

应用型

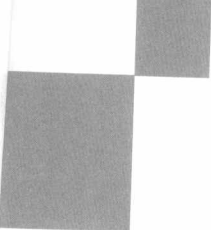
主 编/杨缅昆 方国松

统计学概论

*Introduction To
Statistics*



清华大学出版社



应用型

• 主 编/杨缅昆 方国松

统计学概论

*Introduction To
Statistics*

清华大学出版社
北 京

内 容 简 介

本书针对应用型院校复合型人才的培养目标,适当弱化了统计学中的数学理论推理,更注重统计学的基本概念和基本原理,增强实验教学,培养学生分析问题、解决问题和动手实践的能力,突出了应用统计的特点。全书共十章,包括第一章绪论,第二章统计数据的搜集、整理与显示,第三章统计分布的数值特征,第四章概率基础和抽样分布,第五章统计推断,第六章相关与回归分析,第七章统计指数,第八章时间数列分析,第九章方差分析,第十章马克威统计分析软件在统计中的应用等内容。

本书适用于应用型本科院校统计学专业的学生;在对本书内容酌情取舍后,也适用于应用型本科院校非统计学专业的学生。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

统计学概论/杨缅昆,方国松主编. —北京:清华大学出版社,2009.9

ISBN 978-7-302-20926-3

I. 统… II. ①杨… ②方… III. 统计学-高等学校-教材 IV. C8

中国版本图书馆CIP数据核字(2009)第150649号

责任编辑:杜春杰

封面设计:唐韵设计

版式设计:魏 远

责任校对:王 云 柴 燕

责任印制:杨 艳

出版发行:清华大学出版社

地 址:北京清华大学学研大厦A座

<http://www.tup.com.cn>

邮 编:100084

社 总 机:010-62770175

邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

印 刷 者:清华大学印刷厂

装 订 者:三河市李旗庄少明装订厂

经 销:全国新华书店

开 本:185×230 印 张:17.25 字 数:346千字

版 次:2009年9月第1版 印 次:2009年9月第1次印刷

印 数:1~5000

定 价:35.00元

本书如存在文字不清、漏印、缺页、倒页、脱页等印装质量问题,请与清华大学出版社出版部联系调换。联系电话:(010)62770177 转 3103 产品编号:029330-01

前 言

统计学是关于数据搜集、整理、归纳与分析，并从中探寻客观对象内在规律的方法论科学。根据教育部的要求，统计学是经济、管理类本科生必修的核心课程。随着人们对统计学重要性的认识日益增强，现在已经有越来越多的人文学科的有关专业也开设了这门课程。可以预见，随着人们对统计分析和数量分析提出更高要求，统计学作为一门基础性课程，其重要作用将与日俱增。

众所周知，我国实行扩招政策以来，在高等教育领域已经形成了一个新的与研究型本科教育群体不同的教育群体——应用型本科教育群体。然而，我们注意到，我国教育界并没有专门为这部分教育群体准备相适应的教材，他们所使用的教材绝大多数仍然是以培养研究型人才为目的的教材。统计学科也不例外。由于现有的统计学教材在设计上过于强调理论性和系统性，忽视基础性和应用性，导致数学证明及各种数理统计方法在课程中占有较大比重，教材的难度明显与这个教育群体的文化结构和数学基础不相适应。面对晦涩难懂、不易理解的教学内容，这个教育群体容易产生畏难情绪，陷入一种厌学的困境，无法保证教学效果。面向该教育群体，重视理论的应用性是本书编写的初衷。

本书是由仰恩大学从事统计学教学多年且授课经验丰富的教师编写而成的。全书根据应用型本科教育群体自身的特点，重新谋篇布局，重点放在体例结构和教学内容的调整上，形成以下几个特点：

1. 淡化统计公式的数学证明，注重统计思想的阐述。对于应用型本科学生来说，学习统计学课程，目的在于让学生在正确把握统计思想本质的基础上，根据不同的社会经济现象熟练地运用正确的统计方法。因此，在本书的许多章节中，在不失严谨的前提下，力求淡化数学推导。我们主要通过大量的图示、表格和实例，生动形象地把统计分析的理论和方法呈现在读者面前。

