

GAODENGYUANXIAO

TONGJIXUE

JINGPINJIAOCAI



高等院校统计学精品教材

统计学原理

编 著/赵 煜 申社芳
郭精军 高海燕



中国统计出版社
China Statistics Press

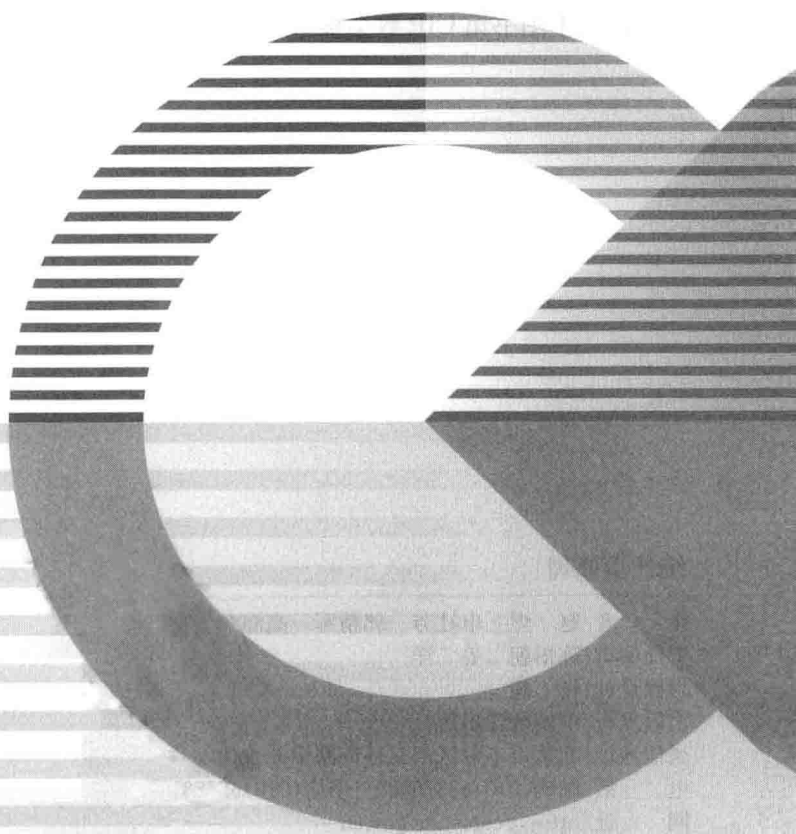
GAODENGYUANXIAO TONGJIXUE JINGPI

015006870

高等院校统计学精品教材

统计学原理

编 著/赵 煜 申社芳
郭精军 高海燕



中国统计出版社
China Statistics Press

图书在版编目(CIP)数据

统计学原理 / 赵煜等编著. — 北京 : 中国统计出版社, 2014.5

高等院校统计学精品教材

ISBN 978-7-5037-7074-6

I. ①统… II. ①赵… III. ①统计学—高等学校—教材 IV. ①C8

中国版本图书馆 CIP 数据核字(2014)第 079632 号

统计学原理

作 者/赵 煜 申社芳 郭精军 高海燕

责任编辑/陈悟朝 姜 洋

封面设计/杨 超

出版发行/中国统计出版社

通信地址/北京市丰台区西三环南路甲 6 号 邮政编码/100073

电 话/邮购(010)63376909 书店(010)68783171

网 址/<http://csp.stats.gov.cn>

印 刷/河北天普润印刷厂

经 销/新华书店

开 本/787×1092mm 1/16

字 数/500 千字

印 张/21.5

印 数/1—2000 册

版 别/2014 年 5 月第 1 版

版 次/2014 年 5 月第 1 次印刷

定 价/42.00 元

版权所有。未经许可,本书的任何部分不得以任何方式在世界任何地区以任何文字翻印、仿制或转载。

中国统计版图书,如有印装错误,本社发行部负责调换。

前言

统计学是帮助我们有效利用信息并进行判断决策的一门方法论学科,其主要目的就是以科学的方法来收集、分析、推断和应用统计信息,以达到以简驭繁,清楚客观地认识我们所面对的不确定性现象。当前,越来越多的人认可统计学是除哲学、数学以外,人类认知客观世界的第三种重要方法。这种通过对数据的研究,达到抽象出有关客观现象数量方面特征经验结论的统计分析方法已经越来越广泛地应用于科学研究、生产经营管理和普通人的日常生活之中,具有统计思维方式已是现代社会发展的必然趋势。因此,统计学原理作为高等院校统计学专业的基础课程,其重要作用是不言而喻的。

