考试成绩(分)	人数(个)(标数)	比率(%) (频率)
60 分以下	5	10.0
60~70	11	22.0
70~80	17	34.0
80~90	11	22.0
90~100	6	12.0
合计	50	100.0

上表各组的临界值的归属规则是,各组只含本组的下限,而不含本组的上限,即上限不在组内的原则。例如,对 90 分的同学而言,只能统计在 90~100 这一组,而不能统计在 80~90 这一组。

从表中可以看出,50 个学生成绩次数分配基本呈单峰对称分布,即标志值小的和标志值大的两端单位较少,居中的标志值的单位多,符合正态分布,说明该科试卷的难易深浅程度是合理的,是一份较理想的考试卷,同时经过分组,我们可以看到不及格人数有 5 人,占总体的 10%,合格人数有 45 人,占 90%,其中优良率为 74%,人数 17 人,说明该班该门课学习成绩也是较理想的。

此例中,调查对象是 50 名同学组成的整体,调查单位是每一个同学,调查项目是应用统计科的成绩。报告单位根据提交成绩的不同,可以是班集体,也可以是大学教务处。

通过此例,我们可以粗略地了解统计资料搜集、统计资料整理、统计分析的工作过程。

二、资料收集的方法

统计调查的目的是为适应管理或研究的需要搜集有关的统计资料。统计资料收集的方法有许多,但在组织形式上主要有普查、报表制度、抽样调查、重点调查和典型调查。

1. 普查

普查是专门组织的一次性全面调查。从宏观上讲,普查是了解国情、国力的重要调查方法。如1978年进行了全国科学技术人员普查,1985年进行了全国工业普查,1990年又进行了第四次全国人口普查等。

2. 统计报表

统计报表是我国搜集统计资料的一种主要方式方法,同时也是我国政府统计工作的一项基本内容,它是按照国家统一规定的表格形式、统一规定的内容、统一的报送程序和报送时间,由填报单位自下而上地逐级提供统计资料的一种统计调查方法。

我国现行的统计报表按内容的不同可分为基本统计报表和专业统计报表。基本统计报表是由国家统计局统一制发,用来搜集国有经济各部门各行业最基本的统计资料。专业统计报表是各有关部门为专业管理工作的需要而印制的,用以搜集本部门的业务技术资料,是基本报表的补充。

定期统计报表还可以按报送周期长短不同分为日报、周报、月报、季报、年报等。按报送方式的不同,可分为电讯报和邮寄报等。

3. 抽样调查

抽样调查是一种非全面调查,它是在一定范围内按照随机原则从所研究对象的全体中抽取部分单位作为样本进行观察,然后依据所获得的样本数据,推算全体数量特征的一种调查方法。所谓随机原则,就是同等机会原则,在确定的范围每一个单位都有同

等抽中的机会。

抽样调查是科学研究和现代管理决策的一种十分重要的调查方法,有广泛的应用范围,特别适用于那些不可能或不必要进行全面调查的场合。如:带有破坏性的工业产品质量检验、个体工商业营业状况调查、城市居民家庭生活情况调查等等。

4. 重点调查

重点调查是一种非全面调查,它是从所要调查对象的全体中选择一部分重点单位进行调查。重点单位,是着眼于现象量的方面,尽管这些单位在全部单位中只是一部分,但是它们的某一主要标志的标志总量在总体标志总量中占有绝大比重。例如:对大型钢铁公司的产量(产量作为钢铁的主要标志)和生产成本的调查。

5. 典型调查

根据调查的目的和任务,在对所研究的对象进行分析的基础上,有意识地选取若干具有代表性的单位进行深入细致的调查和研究,其目的是希望将典型单位的结论推广到一般。典型调查其调查单位的确定与其他调查单位的确定相比较,更多地取决于调查者主观的判断。