2. 编写体例新颖，贴近现实生活，有助于培养与提高学生的学习兴趣。统计学的教学涉及大量的数学公式和繁琐的计算，容易使学生产生畏难心理，为此，我们在相关的各章节中将各种统计分析方法划分成若干类型，分别给出规范化的计算分析程序，从而简化统计方法的计算。例如，统计推断中的区间估计、假设检验，我们按总体平均数和总体比率归纳出若干不同的分析计算程序，不仅使学生对统计推断的理解更加深化，而

且有助于培养和提高学生的学习兴趣。另一方面,我们力求将统计学原理和方法的阐述,贴近现实的社会经济生活。例如,在时间数列分析和统计指数部分,我们结合当地的经济发 展实例来阐述统计分析方法,不仅具体、容易理解,而且有助于培养学生理论联系实际的学风。

3. 引入“马克威”软件,加强实验教学,提高学生的动手能力。统计学是一门应用性很强的科学,因此,改变传统的统计教学模式,将实验教学引入统计学课程,是统计教学改革的重要方向。我们通过引入统计分析软件——马克威分析系统,将统计实验教学作为统计学教学的重要内容,目的在于强化学生的动手能力,在学习统计方法的过程中体会统计思想的实质,从而有助于全面提高应用型人才的综合素质。

4. 将教学内容进行分类,以满足不同的授课对象的教学需要。本书适用于应用型本科院校统计专业的学生,而对于经济、管理类专业以及人文科学相关专业等非统计专业的学生,由于授课对象不同,更由于教学课时少,不可能也没有必要讲授本书的全部内容。基于这个原因,我们将教材内容进行了一定程度的分类,建议在授课对象为非统计专业的学生或者教学课时较少的情况下,部分内容可以不讲。例如,第六章的多元线性回归分析和非线性回归部分;第九章的方差分析部分等。这些地方我们在目录中以“※”标出。当然这仅仅是我们的建议,任课教师可以根据具体情况对授课内容进行选择。

本书是在参阅国内外统计学专著和教材的基础上完成的。全书由杨缅昆、方国松担任主编,负责全书框架的设计、确定各章节的基本内容,提出编写的要求,并进行总纂定稿;全书由十章组成,各章的内容及编写的具体分工如下:严斌(第一章)、都巍巍(第二章、第三章、第十章)、王玲玲(第四章、第九章)、张俊霞(第五章、第六章)、方国松(第七章)、冯海燕(第八章);最后由陶立新、钟翼民审定。

本书的编写得到了仰恩大学领导的鼓励和关心,清华大学出版社以及本书责任编辑的大力支持使本书能够顺利出版,在此一并表示衷心的感谢。

作为一种探索,本书难免存在一些问题与偏差。一本内容科学实用、适应教学相长的好教材,不仅需要编写人员专业知识的积累,而且需要在总结教学实践经验的基础上不断修改与完善,为此我们恳请各位同行专家和广大师生提出宝贵意见。

杨缅昆 方国松
于福建泉州仰恩大学
2009年7月

目 录

第一章 绪论	1
第一节 统计学的产生与发展	1
一、古典统计学时期	1
二、近代统计学时期	2
三、现代统计学时期	3
第二节 统计学的研究对象和性质	3
一、统计学的研究对象	3
二、统计学的性质	5
第三节 统计学的基本概念	5
一、统计总体、总体单位与样本	5
二、标志、变量和统计指标	7
三、参数与统计量	10
思考与练习	10
第二章 统计数据的搜集、整理与显示	11
第一节 统计数据的搜集	11
一、统计数据的直接来源	11
二、调查方案设计	15
第二节 统计数据的整理	17
一、统计分组	17
二、变量数列	20
第三节 统计数据的显示	21
思考与练习	25
第三章 统计分布的数值特征	27
第一节 集中趋势指标	27
一、算术平均数	27

