

应用统计学

主 编 王淑芬



北京大学出版社
PEKING UNIVERSITY PRESS

中国林业出版社
China Forestry Publishing House



清华大学出版社 中国统计出版社 中国统计出版社

应用统计学

（第二版）

清华大学出版社 中国统计出版社
Tsinghua University Press China Statistical Publishing House

21 世纪全国应用型本科财经管理系列实用规划教材

应用统计学

主 编 王淑芬

参 编 常 明 卯 越 盖 兵



北京大学出版社
PEKING UNIVERSITY PRESS

中国林业出版社
China Forestry Publishing House

内 容 简 介

本书结合作者多年的教学实践经验和国内外优秀统计学教材的成果编写而成。基本内容包括描述性统计、推断性统计及多元统计分析,其中包括近年来的一些较新进展。大部分统计方法都给出了 SPSS 软件的操作过程及输出结果的解读。每章后均附有习题,书后附有全部习题的参考答案。便于教学、自学和应用。

本书可作为高等院校财经管理类和理工科各专业本科生、研究生的教材,也可供各领域需要进行数据分析处理的实际工作者自学参考。

图书在版编目(CIP)数据

应用统计学 / 王淑芬主编. —北京:中国林业出版社:北京大学出版社, 2007.2

(21 世纪全国应用型本科财经管理系列应用规划教材)

ISBN 978-7-5038-4748-6

I. 应… II. 王… III. 应用统计学 - 高等学校 - 教材 IV.C8

中国版本图书馆 CIP 数据核字(2007)第 023149 号

书 名: 应用统计学

著作责任者: 王淑芬 主编

策划编辑: 李 虎

责任编辑: 周宏钦 肖基洪

标准书号: ISBN 978-7-5038-4748-6

出 版 者: 中国林业出版社(地址:北京市西城区德内大街刘海胡同 7 号 邮编:100009)

网址: <http://www.cfph.com.cn> E-mail: cfphz@public.bta.net.cn

电话: 编辑部 66170109 营销中心 66187711

北京大学出版社(地址:北京市海淀区成府路 205 号 邮编:100871)

网址: <http://www.pup.cn> <http://www.pup6.com> E-mail: pup_6@163.com

电话: 邮购部 62752015 发行部 62750672 编辑部 62750667 出版部 62754962

印 刷 者: 北京市昌平百善印刷厂

发 行 者: 北京大学出版社 中国林业出版社

经 销 者: 新华书店

787mm×960mm 16 开本 24.75 印张 480 千字

2007 年 2 月第 1 版 2007 年 2 月第 1 次印刷

定 价: 32.00 元

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,侵权必究

举报电话: 010-62752024

电子邮箱: fd@pup.pku.edu.cn

前 言

在人类迈进知识经济时代的今天,面对日益增多的信息需求量和处理量,学习和掌握数据处理和分析技术的统计方法就显得十分重要。应用统计学就是一门运用统计学的原理和方法,研究各领域中有数据搜集、整理、分析和解释的科学,是对客观现象进行定量分析的重要工具。根据多年的教学实践经验和科学研究,并考虑到 21 世纪财经管理专业本科生应用统计学课程教学改革和实际应用的需要,我们编写了这本教材。

本书从实用性的角度出发,结合 SPSS 统计分析软件,系统地阐述了统计学的基本原理和方法。在统计内容的选择和安排上,本书以统计知识介绍为主、软件操作介绍为辅,加强了推断性统计,也适当增加了多元统计分析的内容。在编写过程中,本书的主要指导思想有以下 4 点:

第一,注重统计原理、方法和应用的有机结合。考虑到财经管理专业本科生的数学基础和教学特点,对统计内容力求达到整体框架合理,内容上循序渐进,阐述通俗易懂、思路清晰,且具有启发性、可操作性强的目的。同时,强调各种统计方法的实际背景和统计思想。

第二,注重统计软件的先进性和输出结果的解读。统计软件选用当今世界上最流行的软件包 SPSS 12.0,同时兼顾其他版本,并结合实际案例,每种统计方法都给出了 SPSS 统计软件的操作步骤和输出结果的解释。