第二节 资料表达

资料搜集取得的原始资料是分散的、杂乱的、不系统的,大多表明各个具体单位的具体情况,反映事物的表面现象或一个侧面,不能说明事物的全貌、总体情况。因此,只有对这些资料进行加工整理才能认识事物的总体及其内部联系。例如:工业普查的每个工业企业资料,只能说明每个企业的情况,诸如企业所有制类型、资金多少、设备状况、职工人数、销售收入和利润等等。必须通过对所有企业资料进行整理、汇总、分组等加工处理后,才能得到全国工业的综合情况,从而分析工业的构成、经营状况、规模水平等,

达到对全国工业的全面、系统的认识。

一、资料整理

统计资料的整理是按照研究的目的和要求,将搜集到的原始资料进行科学的排序、分组、汇总等,从中提取有用的信息并探索其中的数量规律性。统计资料的分组是资料整理的重要的一步,即按不同的分组标志将数据资料划分为几个部分。按事物的某种属性对数据资料分类,如企业按经济类型、行业分组;人口按性别、民族分组;大学生按专业分组等。这种分类称为按品质标志分组。这种分组可以反映总体的构成和不同属性事物在总体中的地位和作用。总体是研究对象全体的总和。按数量标志对数据资料分类,如企业按生产能力、劳动生产率分组;商店按商品流转额、职工人数分组。这种分组的目的在于通过事物在数量上的差异来反映事物在性质上的区别。

统计数据按照某种规则排序、分组以后,可以形成一定形式的统计数列。粗略地看,统计数列可表现为四种类型。即:空间数列、品质数列、时间数列和变量数列。空间数列是按不同地区标志进行的分组。例如:人口按省、市、自治区分组所形成的数列即为空间数列。品质数列是按现象的性质、类别标志进行的分组,例如人口数按性别或民族分组所形成的数列即为品质数列。时间数列是按照时间发生的先后顺序所形成的数据数列,例如我国解放后各年的人口数字就是时间数列。变量数列是按某一数量标志大小顺序对一组数据进行排序或分组的结果,例如某企业按工资收入多少进行的分组就构成变量数列。若仅按时间特征来划分,统计数列可分为动态数列和静态数列。动态数列也即时间数列。

二、资料表达

统计调查往往给出大量的资料,要使这些资料被人看懂,或较清晰地加以显示,就必须以某种方式对它进行概括,清晰而有力地展示出来。表达统计资料常用的方式有三:一种是把数字本身加

以概括,一般是用表格形式;第二种是用图的形式——图形、图解的形式;第三种是次数分布形式。用其中任一种方法将大量资料展示出来,都是统计分析的有机部分,由此可对所要研究的资料有更好的全面的理解。

(一) 统计表

经过整理的数字资料,用统计表加以概括表述,不仅节省大量的文字叙述,而且更为集中醒目,条理分明,便于比较分析和资料积累。

1. 统计表的构成

把大量的统计数字资料,按一定顺序和格式列在表上,该表就是统计表。统计表的结构,从形式上看,由总标题、横行标题、纵列标题和数字资料等要素构成。从统计表的内容及所处的地位看:包括主词和宾词两部分。统计表的结构形式如表 1-3。表中的总

表 1-3 统计表的结构形式

表 × × 广东省 1997 年国内生产总值		总标题	
主 词		宾 词	
按三次产业分		国内生产总值(亿元)	百分比(%) 纵列标题
横 列 标 题	第一产业	986.82	13.49
	第二产业	3647.82	49.86
	第三产业	2680.87	36.65
	总计	7315.51	100

标题是:“广东省 1997 年国内生产总值”,横行标题从第二行开始依次是:“第一产业、第二产业、第三产业、总计”,纵列标题从第一列开始依次是:“按三次产业分、国内生产总值、百分比”。表中的主词是三次产业,它们是统计表所反映的对象,在统计表的内容中处于主导地位。表内的宾词是国内生产总值、百分比及相关的数