二、调和平均数	30
三、几何平均数	34
四、算术平均数、几何平均数、调和平均数的比较	35
五、众数	36
六、中位数	37
七、四分位数	40
八、平均指标的比较	42
第二节 离散程度指标	43
一、极差	43
二、四分位差	44
三、平均差	44
四、方差与标准差	45
五、变异系数	47
※第三节 分布形态的描述指标	49
一、矩的概念	49
二、分布的偏态	49
三、分布的峰度	50
思考与练习	51
第四章 概率基础和抽样分布	55
※第一节 随机事件及其概率	55
一、试验与事件	55
二、事件的概率	56
三、概率的性质及运算法则	57
第二节 随机变量的概率分布	59
一、离散型随机变量的概率分布	60
二、连续型随机变量的概率分布	62
三、随机变量的数字特征	64
第三节 抽样分布	65
一、抽样以及抽样分布的含义	65
二、重置抽样下的抽样分布	66
三、不重置抽样下的抽样分布	69
第四节 正态分布	72

一、正态分布在统计学中的地位	72
二、正态分布的密度函数及其数学性质	72
三、正态分布函数及其标准化	74
四、关于抽样分布的定理	76
五、几个与抽样有关的概率分布	79
思考与练习	82
第五章 统计推断	84
第一节 抽样推断及其特点	84
一、抽样推断的特点	84
二、总体参数和样本统计量	84
三、抽样推断的基本条件	85
四、抽样推断的误差	86
第二节 总体参数估计	88
一、总体参数估计概述	89
二、点估计	89
三、区间估计	91
四、样本容量的确定	95
第三节 假设检验概述	97
一、假设检验的基本概念	97
二、假设检验中的基本问题	98
三、总体平均数的检验	100
四、总体成数的检验	103
五、 P 值检验	104
思考与练习	105
第六章 相关与回归分析	109
第一节 相关与回归分析的基本概念	109
一、函数关系与相关关系	109
二、相关关系的种类	109
三、相关分析及其内容	111
四、回归与回归分析	112
五、相关分析与回归分析的关系	112

第二节 简单线性相关分析	113
一、相关系数的计算公式	113
二、相关系数的特征及判别标准	114
三、相关系数的检验	116
第三节 一元线性回归分析	117
一、一元线性回归模型的建立	117
二、一元线性回归模型的参数估计	118
三、一元线性回归拟合优度的评价	120
四、回归估计标准误差	122
五、显著性检验	122
六、一元线性回归模型预测	123
※第四节 多元线性回归分析	125
一、多元线性回归模型的建立	125
二、多元线性相关分析与拟合优度的评价	128
三、多元回归的回归估计标准误差	130
四、显著性检验	131
五、多元线性回归模型预测	133
六、应用统计软件进行回归分析	133
※第五节 非线性回归分析	136
思考与练习	140
第七章 统计指数	144
第一节 指数的概念与分类	144
一、统计指数的概念	144
二、统计指数的分类	145
三、总指数编制的两个基本方式	148
第二节 综合指数	148
一、综合指数的概念及编制原理	148
二、拉氏指数	149
三、帕氏指数	151
四、综合指数的其他形式	152
第三节 平均指数	153
一、平均指数的概念及编制原理	153

二、算术平均指数	154
三、调和平均指数	155
第四节 指数体系与因数分析	158
一、指数体系的概念及其作用	158
二、指数因素分析方法	159
三、总量指标的因素分析	159
四、平均指标的因素分析	164
※第五节 其他常用指数	168
一、工业生产指数	168
二、农副产品收购价格指数	169
三、股票价格指数	169
四、居民消费价格指数	170
思考与练习	174
第八章 时间数列分析	179
第一节 时间数列概述	179
一、时间数列的概念和作用	179
二、时间数列的种类	180
三、时间数列的编制原则	182
第二节 时间数列的水平指标	183
一、发展水平	183
二、平均发展水平	184
第三节 时间数列的速度指标	190
一、增长量和平均增长量	190
二、发展速度与增长速度	193
三、平均发展速度和平均增长速度	197
※第四节 时间数列的构成因素分析	199
一、时间数列的构成因素和分析模型	199
二、长期趋势测定	200
※第五节 季节变动分析	207
一、季节变动及其测定目的	207
二、季节变动分析的原理和方法	207
三、季节变动的调整	210