第三,注重习题类型的多样化和实用性。各章节均附有一定量的思考题、练习题和上机实验题,并在书后附有参考答案。尤其是上机实验题,本书较多地采用了实际数据,旨在使读者能够巩固所学的统计知识,提高利用计算机处理数据的能力。

第四,注重案例的专业性和前沿性。本书精心选取了财经管理领域的案例,尤其案例阅读部分内容所选择的皆是相关作者公开发表的论文及科研成果,从而使读者能够真正理解应用统计学的分析方法,学会如何分析问题和解决问题。

总之,通过本书的学习,使读者既能够根据实际需要设计统计调研方案,将实际问题转化为统计问题,又能够利用 SPSS 软件的强大功能处理分析数据,并通过科学的数据解释,将统计结果应用到实践中去。

本书可作为高等院校统计、财经、管理等专业的本科生教材,也可作为非数学专业的研究生、MBA、工程硕士的教材或参考书,对各个领域中进行数据分析处理的实际工作者也极具参考价值。

本书由天津外国语学院管理学院王淑芬教授主编并负责全书整体框架的设计及全书的修改、总撰和定稿。参与编写的还有天津工业大学卯越(编写第 7、8 章)、天津外国语学院王淑芬(编写第 4、5、6、9、10、11 章)、天津外国语学院常明(编写习题答案、附录)、太原科技大学盖兵(编写第 1、2、3 章)。另外,本书的编写和顺利出版,承蒙天津外国语学院管理学院院长冯雷鸣教授、天津工业大学理学院黄东卫教授、天津昂赛瑞企业管理咨询有限公司张国华总经理、北京大学出版社及中国林业出版社编辑的热情支持和帮助,在此表示诚挚的谢意。同时,编者在引用习题和例题的过程中,曾参阅了国内外相关统计学书籍,在此特向相关作者表示衷心的感谢。

由于编写水平有限,书中难免有纰漏或不妥之处,敬请广大同仁、读者批评指正。

编 者

2007 年 1 月

目 录

第 1 章 统计学与统计数据.....1	3.1.3 中位数.....50
1.1 统计学的基本原理与内容.....1	3.1.4 分位数.....53
1.1.1 统计与统计学.....1	3.1.5 集中趋势各测定指标之间 的关系.....55
1.1.2 总体与样本.....5	3.2 离散程度的描述.....56
1.1.3 参数与统计量.....6	3.2.1 极差和四分位差.....56
1.1.4 标志与指标.....7	3.2.2 方差和标准差.....57
1.2 统计数据的来源与类型.....9	3.2.3 变异系数和异众比率.....60
1.2.1 统计数据的来源.....9	3.3 分布形态的描述.....61
1.2.2 统计数据的类型.....10	3.3.1 偏度.....61
习题.....11	3.3.2 峰度.....61
第 2 章 统计数据的收集、整理与 显示.....14	习题.....62
2.1 统计数据的收集.....14	第 4 章 SPSS 的简单应用.....67
2.1.1 统计调查方案的确定.....14	4.1 SPSS for Windows 的界面.....67
2.1.2 常用的统计调查方法.....15	4.1.1 SPSS 的主窗口.....67
2.1.3 调查问卷设计与问卷调查 表的编码.....17	4.1.2 数据编辑窗口.....67
2.1.4 统计数据的误差.....23	4.2 建立数据文件.....69
2.2 统计数据的整理.....24	4.2.1 变量的定义.....69
2.2.1 统计数据分组.....24	4.2.2 数据的输入.....71
2.2.2 频数分布.....28	4.2.3 数据的编辑.....71
2.2.3 统计表与统计图.....31	4.2.4 数据文件的合并.....72
习题.....37	4.3 用 SPSS 进行基本统计分析.....76
第 3 章 统计数据的特征描述.....41	4.3.1 数据整理.....76
3.1 集中趋势的描述.....41	4.3.2 数据分组.....79
3.1.1 平均数.....41	4.3.3 计算基本描述统计量.....83
3.1.2 众数.....47	4.3.4 频数分析.....85
	4.4 统计绘图.....91
	4.4.1 统计图的绘制.....92
	4.4.2 统计图的编辑.....100