本书是一本为高等院校统计学专业本科学生介绍统计学基本原理的教材,也可作为财经类院校非统计专业学生学习统计学的教材或参考书。作为一本介绍原理的教材,为达到开蒙领航的作用,本书内容涵盖统计学基本原理和应用方法两个方面。编写过程中更注重在统计思想、观点和概念的阐述上,力求做到以统计学科在各领域中的普遍应用为出发点,以解决实际问题为导向,重点培养学生形成用正确的统计观点去观察和研究事物的能力和习惯,进而初步建立统计思维体系。全书共分十章,以统计工作流程为主线,顺序介绍描述统计(统计信息的收集与提取)与推断统计(统计估计与统计假设检验)这两个核心环节。进一步结合经济统计经典方法,系统介绍了统计分析基本工具(统计变量体系、统计图表、统计模型)的应用原理。

本书由赵煜、申社芳、郭精军、高海燕四人分工合作完成。具体分工如下:第1、2、4章由赵煜编写;第3、9、10章由申社芳编写;第5、6章由郭精军编写;第7、8章由高海燕编写。在本书的编写过程中,兰州商学院统计学院庞智强教授、王永瑜教授、郭海明副教授

给予了大力支持,统计学院所有教师均对本书的内容提出了许多宝贵的修改意见,编者参考和引用了国内外专家学者的有关著作,同时也参考了一些网络资源提供的内容,在此表示衷心感谢!本书曾以讲义的形式在兰州商学院统计学院 2009—2011 级统计学专业统计学原理课堂上使用.感谢各位同学积极投入这门课程的学习和实践,感谢及时反馈诸多有价值的意见和建议.特别感谢崔琳爽、王茹两位同学对本书内容的积极建言和认真核对.

由于编者学识水平及教学经验有限,书中不妥或错漏之处在所难免,恳请专家及读者给予批评指正.

编 者

2013 年 12 月

目 录

第 1 章 导论	1
§ 1.1 什么是统计	1
§ 1.2 统计学基本概念	3
§ 1.3 统计学研究体系	8
§ 1.4 统计学的产生与发展	12
本章小结	15
思考与实践	16
习题 1	16
第 2 章 统计学的概率基础	18
§ 2.1 统计学常用分布	18
§ 2.2 总体分布与样本分布	27
§ 2.3 统计量	28
§ 2.4 抽样分布	33
本章小结	41
思考与实践	41
习题 2	42
第 3 章 统计调查与整理	45
§ 3.1 统计调查的设计	45
§ 3.2 统计数据的搜集	49
§ 3.3 统计数据的质量	57
§ 3.4 统计数据的整理	60

本章小结	71
思考与实践	72
习题 3	72
第 4 章 样本信息的提取	77
§ 4.1 位置信息的提取	77
§ 4.2 尺度信息的提取	88
§ 4.3 分布形态信息的提取	92
§ 4.4 相对信息的提取	94
§ 4.5 统计图	95
本章小结	109
思考与实践	109
习题 4	110
第 5 章 参数估计	115
§ 5.1 点估计	115
§ 5.2 点估计优劣的评价标准	120
§ 5.3 区间估计	126
§ 5.4 单侧置信限	136
§ 5.5 贝叶斯估计	138
本章小结	143
思考与实践	144
习题 5	145
第 6 章 假设检验	147
§ 6.1 假设检验概述	147
§ 6.2 单个正态总体参数的假设检验	150
§ 6.3 两个正态总体参数的假设检验	155
§ 6.4 非正态总体参数的假设检验	159
§ 6.5 假设检验的引申问题	161
§ 6.6 关于分布的检验	165
本章小结	171
思考与实践	174