※第六节 循环变动分析	211
一、循环变动及其测定目的	211
二、循环变动的测定方法	211
思考与练习	215
※第九章 方差分析	218
第一节 引言	218
一、问题的提出	218
二、方差分析的基本概念	218
三、方差分析的基本思想和原理	220
四、方差分析的基本假设	221
第二节 单因素方差分析	221
一、单因素方差分析的数据结构	221
二、方差分析的数学模型	222
三、方差分析的基本方法	223
四、方差分析案例	225
第三节 双因素方差分析	226
一、双因素方差分析概述	226
二、无交互作用的双因素方差分析	228
三、有交互作用的双因素方差分析	231
思考与练习	235
第十章 马克威统计分析软件在统计中的应用	238
实验一 统计数据的整理	238
实验二 描述性统计指标的计算	241
实验三 总体参数的假设检验	245
实验四 相关与回归分析	248
实验五 一元方差分析	252
※实验六 多元方差分析	255
附录 A	259
参考文献	266

注：标“※”章节为选择性使用内容，授课老师可根据教学要求自主决定。

第一节 统计学的产生与发展

统计学的产生与统计实践活动是密不可分的,统计作为一种社会实践活动,已有四五千年的历史。早在原始社会,人们按部落居住在一起,打猎、捕鱼后就要算算有多少人、多少食物,以便分配食物;我国夏禹时代就有了人口数据的记载;为了赋税、徭役和兵役的需要,历代都有田亩和户口的记录。而统计学的理论和方法,则是在长期统计实践活动的基础上发展起来的,距今已有三百多年的历史。从统计学的发展过程来看,可以大致分为三个阶段。

一、古典统计学时期

17 世纪中叶至 18 世纪中叶是古典统计学时期,在这一时期,统计学理论初步形成了一定的学术派别,主要有国势学派和政治算术学派。

1. 国势学派

国势学派又称记述学派,产生于 17 世纪的德国。由于该学派主要以文字记述国家的显著事项,故称记述学派。其主要代表人物是海尔曼·康令(Hermann Conring, 1606—1681)和阿亨华尔(Gottfried Achenwall, 1719—1772)。康令于 1660 年把国势学从法学、史学和地理学等学科中独立出来,在大学中讲授“实际政治家所必需的知识”;阿亨华尔在哥廷根大学开设“国家学”课程,其主要著作是《近代欧洲各国国势学纲要》,书中讲述“一国或多数国家的显著事项”,主要用对比分析的方法研究了国家组织、领土、人口、资源财富和国情国力,比较了各国实力的强弱,为德国的君主政体服务。因在外文中“国势”与“统计”词义相通,后来正式命名为“统计学”。国势学派只是对国情的记述,偏重事物性质的解释,未能进一步揭示社会经济现象的规律,也不研究事物的计量分析方法,不注重数量对比和数量计算,只是用比较级和最高级的词汇对事物的状态进行描述。所以,人们也把它叫做记述学派(旧学派或德国学派),并认为国势学派有统计学之名而无统计学之实。

2. 政治算术学派

政治算术学派产生于 19 世纪中叶的英国,创始人威廉·配第(William Petty, 1623—

1687)，其代表作是他于1676年完成的《政治算术》一书，这本书是经济学和统计学史上的重要著作，这里的“政治”是指政治经济学，“算术”是指统计方法。在这部书中，他利用实际资料，运用数字、重量和尺度等定量分析工具对英国、法国和荷兰三国的国情国力，作了系统的数量对比分析，其所采用的方法是前所未有的，为统计学的形成和发展奠定了方法论基础。因此马克思说：“威廉·配第——政治经济学之父，在某种程度上也是统计学的创始人。”政治算术学派的另一个代表人物是约翰·格朗特（1620—1674），他以1604年伦敦教会每周一次发表的“死亡公报”为研究资料，在1662年发表了《关于死亡公报的自然和政治观察》的论著。书中通过大量观察发现了人口各年龄组的死亡率、性别比例等重要的数量规律，并对人口总数进行了较为科学的估计；并且第一次编制了“生命表”，对死亡率与人口寿命作了分析，从而引起了普遍的关注。因此，他被认为是人口统计学的创始人。