4.5 多选题分析.....	103	6.2 一元线性回归分析.....	152
4.5.1 多选问题的分解.....	103	6.2.1 回归分析的含义.....	152
4.5.2 多选题的频数分析.....	104	6.2.2 一元线性回归模型.....	153
习题.....	109	6.2.3 参数的最小二乘估计.....	154
第5章 统计假设检验	114	6.2.4 一元线性回归的统计检验.....	157
5.1 统计假设检验的基本问题.....	114	6.2.5 利用回归方程进行预测.....	162
5.1.1 统计假设检验的基本思想.....	114	6.3 多元线性回归分析.....	163
5.1.2 统计假设检验的基本步骤.....	116	6.3.1 多元线性回归模型.....	163
5.1.3 单边检验.....	117	6.3.2 参数的最小二乘估计.....	164
5.1.4 统计假设检验中的两类判 断错误.....	118	6.3.3 多元线性回归的统计检验.....	164
5.1.5 统计假设检验中的 P 值.....	118	6.4 非线性回归分析.....	166
5.2 正态总体均值和方差的统计假 设检验.....	120	6.4.1 非线性回归分析概述.....	166
5.2.1 单样本的 t 检验.....	120	6.4.2 非线性回归模型及其线性 化方法.....	167
5.2.2 两个独立样本的 t 检验.....	121	6.5 用 SPSS 进行相关及回归分析.....	168
5.2.3 两个配对样本的 t 检验.....	122	6.5.1 相关分析 SPSS 操作步骤.....	168
5.2.4 两个独立样本的 F 检验.....	124	6.5.2 回归分析 SPSS 操作步骤.....	170
5.3 单因素方差分析.....	125	6.5.3 曲线回归分析 SPSS 操作步骤.....	176
5.3.1 方差分析的基本概念.....	125	习题.....	180
5.3.2 单因素方差分析的基本 原理.....	125	第7章 聚类分析	190
5.4 用 SPSS 进行统计假设检验.....	131	7.1 聚类分析概述.....	190
5.4.1 One-Sample T Test 过程.....	131	7.1.1 什么是聚类分析.....	190
5.4.2 Independent-Samples T Test 过程.....	132	7.1.2 聚类分析的目的.....	190
5.4.3 Paired-Samples T Test 过程.....	134	7.1.3 聚类分析的基本步骤.....	191
5.4.4 One-Way ANOVA 过程.....	135	7.2 分类统计量.....	191
习题.....	137	7.2.1 数据资料矩阵的标准化 处理.....	191
第6章 相关分析与回归分析	147	7.2.2 距离和相似系数.....	192
6.1 相关分析.....	147	7.3 系统聚类法.....	194
6.1.1 变量间的关系.....	147	7.3.1 类间距离.....	194
6.1.2 相关分析.....	148	7.3.2 系统聚类分析.....	196
		7.4 用 SPSS 进行聚类分析.....	199



7.4.1 聚类分析 SPSS 操作过程	199	9.4 用 SPSS 进行对应分析	270
7.4.2 聚类分析 SPSS 输出结果 解释	202	9.4.1 对应分析数据的预处理	270
习题	214	9.4.2 对应分析 SPSS 操作步骤	272
第 8 章 主成分分析与因子分析	219	习题	287
8.1 主成分分析	219	第 10 章 时间序列分析和预测	292
8.1.1 主成分分析的数学模型	219	10.1 时间序列分析概述	292
8.1.2 主成分分析的步骤及应用	222	10.1.1 时间序列的相关概念	292
8.2 因子分析	227	10.1.2 时间序列分析的一般步骤	294
8.2.1 因子分析的数学模型	227	10.2 时间序列预测方法	295
8.2.2 因子载荷	228	10.2.1 趋势外推法	295
8.2.3 因子旋转与因子得分	231	10.2.2 指数平滑法	305
8.2.4 因子分析的步骤及应用	233	10.2.3 时间序列分解法	310
8.3 主成分分析和因子分析的区别	237	10.3 用 SPSS 进行时间序列分析	316
8.4 用 SPSS 进行因子分析	238	10.3.1 时间序列的序列图的绘制	316
8.4.1 因子分析 SPSS 操作过程	238	10.3.2 时间序列的预处理	318
8.4.2 因子分析 SPSS 输出结果 解释	240	10.3.3 指数平滑 SPSS 操作步骤	323
习题	252	10.3.4 时间序列分解 SPSS 操作 步骤	326
第 9 章 对应分析	258	习题	332
9.1 对应分析概述	258	第 11 章 案例阅读	339
9.1.1 问题的提出	258	案例 1 天津市房地产市场调查 问卷设计	339
9.1.2 对应分析的基本思想	259	案例 2 天津市城市居民消费 结构变动的因子分析	350
9.2 对应分析的原理和方法	259	案例 3 全国百强县教育与经济 协调发展的统计分析	355
9.2.1 对应分析中的数据变换 方法	259	案例 4 对应分析在天津市城市 居民住房消费中的应用	361
9.2.2 对协方差矩阵 R 与 Q 进行 因子分析	263	习题	365
9.2.3 绘制变量和样品的对应 分布图	263	附录 常用统计表	366
9.3 对应分析的计算与应用	264	习题答案	373
9.3.1 对应分析的基本计算步骤	264	参考文献	386
9.3.2 对应分析的应用	265		