习题 6	174
第 7 章 方差分析	178
§ 7.1 方差分析概述	178
§ 7.2 单因素方差分析	182
§ 7.3 多重比较	192
§ 7.4 双因素方差分析	195
§ 7.5 试验设计初步	203
本章小结	208
思考与实践	210
习题 7	210
第 8 章 相关与回归分析	214
§ 8.1 相关关系	214
§ 8.2 相关分析	217
§ 8.3 回归分析的基本问题	225
§ 8.4 一元线性回归	230
§ 8.5 可线性化的非线性回归	247
本章小结	250
思考与实践	251
习题 8	251
第 9 章 时间序列分析	254
§ 9.1 时间序列分析的基本问题	255
§ 9.2 时间序列的指标分析	258
§ 9.3 时间序列的结构分析	267
本章小结	280
思考与实践	281
习题 9	281
第 10 章 统计指数	286
§ 10.1 对比分析	286
§ 10.2 指数的概念和种类	289

§ 10.3 综合指数	292
§ 10.4 平均指数	297
§ 10.5 指数体系和因素分析	300
§ 10.6 几种常用的经济指数	305
本章小结	311
思考与实践	312
习题 10	312
附录 统计用表	318
参考文献	334

第1章 导论

在终极的分析中,一切知识都是历史;在抽象的意思下,一切科学都是数学;
在理性的基础上,所有判断都是统计学。——C. R. Rao

众所周知,当今的我们身处信息时代,报纸、电视、网络等各类媒体每天都会把各种各样的信息推送到我们面前。事实上,为了解某一现象或者为达到某种目的,我们经常需要立足于这些信息,为进一步的行动进行判断。那么,如何去芜存菁,让这些信息为我们所用,是我们面临的一大挑战。正如统计学家 C. R. Rao 在其论著《统计与真理》中所述“在理性的基础上,所有判断都是统计学”,统计学就是帮助我们有效利用信息并进行判断决策的一门方法论学科,其主要目的就是以科学的方法来收集、分析、推断和应用统计信息,以达到以简驭繁,清楚客观地认识我们所面对的不确定性现象。

§ 1.1 什么是统计

1.1.1 生活中的统计问题

统计及相关话题存在于我们生活的方方面面,我们总是自觉或不自觉地借用统计数据、统计方法来描述对事物的看法或论证自己的观点,以下问题将引领我们步入统计的殿堂。

【例 1.1.1】 无处不在的统计

(1)食品安全是老百姓关注的话题,欲了解市场上某种食品的某种添加剂的含量是否符合国家标准,我们首先考虑到的是对其进行调查。那么,怎样展开调查呢?可不可以选取几个市场进行研究?如果可以的话,在选定的市场中又该选取多少该种食品进行调查呢?——抽样调查问题

(2)农民在种地之前,会选择适当的种子。假如一个农业试验站共有五种小麦种子可供选择,请问此试验站应向农民提供哪种小麦种子以备来年的播种?——试验设计问题

(3)为了解现阶段护士对于工作的满意程度,某医护协会对全国的医院护士进行了一项调查,获得一组对工作、工资和升职机会满意程度评分的数据,那么如何整理与分析这组数据呢?——统计信息的提取问题

(4)寿命试验是破坏性试验,只能抽取少量产品做试验。例如,某工厂生产一种晶体管,由于在批量生产中各种随机因素的干扰,所生产的晶体管寿命不同。如何估计其平均寿命呢?——参数估计问题

(5)某厂进行技术革新,怎么考察新、旧技术是否存在差别呢?差别是技术革新造成的还是由于随机误差造成的呢?鉴别方法是什么?——假设检验问题

(6)为了探讨吸烟与患慢性支气管炎是否有关联,通过调查获得一组被调查者是否吸烟及是否患有慢性支气管炎的频数数据,问:从这批数据能否断定患慢性支气管炎与吸烟有关?——变量间的关系问题

(7)股票是大家常用的投资工具之一,如果把每天的收盘价记录下来,则人们关心的问题就是如何利用过去 n 天的收盘价预测未来?——时间序列的预测问题

(8)从遗传学的角度看,遗传会把一种性状(如身高)的优势传递给下一代.如果真是这样的话,人类的身高将会出现两级分化的现象.但实际上,人类身高相对稳定地表现为正态分布,如何解释这种情况?——回归分析问题

(9)当你买了一台电脑时,被告知三年内可以免费保修.那么,这个保修期是怎样决定的呢?——质量控制问题

上述问题来自我们生活的各个方面,其解决之道,正是统计学研究的主要内容.显然,无论是在自然科学还是人文社会科学的研究领域中,统计学都有着十分广泛的应用,统计方法已成为人类认识世界、控制实践活动和进行科学研究的重要工具之一.