二、近代统计学时期

18世纪末至19世纪末是近代统计学时期，在这个时期，各种学派的学术观点已经形成，并且形成了两个主要学派，即数理统计学派和社会统计学派。

1. 数理统计学派

在18世纪，概率理论日益成熟，为统计学的发展奠定了基础。19世纪中叶，概率论被引进统计学从而形成数理学派，其奠基人是比利时的阿道夫·凯特勒（Lambert Adolphe Jacques Quetelet, 1796—1874），在其《社会物理学》中将古典概率论引入统计学，使统计学进入一个新的发展阶段。他认为概率论是适于政治及道德科学中以观察与计数为基础的方法，并以此方法对自然现象和社会现象的规律性进行观察，并认为要促进科学的发展，就必须更多地应用数学。总之，他把概率论引入统计学，为数理统计学的形成与发展奠定了基础。

2. 社会统计学派

社会统计学派产生于19世纪后半叶，创始人是德国经济学家、统计学家克尼斯（K.G.A.Kn-ies, 1821—1898），主要代表人物有厄恩斯特·恩格尔（Christian Lonrenz Ernst Engel）、乔治·冯·梅尔（Georg von Mayr, 1841—1925）等人。他们融合了国势学派与政治算术学派的观点，沿着凯特勒的“基本统计理论”向前发展，但在学科性质上认为统计学是一门社会科学，是研究社会现象变动原因和规律性的实质性科学，以此同数理统计学派通用方法相对立。社会统计学派在研究对象上认为统计学是研究总体而不是个别现象，而且认为由于社会现象的复杂性和整体性，必须对总体进行大量观察和分析，研究其内在联系，才能揭示现象内在规律。这是社会统计学派的“实质性科学”的显著特点。

三、现代统计学时期

20 世纪至今为现代统计学时期,这一时期的主要特征是描述统计学已转向推断统计学,1907 年,英国人戈塞特(1876—1937)提出了小样本 t 统计量理论,丰富了抽样分布理论,为统计推断奠定了基础。英国的 R.A.Fisher(1890—1962)提出了极大似然估计量的概念,迅速成为了估计参数的重要方法,他还提出样本相关系数的分布、实验设计和方差分析等方法。英国科学家弗朗西斯·高尔顿(Francis Galton)提出了相关与回归思想,并给出计算相关系数的明确公式。英国统计学者 K.皮尔逊发展了拟合优度检验,还给出了卡方统计量及其极限分布,波兰学者奈曼(J.Neyman, 1894—1981)创立了区间估计理论,并和 E.皮尔逊发展了假设理论。美国学者瓦尔德提出决策理论和序贯抽样方法。美国化学家威尔科克森(Frank Wilcoxon)发展了一系列非参数统计方法,开辟了统计学的新领域。由马哈拉诺比斯领导的印度统计研究所和 20 世纪 30 年代后期奈曼发表的两篇论文,使抽样的数学理论在 20 世纪 30 年代得到了迅速发展。

统计学大致经过以上三个阶段发展到今天,随着统计学理论知识的发展与健全,统计学的应用领域将会进一步扩大,将出现许多新型的交叉学科,比如统计应用到法律、文学等学科。同时,伴随着计算机技术的飞速发展,统计学还将在模糊现象、突变现象及混沌现象等方面开辟新的研究领域。

第二节 统计学的研究对象和性质

一、统计学的研究对象

空间是三维的,时空是四维的,人们所面临的客观事物更是多维的,统计学正是要从复杂多变的客观事物中,挖掘出其蕴涵的客观规律,为人们的各种行为活动提供有力的参考,避免行为的盲目性。由于客观事物往往可以用其数量表现,因此,要找出客观事物的内在规律,首先要认识客观事物,那么就必须通过试验或调查来搜集有关数据,并且加以整理、归纳和分析,以便对客观事物规律性的数量表现作出合理的描述。

由此可见,统计学的研究对象是客观事物的数量特征和数量关系。统计学也就是关于数据搜集、整理、归纳、分析的方法论科学,其目的是探索数据的内在数量规律性。

统计学的研究对象具有以下特点:

1. 数量性

客观现象有着质和量两个方面的表现,根据质和量的辩证统一研究现象的数量特征,

从数量上认识现象的性质和规律性,这是统计研究的基本特点。统计运用科学方法搜集、整理、分析反映现象特征的数据,并通过统计指标反映现象的规模、水平、比例、速率及其变动规律。认识现象的数量表现,是深入研究现象质的表现的前提和基础。现象的数量方面包括数量多少、数量关系、质和量互变的数量界限等。数量关系指各种平衡关系、比例关系和依存关系,例如总供给与总需求的平衡关系,各产业间的比例关系,消费与收入之间的依存关系等。客观现象往往具有复杂性的特点,现象之间具有多方面的联系。在研究现象的数量方面时,我们必须把握现象的全貌,反映现象发展变化的过程,必须紧密联系现象的具体内容和联系质的特征,这是统计学与数学的一个重要区别。例如,一个国家的人口数量、结构和分布;国民经济的规模、发展速度;人们的生活水平等,都是反映基本国情和基本国力的基本指标,通过这样的一系列指标才能对整个国家有一个客观清晰的认识。由此可见,数量性是统计研究对象的特点之一。对于用定性方式表述的客观现象,则应该将其转换为数量形式,例如为了反映某产品的质量情况,用合格与不合格来表示,这时可分别记合格与不合格为1与0等。

2. 大量性

大量性也称总体性,统计研究的对象总是由大量同类事物构成的总体现象的数量特征。个别和单个事物的数量表现是可以直接获取的,一般不需运用统计研究方法。例如,要了解某名工人的工作情况,查一查生产记录就可以了,可如果要了解全体工人产量的分布、差异和一般水平等,就要用统计方法来进行计算和分析。统计对总体现象的数量特征进行研究时,是通过对组成总体的个别事物量的认识来实现的。例如,在人口普查中我们通过对每一户家庭的人口状况进行调查,根据所取得的资料,编制人口总数、人口结构(性别、年龄、民族、职业等结构)、人口分布、人口出生率、人口死亡率等指标来反映一个国家或一个地区的人口总体状况。个别事物有很大的偶然性,大量事物具有共性,统计学正是要从大量的客观事物中找出其共性,即规律性。从对个体数量特征的观测入手,运用科学的统计方法获得反映总体一般特征的综合数量,这是统计的又一基本特征。

3. 变异性

变异性是指组成研究对象的各个单位在特征表现上存在差异,并且这些差异是不可以按已知条件事先推断的。例如,要研究某地区大学生的消费行为,每个学生的家庭收入有差异,学生的消费偏好有差异,消费品的市场价格也不稳定。这时就需要研究大学生的平均消费、家庭平均收入、消费偏好和消费品的市场价格等因素,如果每个大学生不存在这些差异,我们只要调查一个学生相关消费行为,就可以知道整个地区的大学生消费行为,这时也就不需要做统计了。正是因为研究对象的各单位存在差异性,统计方法才有了用武之地。

二、统计学的性质

根据统计学前面的定义,我们很容易知道统计学的性质:统计学是一门认识方法论科学,具体说它是研究如何搜集数据、整理数据并分析数据,以便从中作出正确推断的认识方法论科学。

之所以统计学具有这样的性质,是因为:首先,统计学是为了揭示客观事物的规律性;其次,为了达到这个目的,需要各种统计方法来认识事物的真面目。因此,统计学是认识客观事物的方法论科学。

统计学和数学都是研究数量关系的科学,它们之间既有联系又有区别。一方面,数学以抽象的概念和方法研究各种数量关系和空间形式,而统计学则是对客观现象在质和量的相互联系中研究其数量方面,揭示其数量变动的规律性,这是它们之间的本质区别。另一方面,数学又为统计学提供大量的计算分析方法,尤其是数理统计不仅用于研究社会经济现象,也可用于研究自然技术现象。工业产品、农副产品的抽样调查、生产过程的检验和控制等就是数理统计方法在社会经济领域中的应用。