第 1 章 统计学与统计数据

统计学是一门通过定量认识的方法对总体现象进行研究的科学，其目的在于探索客观现象内在的数量规律性，从而认识客观事物内在的本质的规律性。统计学是在统计工作实践的基础上产生的，而它一经形成，又对统计工作中的数据统计、资料搜集、整理、分析等起到理论指导作用，并且得到不断的丰富和发展。

1.1 统计学的基本原理与内容

1.1.1 统计与统计学

在日常生活中，人们经常会接触到“统计”这一术语，在有关媒体中也经常会看见一些报道使用统计数据、图表等。例如，大学每学年末要统计任课教师的教学工作量；直播篮球比赛时电视屏幕上不时会出现球队的技术统计数据；对某城市的综合竞争力进行分析和评价研究，则要构造影响该城市的综合竞争力的指标体系，并搜集相关统计数据来进行研究。可以这样说，统计已经渗透到日常生活、社会经济活动和科学研究的方方面面，统计无处不在，并且正在发挥越来越重要的作用。

1.1.1.1 统计的含义

“统计”作为一种社会实践活动已有悠久的历史。据历史记载，我国在西周时期就已经建立了统计制度。在英文中，统计为 statistics，它与“国家”为同一词根。可以说，自从有了国家，就有了统计活动，而最初的统计活动就是为统治者管理国家的需要而进行的搜集资料的工作，涉及计算国家的人力、物力和财力等。随着社会经济和科技的发展以及统计学自身的进步，统计的应用领域不断扩大。现在，统计不仅被用于经济管理领域，而且在其他许多领域也得到了广泛的应用。

那么，究竟何为统计？这里有必要给出一个比较准确的科学定义。所谓统计，就是人们认识客观世界总体数量变动关系和变动规律的活动的总称，是人们认识客观世界的一种有力工具。统计主要有以下 3 个方面的含义。

1. 统计工作(statistical work)

统计工作即统计实践,是对社会经济现象客观存在的现实数量方面进行搜集、整理和分析的活动过程,属于统计中最基础的工作。统计工作全过程分为4个阶段:

1) 统计设计

统计设计是指根据研究对象、内容和目的对整个统计过程的各个方面和各个环节进行全盘考虑和安排,同时提出收集、整理和分析数据的方案和工作进度等。统计设计的主要内容是指标设计。统计的目的是反映总体的数量特征,因此设计相应的指标来反映总体的数量特征是首要的任务,否则下面的工作无从做起。

统计设计是整个统计研究的前期工程,其完成质量直接关系到整个统计研究的质量。

2) 统计调查

经过统计设计形成方案之后,就可以确定统计调查方法,搜集数据。调查是在社会经济统计中获得原始数据的主要手段。随着市场经济的发展,调查在经济活动中所起的作用越来越大,企业的经营、政府的决策,都离不开调查的第一手数据。如何科学地进行调查是统计学研究的重要内容。

3) 统计整理

原始的统计数据收集后还必须经过整理、加工才能发挥其作用。统计整理就是对所收集到的数据资料进行审核、汇总,使之科学化、系统化、条理化,并且可以用各种统计图表表示整理后的结果。统计调查的主要内容是统计分组,这是一种重要的统计方法。