1.1.2 统计的含义

在各种实践活动和科学研究领域中经常出现的“统计”一词,一般具有三种含义:统计活动、统计数据和统计学.

统计活动也称统计工作,是指收集、整理和分析统计数据,提取统计信息,并探索研究对象内在数量规律性的实践过程.例如,开会时主持人要统计一下出席会议的人数,企业每年要“统计”产量和产值,体育比赛要进行技术“统计”,电视台收视率的调查等,这些均属于最基础的统计实践活动.

统计数据也称统计资料,是通过统计活动获得的各种资料的总称,其中蕴含着大量丰富的有关研究对象的数量信息.统计数据的表现方式多种多样,如科学试验或社会经济现象数量特征的原始记录、统计表、政府统计公报、统计年鉴等各种数字和文字资料.

统计的第三种含义即作为一门科学的“**统计学**”,它系统阐述统计工作基本理论和基本方法,是对统计工作实践的理论概括和经验总结,也是本书将要探讨的主要内容.

统计活动、统计数据和统计学之间有着密切联系,统计学以统计数据为研究的切入点,是统计实践活动的理论概括和科学总结,而统计活动与统计数据之间则是过程与成果之间的关系.统计活动、统计数据和统计学相互依存、相互联系,共同构成了一个完整的体系,这就是我们所说的“统计”.统计就是从精心设计的方案中获得统计资料,并加以整理分析,从中提取有效信息,进而将自然和社会现象的数量特征客观理性呈现,从而使人们能在更清晰地了解现象本质基础上提出相应对策的全过程.

§ 1.2 统计学基本概念

1.2.1 总体与样本

1. 总体与个体

统计总体,简称**总体**,是指客观存在的在同一性质基础上结合起来的研究对象的全体,特指所研究对象的数量方面,其大小与范围随具体研究与考察的目的而确定.总体确定后,我们称构成总体的每一个元素为**个体**.总体中所包含的个体的个数称为**总体容量**,记为 N .例如,要研究全国城镇居民家庭的收入情况,则全国城镇居民家庭的收入就是一个统计总体.而构成总体的每一个具体城镇居民家庭的收入就是个体.又如,考察某大学一年级新生的体重情况,则该校一年级全体新生的体重就构成了待研究的统计总体,而个体就是一年级每一位新生的体重.

容量为有限的总体称为**有限总体**;当个体数目难以确定,其容量有可能是无限时,便构成了**无限总体**.例如,全国人口普查,尽管其包含的个体数量很大,但仍然是有限的,所以是有限总体;又如,要检验某种新工艺是否真正能够改善产品性能的问题,由于该新工艺有可能一直延续下去,利用该工艺制造的产品包括已经生产和将要生产的产品,其数量难以具体确定,因此属于无限总体.

统计学将统计总体与个体视为随机变量,习惯将总体记为 X ,第 i 个个体记为 X_i .对总体的研究就相当于对一个随机变量 X 的研究,今后将不区分总体与相应的随机变量.

根据同时研究的变量个数的多少,统计总体可分为**一维总体**和**多维总体**.如前例,研究一年级新生的身体素质状况,若关注身高 X 与体重 Y ,则可构成**二维总体** (X, Y) ;若同时考虑其他身体素质变量诸如血压、肺活量之类,则可构成**三维**、**四维**等其他**多维总体**.本书主要介绍一维及二维总体的研究方法.

2. 样本

统计研究关注对统计总体规律性的把握,由于总体往往在时间或空间上具有不可穷尽性,因此不必要或不可能对构成总体的所有个体都进行研究.这时,需要采用一定的方式,从总体中抽取一部分个体,作为总体的代表加以研究.