统计学在研究客观现象的数量特征和数量关系时,必然要以相关的科学的基本理论和基本知识为指导,如经济学、社会学、物理学、生物学、心理学等。而且,统计学的基本理论在各个领域中的应用形成了各种专门统计学,如经济统计学、人口统计学、科技统计学、金融统计学、经营统计学、心理统计学等。统计学与相关科学的结合同时促进了统计理论和方法的发展。

第三节 统计学的基本概念

一、统计总体、总体单位与样本

(一) 统计总体与总体单位

统计总体,简称总体,就是根据一定目的确定的所要研究对象的全体,它是由客观存在的、具有某种共同性质的许多个别单位构成的整体。总体单位,简称单位,是组成总体的各个个体。我们可以把总体看成是集合,而单位则看成是集合中的元素。

例如,研究某厂生产的一批日光灯的寿命,则该厂生产的这批日光灯组成总体,而这批产品中的每一只日光灯就是一个单位。研究某企业职工의年龄结构,则该企业所有职工组成总体,而该企业的每一名职工就是一个单位。

统计总体可分为有限总体和无限总体,有限总体是由有限个单位构成的总体,换句

话说,总体的范围能够明确确定。例如,要了解某大学学生的学习情况,则该大学全部学生所构成的总体就是有限总体。无限总体是由无限个单位组成的总体,比如,要检验某咖啡厂的咖啡装袋重量是否正常,倘若该厂可以无限生产下去,则该厂生产的咖啡可视为无限总体。又如,海里的鱼、天上的星星都可视为无限总体。区分无限总体和有限总体的意义在于对不同的总体应分别采用不同的调查研究方式。对于有限总体既可以采用全面调查方式,也可以采用非全面调查方式;对于无限总体则只能采用抽样调查方式。

总体与单位是两个不同层次的概念,它们之间的关系具有相对性。随着研究目的的不同,总体和单位的关系可以发生变化。例如,要研究某高校各专业的办学情况,该高校的所有专业构成总体,而该校开办的每一个专业就是一个单位。如果要研究某地区高等学校的办学情况,该地区所有高等学校构成总体,而该地区的每一所高校就是一个单位。

(二) 样本

从总体中抽取的部分单位组成的集合称为样本。抽取样本的目的,就在于要用样本的数量特征来估计或推断总体的数量特征。对于无限总体我们不可能对每一单位进行观察,即使是有限总体,由于其大量性的特点,要对所有单位进行观察,要花费大量的人力、物力、财力和时间,是十分不经济的事情。因此,一般情况下,我们都是通过样本来推断总体的特征的。

既然抽样的目的是推断总体的特征,因而从总体中抽取样本时必须遵循随机原则,这样才能保证样本的代表性。例如,我们研究某厂生产的日光灯的寿命,随机抽取 100 支日光灯进行检验,则这 100 支日光灯就是一个样本。总体是统计研究的对象,样本作为总体的代表,也是统计研究的对象,因此样本也符合总体的概念,为了加以区别,通常将所要研究的事物全体构成的总体称为全及总体,而将样本单位组成的总体称为抽样总体。

抽取样本时要注意以下问题:

1. 从一个总体中可以抽取许多个不同的样本。根据研究目的确定的统计总体是唯一确定的,而样本却是随机的,从一个总体中抽取不同样本的数目的多少与样本单位数和抽样方式有关。例如,某校有 $N = 10\,000$ 名学生,从该校学生中抽取 $n = 100$ 名构成样本,按重置抽样方式其可能的样本数目为: $N^n = (10\,000)^{100} = 10^{400}$,按不重置抽样时的情形请读者推算。由此可见,从一个总体中抽样一定样本容量,不同的样本是非常多的。

2. 样本的代表性。抽样的目的是用样本的数量特征去推断总体的数量特征,因此,就要求样本的指标与总体的指标的误差要小,即抽样误差小。抽样误差越小则样本的代表性就越强。样本的代表性的高低与样本单位数、抽样方式和抽样的组织形式有关。提高样本的代表性,降低抽样误差,是抽取样本时要高度关注的问题。

3. 样本的客观性。在抽取样本时,要遵循随机原则,排除主观因素的影响,保持取样的客观性,从而提高样本的代表性。