4) 统计分析

统计分析就是通过统计方法研究数据,从数据中得出规律性的结论。统计分析的主要内容是统计分析报告,统计人员要写出高质量的统计分析报告,不仅需要统计人员对实际问题的掌握和对统计分析理论和方法的熟知程度与运用能力,而且更需要具备大量的相关知识,如国家政治、经济、法律等。

统计工作的4个阶段的关系是:统计设计是统计调查的前提;统计调查是统计的基础工作,没有调查就没有以后的整理与分析;统计整理是统计工作的中间一环,目的是使杂乱无序的调查数据变得有序,为以后的统计分析做准备;统计分析是统计工作中最后一步,也是最重要的一步,分析结果可以作为决策的参考。

2. 统计数据(statistical data)

统计数据即统计资料,是统计工作的成果。统计数据的搜集是取得统计数据的过程,它是进行统计分析的基础。离开了统计数据,统计方法就失去了用武之地。例如,各类统计年鉴公布的反映月度、季度或者年度的经济发展情况的数据等。

统计工作和统计数据的关系是过程和成果的关系。即统计数据是统计工作提供的,是

统计活动的成果。

3. 统计学(statistics)

统计学是一门有关统计数据的科学。它是研究如何搜集、整理数据和进行数量分析、推断的一门方法论科学，是统计工作经验的总结和概括。

统计工作和统计学是实践和理论的关系。统计理论来源于统计实践，它是统计工作经验的总结和概括。反过来，统计理论又是指导统计工作的原则和方法。

总之，在统计工作、统计数据和统计学三者之中，统计数据是统计工作的成果，统计工作的水平、质量又直接影响统计数据的质量和效用。统计学和统计工作是理论和实践的关系，即统计学是在不断对统计工作的经验加以总结，而得到发展。

1.1.1.2 统计学的含义

虽然人类统计的实践活动可以追溯到相当遥远的古代，但是，将统计从实践上升到理论并加以总结和概括，使之成为一门科学——统计学，距今却只有 300 多年的历史。综观统计学的发展历史，可以发现，统计学最初是从设置指标研究社会经济现象的数量开始的。随着社会的发展，为了适应实践的需要，统计方法和理论不断丰富和完善，统计学也在不断发展和演变。从当前世界各国的状况来看，统计学已经成为研究社会经济现象和自然现象数量方面的有力工具，它既研究确定现象的数量方面，也研究随机现象的数量方面。统计学的作用与功能从描述事物现状、反映事物规律，向进行抽样推断、预测未来变化的方向发展，统计学自身也从单一的实质性社会科学演变成横跨社会科学领域和自然科学领域的多科性的方法论科学。

1. 统计学的定义

统计学是一门收集、整理、分析和解释统计数据的方法论科学，其目的是探索数据的内在数量规律性，以达到对客观事物的科学认识。也可以这样说，**统计学**是关于数据的科学，其内容包括数据的收集、分类、汇总、组织、分析、推断和解释。

由统计学的定义知道，统计离不开数据。统计研究的过程首先要有数据，在拿到数据后，为满足分析的需要，还要对数据进行一定的整理，而后再对数据进行分析 and 解释。统计研究的过程如图 1.1 所示。

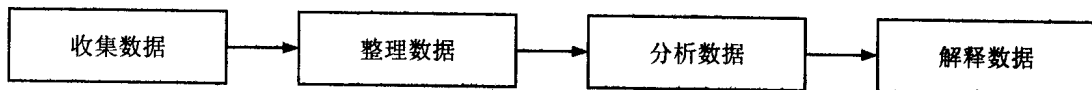


图 1.1 统计研究的过程

2. 统计学的分类

根据统计学的方法的构成, 可以将统计学分为描述统计学和推断统计学。

1) 描述统计学(descriptive statistics)

描述统计学研究如何取得反映客观现象的数据, 并以图表的形式对所收集的数据进行加工处理和显示, 进而通过综合、概括与分析, 得出反映客观现象的规律性特征。内容主要包括统计数据的收集方法、数据的加工处理方法、数据的显示方法、数据分布特征的概括与分析方法等。描述统计学属于初等统计学。