从总体中抽取部分个体的过程称为**抽样**,所抽取的部分个体构成的集合称为**样本**.例如,若在同一架天平上将一个物体称量 n 次,会得到一组结果 $X_1, X_2, \dots, X_n$,则它们的全体,即 $X = (X_1, X_2, \dots, X_n)$,就称为**样本**.在不致引起混淆的前提下,数据 $X_1, X_2, \dots, X_n$ 中的任一 X_i 也称为**样本**.样本中所含个体数目 n 称为**样本容量**.

从一维总体中获得的样本为一维样本;从二维总体中获得的样本 $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ 为二维样本;类似地,有其他高维样本.多维样本中的任意一个成分称为**样本的分量**.例如, X_1 是样本 (X_1, Y_1) 的一个分量.

统计学将来自总体 X 的容量为 n 的样本视为一个 n 维随机变量 $(X_1, X_2, \dots, X_n)$.一旦抽样发生,则可得到样本的一次具体观察值 $(x_1, x_2, \dots, x_n)$,称其为**样本值**或**样本数据**.全体样本值组成的集合称为**样本空间**.显然,样本具有二重属性.从实用者的眼光看,样本就是一片

已知的数字;用概率论的语言说,样本就是随机变量.读者在学习时即使面对样本数据,也不应忽视“样本是随机变量”这一背景.这意味着由此所发展的统计方法具有一定的普遍性,而不只是可用于某些具体样本值.否则,样本就不过是一堆杂乱无章的、毫无规律性可言的数字,无法进行任何统计处理.

【例 1.2.1】 样本的一些例子

(1)从一批机器零件毛坯中随机地抽取 10 件,测得其重量(单位:克)为:

210 243 185 240 215 228 196 235 200 199

这是一个容量为 10 的样本数据,它是来自一批机器零件毛坯重量这一总体的一个样本观察值.

(2)某大学为了解学生每天的上网时间,在全校学生中随机抽取 30 人,调查他们每天的上网时间,得到数据(单位:小时)如下:

3.3	6.2	2.1	1.9	5.5	2.5	4.2	4.5	3.4	0.8	2.5	4.1	3.2	4.4	1.3
0.5	3.6	2.8	1.6	4.7	5.1	4.3	4.5	4.0	1.7	2.4	2.5	3.0	3.2	1.5

这是一个容量为 30 的样本数据,对应的总体是该校大学生每天的上网时间.

(3)对 300 个零售商店调查周零售额的结果(单位:元)如下:

零售额	0~1000	1000~5000	5000~10000	10000~20000	20000~30000
商店数	54	115	90	30	11

这是一个容量为 300 的样本数据,对应的总体是所有零售店的周零售额.这里并没有给出每一个个体具体的观察值,而是给出了样本观察值所在的区间,称为分组样本数据.这样会导致损失一些信息,但在样本量较大时,这种经过整理的数据更能使人们对总体有一个大致的印象.

1.2.2 统计信息、统计变量与统计数据

1. 统计信息

信息是关于事物特征的表述,是事物的结构和运动状态以及与其他事物之间关系的综合反映,提取并利用信息是人类减少对未知对象的无知程度及降低不确定因素的基本方式.统计信息是信息的一种,是研究者运用科学的手段,经过一系列统计活动所取得的关于研究对象的状态、结构、功能及其发展变化规律的客观描述.统计信息反映的是研究对象的数量方面,主要包括数量的多少,现象之间的数量关系和质量互变的数量界限等.因此,数量性是统计信息的基本特征.

统计活动所涉及的信息主要有三类:总体信息,样本信息与先验信息.

总体信息,即总体分布或总体所属分布族带给我们的信息.譬如,“总体是正态分布”这一句话就给我们带来很多信息.总体信息是很重要的信息.为了获取此种信息往往耗资巨大.例如,美国军界为了获得某种新的电子元器件的寿命分布,常常购买成千上万个此种元器件,做大量寿命试验、获得大量数据后才能确认其寿命分布是什么.我国为确认国产轴承寿命分布服从两参数威布尔分布,前后也花了五年时间,处理几千个数据后才确定下来的.

样本信息,即从总体中抽取的样本所提供的信息.这是最“新鲜”的信息,人们希望通过所提取的样本信息对总体的某些特征作出较为精确的统计推断,因此,样本信息越多越好.可以

说,没有样本信息就没有统计学可言。

先验信息,即在抽样之前有关统计问题的一些信息.一般说来,先验信息主要来源于经验和历史资料.先验信息是客观存在的,在日常生活和工作中也经常可见,不少人在自觉或不自觉地使用它.以下是关于先验信息使用的两个实例.