2) 推断统计学(inferential statistics)

推断统计学主要是研究如何根据样本信息来推断总体的特征, 所应用的知识主要是概率论与数理统计, 属于较高级的统计学。在有些情况下, 人们获得的统计资料并非事物整体的状况, 而是来自事物的一个局部。在保证推断具有一定的有效性和可靠性的前提下, 如何利用局部的数据去推断整体的情况, 即是推断统计学所要研究的内容。它包括抽样分布理论、参数估计、假设检验、方差分析、回归分析及多元统计分析等。

描述统计学不必深入一层地去试图推论数据本身以外的任何事情, 而推断统计学则在样本数据的基础上深入一步地分析、研究和推断, 以推知资料本身外的情况和数量关系; 描述统计学用的是总体数据, 而推断统计学则往往用样本数据。在现实问题中, 得到的数据主要是样本数据, 因此, 推断统计学越来越重要, 是统计学的核心内容。统计学的发展过程中, 先有描述统计学, 后有推断统计学, 从描述统计学发展到推断统计学, 是统计学发展成熟的标志。

此外, 根据统计学的学科分类, 还可以将统计学分为理论统计学和应用统计学。

理论统计学(theoretical statistics)是指统计学的基本原理, 主要研究统计学的一般理论问题, 尤其是各种统计方法的数学理论问题。

应用统计学(applied statistics)是研究如何应用统计方法去解决实际问题的, 应用统计学一般都与特定的领域相联系。例如, 统计学在教育领域的应用, 称为教育统计学; 在经济领域的应用, 称为经济统计学; 等等。

3. 统计学研究对象的特点

1) 数量性

统计最基本的特点就是以数字为语言, 用数字说话。具体地说, 就是用规模、水平、速度、结构和比例关系等, 去描述和分析客观事物的数量表现、数量关系和数量变化, 揭示事物的本质, 反映事物发展规律, 推断事物发展前景。

2) 具体性

统计所研究的量不是抽象的量, 而是与客观事物的质密切相联系的量, 是体现事物相



互关系和发展变化的量,具有明显的时空特点和事物属性的特点,这一点是统计学与数学最本质的区别。因此,具体性是指除数字外,还要有说明该数字所表示的内容、所指的时间(或时点)、所指的空间以及计量单位和计算方法,这就是后面要讲的统计指标的构成要素。只有这样统计的数字,才有意义。

3) 综合性

作为认识客观事物的统计,是从总体上来认识其数量特征,它虽然也研究个体,但是其目的是为了通过个体来推断总体。在实际中,综合性体现为要把研究对象作为一个整体来描述,揭示或推断它的数量特征。

1.1.2 总体与样本

1. 总体(population)

通常,总体是指根据研究目的确定的所要研究的同类事物的全体。它通常是由具有某种共同性质的许多个体组成。例如,全国高校教师组成的全体可以作为一个总体,而其中的每位教师就是个体;天津市所有国有企业的全体也可以作为一个总体,而其中的每个企业就是个体。

在研究问题时,人们对于总体中的人或事物本身并不关心,感兴趣的是表征总体状况的人或事物的某一个或某几个数量指标的分布特征。例如,要检验一批产品的合格率,这一批产品的全体就是总体。但在统计研究中,只是关心这批产品的合格率,而不是产品本身。因此,也可以把这批产品的合格数的全体作为一个总体,这时总体就是一组观测数据。这就是数理统计中所涉及的总体的概念。

在数理统计中,总体是指研究对象的某项数量指标的值的全体。总体中的每一个可能观察值称为个体,它是某一随机变量的值。这样,一个总体对应于一个随机变量,因此总体可用一个随机变量 X 或 Y 来表示。例如,要研究全国高校教师的收入情况,就将全国高校教师的收入的全体作为一个总体 X ,而其中每一位教师的收入就是个体,它是总体 X 的一个可能观察值;要研究某大学本科生的年消费情况,就将该大学本科生的年消费的全体作为一个总体 Y ,而其中每一位本科生的年消费是个体,它是总体 Y 的一个可能观察值。