【例 1.2.2】英国统计学家 Savage(1961)曾考察如下两个统计试验:

试验 1 一位常饮牛奶加茶的妇女声称,她能辨别先倒进杯子里的是茶还是牛奶.对此做了十次试验,她都正确地说出了.

试验 2 一位音乐家声称,他能从一页乐谱中辨别出是海顿还是莫扎特的作品.在十次这样的试验中,他都能正确辨别.

在这两个统计试验中,假如认为被试验者是在猜测,每次成功概率为 0.5,那么十次都猜中的概率为 0.0009766,这是一个很小的概率,是几乎不可能发生的.所以“每次成功概率为 0.5”的假设应被拒绝.被试验者每次成功的概率要比 0.5 大得多.这就不是猜测,而是他们的经验帮了他们的忙.可见经验(先验信息的一种)在决策中不可忽视,应加以利用.

统计总体往往是未知的,获取总体信息并加以应用是我们的终极目的,样本信息与先验信息正是帮助我们获取总体信息的有力工具.根据是否在统计研究中使用先验信息,统计学可分为经典统计学和贝叶斯统计学,而获取并借助样本信息对未知的总体信息进行推断与决策的经典统计学,正是本书的核心内容.

2. 统计变量

(1) 统计变量

变量即指取值可变的量,是说明现象某种数量特征的概念,其特点是从一次观察到下一次观察结果会呈现出差别或变化.根据取值确定与否,变量又可分为确定性变量和随机变量.

确定性变量是受确定性因素影响的变量,即影响变量值变化的因素是明确的,是可解释和可控制的;而**随机变量**则是受许多微小的不确定因素影响的变量,变量的取值无法事先确定.例如,除了某种正常的、起决定性的因素外,影响某企业生产的同一批次灯泡质量的还有许多其他因素,如果抽取一部分灯泡进行检验,这些灯泡的寿命值不尽相同,数值的大小带有偶然性的波动,检验前是不能预先确定的,这说明灯泡寿命是一个随机变量.随机变量,简单理解就是取值带有随机性的变量,其取值虽具有随机性,但它的数值变动却有一定的规律性,通过大量观察,应用统计技术方法,可以揭示和描述其数量特征以及变动的规律性.

统计变量是统计研究中所构建的各类变量的统称,是提取并分析统计信息的一个有效工具,是研究对象的具体化与数量化.其本质是随机变量.随机变量的主要特征有两个——**变异性**与**随机性**.作为变量,需研究其取值范围;作为随机变量,则需进一步研究其取各值或在某一区域内取值的概率.基于此对随机变量研究的切入点主要有两个:

其一是进行全面研究,其基本工具为随机变量的分布函数及概率分布列(概率密度函数).了解了随机变量的概率分布,即从总体上把握了其取值范围及取值的概率分布状况,就等于把握了所研究对象的全貌.

其二是进行重点研究,其基本工具是随机变量的数字特征.数字特征反映随机变量取值及分布状况的主要特征,使所研究对象的数量特征马上生动、鲜明起来,有助于我们对所研究对象进行直观的把握和具体的比较.

例如,我们说某门课程的考试成绩服从正态分布时,是对其进行全面描述,你脑海中可能

呈现一条钟形曲线;而当我们强调该门课程的考试成绩期望 $\mu=70$,方差 $\sigma^2=100$ 时,则是对其进行重点刻画,此时你脑海中的钟形曲线位置与形状完全明晰,你可能已经根据你的经验对该门课程的难易程度进行了定位.重点研究是对随机变量全面研究的有力补充.实际中,我们总是在全面研究的基础上,对随机变量进行重点研究,以概括描述所研究对象在某一特定方面表现出的数量特征.

(2) 统计变量的分类

结合变量的本质属性与取值特征,统计变量可分为**定性变量**与**定量变量**.

定性变量也称属性变量,其取值为一些性质描述,用来说明研究对象的属性特征.定性变量根据其取值的递进层次,又可以分为**定类变量**与**定序变量**.