总体根据所包含的个体的数量是否有限可以分为有限总体和无限总体。有限总体是由有限的个体构成的总体。例如,全国高校教师的收入就是有限总体。无限总体是有无限个个体构成的总体。例如,在科学实验中,每一个实验数据可以看作是一个总体的个体,而实验则可以无限地进行下去,因此由实验数据构成的总体就是一个无限总体。

2. 样本(sample)

统计研究的目的是要确定总体的数量特征。但是,当总体中的个体数量很多甚至无限

时,不必要也不可能对构成总体的所有个体都一一进行调查。因此常常从总体中抽取一部分个体进行研究,进而根据所抽得的部分个体的数量特征来推断总体。

样本就是由总体中抽取部分个体组成的集合,构成样本的个体的数目称为**样本容量**,也叫**样本大小**。

在数理统计中,所谓从总体中抽取一个个体,就是对总体 X 进行一次观测并记录其结果。在相同的条件下对总体 X 进行 n 次重复的、独立的观察,将 n 次观察结果按实验的次序记为 $X_1, X_2, \dots, X_n$, 有理由认为 $X_1, X_2, \dots, X_n$ 是相互独立的,且都是与总体 X 具有相同分布的随机变量。这样得到的 $X_1, X_2, \dots, X_n$ 称为来自总体 X 的一个**简单随机样本**, n 称为**样本容量(或样本大小)**。一个样本对应于一组随机变量。以后如无另加说明,所提到的样本都是指简单随机样本。

当 n 次观察一经完成,我们就得到一组实数 $x_1, x_2, \dots, x_n$, 它们依次是样本 $X_1, X_2, \dots, X_n$ 的观测值,称为**样本观测值**,简称**样本值**。

需要说明的是,前面对于有限总体和无限总体的分类,对于抽样有着重要的意义。从无限总体中抽取的样本可以看作是独立的,而从有限总体中抽取的样本将受到前一次所抽取的个体的影响。若使得所抽取的样本是简单随机样本,对于有限总体,采用放回抽样就可以做到,但放回抽样使用起来不方便,因此,当总体中个体的总数 N 比要得到的样本的容量 n 大得多时,在实际中可将不放回抽样近似当作放回抽样来处理。

也就是说,当总体中个体的总数 N 比要得到的样本的容量 n 大得多时,不管采用什么样的抽样方式,所得到的样本都可以近似看成简单随机样本,即样本满足独立性及与总体同分布的性质。

1.1.3 参数与统计量

1. 参数(parameter)

参数是指研究者想要了解的总体的某种特征值,有时候又称为**总体指标**。通常所要研究的参数主要有总体平均数、方差、总体比例等,用希腊字母来表示。如总体平均数用 μ 表示,总体方差用 σ^2 表示,总体比例用 π 表示。

由于总体的某种特征值通常是未知的,所以参数是一个未知的常数。比如不知道全国高校教师的平均收入以及所有高校教师的收入差异有多大;不知道某工厂所生产的一批产品的合格率;等等。正因为如此,才进行抽样,根据样本计算出某些值去估计总体参数。

2. 统计量(statistic)

统计量是由样本 $X_1, X_2, \dots, X_n$ 构成的一个新的函数 $T(X_1, X_2, \dots, X_n)$, 不含未知参数。统计量又称为**样本指标**。通常所要研究的样本指标主要有样本平均数、样本方差、样

本比例等,用英文中 26 个字母来表示。如样本平均数用 $\bar{X}$ 表示,样本方差用 S^2 表示,样本比例用 P 表示。

由于样本是已经抽取出来的,所以统计量总是可以计算出来的具体值。抽样的目的就是用样本指标去估计总体指标。比如用样本平均数 $\bar{X}$ 去估计总体平均数 μ ,用样本方差 S^2 去估计总体方差 σ^2 ,用样本比例 P 去估计总体比例 π ,等等。

需要注意的是,总体各指标都是未知的,但是却是唯一确定的,是确定的变量。而样本指标随着抽样的不同而发生变化,因而都是随机变量。

1.1.4 标志与指标

1. 标志(symbol)

标志是用来说明总体中各个个体普遍具有的属性或特征。每个总体中的个体从不同方面考察都具有许多属性和特征。例如,每个学生都具有性别、年龄、身高、体重等属性和特征,这些就是学生作为总体的每个个体的标志。

标志分为品质标志和数量标志两种。总体中个体的属性特征称为**品质标志**,这类标志的变异不能用数值表示。例如,在天津市的所有工业企业总体中,工业企业的名称、所有制、所在区县等都是品质标志。总体中个体的数量特征称为**数量标志**,这类标志的变异能用数值表示。例如,在天津市的所有工业企业总体中,工业企业的职工人数、固定资产、工业增加值等都是数量标志。

每个总体中的个体既有品质标志,又有数量标志,它们在统计研究中所采用的方法是不同的。

2. 指标(indication)

指标是反映总体数量特征的概念及其数值。它是利用科学的统计方法,对总体中各个个体的数量标志进行综合汇总而形成的。一项完整的统计指标由总体范围、时间、地点、指标数值和数值单位等构成。它具有以下方面的特征:第一,统计指标是一定社会经济范畴的具体表现;第二,统计指标具有可量性;第三,统计指标具有综合性。

例如,在某高等学校本科教学评估中,人们可以查阅到一系列指标。如 2003 年某大学具有高级职称教师的人数为 150 人,有 3 个一级学科硕士点,有 8 门市级精品课程等。这些指标从某一侧面反映了该大学的教学水平的数量特征。

指标按所反映的数量特点不同可分为数量指标和质量指标。

1) 数量指标(quantitative indication)

凡是说明总体规模大小、水平高低、数量多少的指标都称为**数量指标**。由于它反映的是现象的总量,因此也称为总量指标,用绝对数表示,它是汇总的结果。如在天津市工业

企业总体中,天津市工业企业总数、职工人数、工业总产值、商品进出口总额、能源消耗总量等,这些指标反映了现象或过程的总规模和水平。

2) 质量指标(qualitative indication)

凡是说明总体相对水平或工作质量的统计指标都称为**质量指标**,用相对数或平均数表示,它是总量指标的派生指标。如在天津市工业企业总体中,企业劳动生产率、职工总平均工资、工人出勤率、百元产值总能耗等,这些指标用来反映现象之间的内在联系和对比关系,更深刻地阐明现象发展的规律性。

单个指标只能反映总体的某一个数量特征,说明现象的某一个侧面的情况。然而,客观现象是错综复杂的,要反映全貌,描述现象发展的全过程,只靠单个指标是不够的,需要建立统计指标体系。

指标体系(indicative system)是各种相互联系的指标群所构成的整体,用以说明所研究的社会经济现象各方面相互依存和相互制约的关系。例如,为了评估某省各区县的经济水平,需要设立年末总人口、年末全部就业人口数、失业人口占总人口的比例、国内生产总值、人均国内生产总值、第一产业增加值、粮食总产量、工业增加值、社会固定资产投资、社会消费零售总额、地方财政预算收入、城乡居民储蓄存款余额、年末金融机构贷款余额、农民人均纯收入、全部职工年平均工资,第一、二、三产业占总从业人数的百分比等多项指标,来组成该省各区县的经济水平评估指标体系。

指标体系的设置不但是客观现象的反映,也是人们对客观认识的结果。确定指标体系必须有一定的理论依据,同时又必须对计算口径加以具体化,以便达到量化的目的。在指标筛选过程中,应尽量遵循目的性、互补性和可操作性原则。目的性体现在:由于研究目的不同,指标选择自然不同,在研究中应选择与研究目的明显相关的指标并剔除那些与研究目的无关或关系不明确的指标。互补性体现在:力求各项指标分别能反映客观现象的某一方面,综合起来能反映其全貌。可操作性体现在:根据所设计的指标体系,必须能收集到一些必要的数,否则必须放弃或更换相应的指标,反过来重新设置指标体系。

标志与统计指标之间存在着明显的区别和紧密的联系。

区别:

(1) 反映的事物范围大小不同。标志说明总体单位的属性特征;统计指标反映的是总体的数量特征。

(2) 表述形式不同。标志既有能用数值表示的数量标志,又有只能用文字表示的品质标志;统计指标都是用数字来表示的。

联系:

(1) 标志与统计指标之间具有对应关系。标志是统计指标的核算基础,它与统计指标的指标名称往往是同一概念,有相互对应的关系。