定类变量用来说明事物的类别特征.如“性别”就是个定类变量,其取值为“男”或“女”;“行业”也是一个定类变量,其取值可以为“零售业”、“旅游业”、“汽车制造业”等.

定序变量是说明事物有序类别特征.如“产品等级”就是个定序变量,其取值可以为“一等品”、“二等品”、“三等品”、“次品”等;“受教育程度”也是一个定序变量,其取值可以为“小学”、“初中”、“高中”、“大学”等;人们对某种事物的看法往往也是一个定序变量,其取值可以为“同意”、“保持中立”、“反对”等.

定性变量的取值常常被称为变量的水平.如前例中,我们可以说“性别”有两个水平,“产品等级”有四个水平,等等.

定量变量也称数值型变量,其取值为具体的数值,用来说明事物的数量属性.例如:学生身高为160cm、175cm、183cm等.诸如“产品产量”、“商品销售额”、“零件尺寸”、“年龄”、“时间”等都是数值型变量.

根据变量值是否连续出现,定量变量又可分为**离散型变量**和**连续型变量**.离散型变量是指变量的取值为有限或无限可列个,用来表示计数过程的结果.例如,企业数,职工人数、机器设备台数等.连续型变量是指变量的取值充满数轴的某一个或某几个区间,无法一一列举,往往表示测量过程的结果.例如,气象上的温度、湿度,零件的尺寸,人的身高、体重、年龄,电子元件的使用寿命等.

本书以定量变量为主要研究对象,并简要介绍定性变量的基本研究方法.

(3) 变量间的关系

现实世界的问题都是相互联系的,人们每时每刻都在关心事物之间的关系.比如,收入和消费之间的关系、投资和经济增长之间的关系、广告投入和产品销售量之间的关系、治疗手段和治愈率之间的关系等.变量间关系的研究主要包含这样几个问题:变量间是否存在关系?如果存在关系,是函数(确定性)关系还是相关关系?如果是相关关系,是线性关系还是非线性关系?关系的强度如何,方向如何?能不能定量刻画?统计学研究的一个主要任务就是定量地揭示事物之间存在的有别于确定性关系的统计相关关系.

依照取值特征,变量间的关系可分为定性变量间的关系、定量变量间的关系及定性和定量变量间的混合关系.我们可以通过统计图、统计量及统计模型提取变量之间关系的相关信息.本书第8章介绍的就是定量变量间关系的统计研究方法.

统计研究正是依据所构建的一整套科学的统计变量体系,通过对大量的经济、社会和科学技术或自然现象进行系统观察,进而解读、提取上述变量分布、特征及关系等方面的综合信息,采用估计、检验等统计手段,提取有序的、有规律的数量关系和数量界限,进而得到反映所研究

系统及其子系统的变化特性及规律性的数量性信息。

3. 统计数据

统计数据是统计信息的载体,负载研究对象状态、结构、动态特征及发展变化规律等方面的数量信息;**统计数据又是统计变量的具体表现**,是统计变量的实现值。例如,每个企业所拥有的固定资产是一个变量,各企业固定资产的具体数值可能是200万元,也可能是300万元。这些具体数值即为变量的实现值,也就是统计数据,是我们所感兴趣的统计信息。为了区别,除一些惯用表示方法,在本书中,统计变量均用大写的英文字母表示,而统计数据则用小写英文字母表示。

按不同标准,可对统计数据进行不同分类。

(1) 按取值特征可分为定类数据、定序数据及定量数据

统计数据作为统计变量的实现值,根据取值特征,也可以相应地划分为定类数据、定序数据和定量数据。其中,定类数据与定序数据统称为定性数据,数据表现为类别,用文字来表述,为非数字型数据。而定量数据是定量变量的观察值,结果表现为具体的数值,是对事物的精确测度。

从定类数据到定量数据,体现数据测量由低到高的层次。数据层次越高,所蕴含的信息越多,相应的统计分析手段越丰富。

(2) 按收集方法可分为调查数据和试验数据

调查数据是通过调查或观测而收集到的数据,在没有对事物人为控制的条件下而得到的,有关社会经济现象的统计数据几乎都是调查数据。

试验数据是在试验中控制试验对象而收集到的数据。比如,对一种新药疗效的试验,对一种新的农作物品种的试验等。自然科学领域的数据大多数都为试验数据。

(3) 按时间状态可分为截面数据和时间序列数据

截面数据也称为静态数据,是指在相同或近似相同的时间点上收集的数据。描述现象在某一固定时刻的变化情况。比如,2008年我国各地区的国内生产总值数据。

时间序列数据也称为动态数据,是指在不同时间点上收集到的数据。描述现象随时间变化的情况。比如,2000年至2012年我国国内生产总值数据。关于时间序列的统计研究基本方法将在本书第9章及后续课程中讲述。

(4) 按数据的表现形式可分为绝对数形式和相对数形式

现象的规模、水平一般以绝对数形式表现。绝对数的计量单位一般为实物单位或价值单位。实物单位可以是自然计量单位,也可以是物理计量单位,如人口数用人计量,机器数用台计量,对于一些化工产品和燃料,常常还折合成标准实物单位计量。价值单位是以货币形式进行计量,如国内生产总值、进出口总额等就是以价值单位为计量单位。

相对数由两个互相联系的数值对比求得。常用的相对数包括:结构相对数、动态相对数、比较相对数、强度相对数、计划完成相对数等。关于这些相对数的含义及其计算,本书将在第10章作详细介绍。

§ 1.3 统计学研究体系

1.3.1 统计学研究对象

一个统计问题总有它明确的研究对象——统计总体,研究的目的是获取总体信息.由于统计数据是统计信息的载体,所以,统计学的最直接研究对象就是统计数据.从这个角度,统计学可视为是一门研究如何有效地收集、整理和分析受随机因素影响的数据,并对所考虑的问题做出推断或预测,直至为采取决策和行动提供依据和建议的学科,其目的是通过搜集、整理、分析取自某研究总体的统计数据,来认识该统计总体的内在数量规律性.

统计学的研究对象,具有**随机性、总体性和内在规律性**的特点.

统计学研究的是自然和社会不确定性现象内在量的规律性,所研究对象是受随机因素影响的数据,“随机性”是统计学研究对象的一个重要特点,也是统计学采用归纳及经验研究思维的根本原因.

统计研究关注的是对统计总体规律性的把握,后续的学习中我们可以看出统计研究的基本方式是借助捕捉、归纳大量样本数据中存在的普遍特征来反映其所来自的总体的数量特征.例如,对某市中学生体质状况的研究,虽然需要对不同具体的学生进行调查,但其目的并不在于了解每一位中学生的体质状况,而是要反映“某市中学生体质”这一统计总体的特征.因此,“总体性”是统计学研究对象的基本特征.

统计学研究的总体本质上是随机变量,是对随机现象总体普遍存在的事实进行大量观察和综合分析.而随机现象既有其偶然性的一面,又有其必然性的一面,体现为构成总体数据的个体数据之间的同质性和差异性.概率论的学习已明确告诉我们随机现象的统计规律性是其本身所固有的,是不随人们意志而改变的一种客观属性,统计学就是通过已获得的随机试验数据揭示与度量其中所蕴含的“内在规律性”的科学方法.

1.3.2 统计研究过程

针对实际问题,统计研究活动主要包括**统计设计、统计数据的收集与整理、样本信息的提取及总体特征的推断与分析**四个环节.各环节紧密相连,环环相扣,每一步都成为统计学研究的一个有意义的领域.

第一,统计设计.统计设计是整个统计研究工作的基础环节,是对统计工作各个方面进行通盘考虑和安排,其主要内容是明确研究目的、制定变量体系、确定数据收集方案以及研究进度等,最终形成工作方案.统计设计的目的是形成一个有组织、有系统、可分析的研究过程,这一环节的完成好坏直接影响着整个统计研究工作的质量.特别是统计变量体系的建立,这是对现实问题的抽象化与定量化,一方面要立足于所研究实质性学科的理论基础,另一方面也要受统计学基本理论指导,是整个设计环节的重中之重.

第二,统计数据的收集与整理.经过统计设计,形成研究方案之后,就可以开始收集统计数据.如何有效获取统计数据、如何保障统计数据的可应用性及质量,使所需要的统计信息能通过统计数据得到充分传递,是这一环节的主要工作.统计学中有专门的分支——统计调查与试