



普通高等教育“十一五”国家级规划教材
全国统计教材编审委员会推荐使用教材

· 国家精品课程 ·

厦门大学统计学 · 国家双一流学科教材



统计学

STATISTICS

【第六版】

陈珍珍 主编

罗乐勤 副主编

厦门大学出版社 国家一级出版社
全国百佳图书出版单位



普通高等教育“十一五”国家级规划教材

全国统计教材编审委员会推荐使用教材

国家精品课程

厦门大学统计学·国家双一流学科教材

统 计 学 (第六版)

Statistics

陈珍珍 主 编

罗乐勤 副主编



厦门大学出版社
XIAMEN UNIVERSITY PRESS

国家一级出版社
全国百佳图书出版单位

图书在版编目(CIP)数据

统计学/陈珍珍主编. —6 版. —厦门:厦门大学出版社,2018.2

ISBN 978-7-5615-6662-6

I. ①统… II. ①陈… III. ①统计学-教材 IV. ①C8

中国版本图书馆 CIP 数据核字(2017)第 209569 号

出 版 人 郑文礼

责任编辑 陈丽贞

封面设计 蒋卓群

技术编辑 朱 楷

出版发行 厦门大学出版社

社 址 厦门市软件园二期望海路 39 号

邮政编码 361008

总 编 办 0592-2182177 0592-2181406(传真)

营销中心 0592-2184458 0592-2181365

网 址 <http://www.xmupress.com>

邮 箱 xmup@xmupress.com

印 刷 厦门市金凯龙印刷有限公司

开本 720mm×970mm 1/16

印张 24.5

字数 414 千字

印数 1~6 000 册

版次 2018 年 2 月第 6 版

印次 2018 年 2 月第 1 次印刷

定价 45.00 元

本书如有印装质量问题请直接寄承印厂调换



厦门大学出版社
微信二维码



厦门大学出版社
微博二维码

第六版前言

本书自2002年第一版问世以来,在广大同行专家与各高校使用者的关怀与支持下,累计发行量接近14万册。本书的第一版是全国统计教材编审委员会推荐使用教材,于2004年作为福建省精品课程的立项教材。第二版被列为“普通高等教育‘十一五’国家规划教材”,而后一直是国家精品课程的使用教材。本书曾荣获第二届福建省优秀出版物奖,华东地区大学出版社第八届优秀教材、学术专著一等奖,中国大学出版社图书奖首届优秀教材奖一等奖,中国大学出版社图书奖第十届优秀畅销书奖一等奖等奖项。获得这些荣誉只能说明过去,时代在前进,我们也应与时俱进。特别是最近厦门大学统计学科入选“双一流”建设学科名单,我们倍感统计教材的建设、更新任重而道远。

为了适应数据科技时代的发展,我们推出第六版。在大数据、云计算时代,海量的、膨胀式的且又是碎片式的信息蜂拥而来。人们希望能够有效地驾驭各类信息,使之更好地为人类所用。统计作为认识客观世界数量规律的有力工具,为人类提供了驾驭数据信息不可或缺的手段。统计学是一门研究数据的搜集、处理和分析的科学,借助统计学的理论与方法,人们可以更深入地认识现象的数量特征、数量关系及其变动趋势,更有预见地把握未来。

本书的编著者长期潜心于统计教学与科研,通过多年在统计教学第一线的师生互动,从教学实践过程中总结经验,不断探索和调整适合我国高等院校经济类、管理类学生的统计教育模式。根据数据科技时代的特征,我们在第六版的编撰中,坚持统计理论与客观实际应用紧密结合,使教材内容能够更符合我国高等院校经济类、管理类统计教育的要求。在编排形式上,力求版式活泼,做到图文并茂。在教材内容上,我们对原有版本进行了适当的调整、提炼。主要的调整之处:第一,针对数据科技时代的特点,补充适量的关于大数据与云计算的基础理论;第二,厘清统计数据与大数据在性质上的差异;第三,阐明统计在大数据时代的地位与作用;第四,引用最新的有关社会经济数据编制例题并演示相应的统计软件的处理方法,以期更有效地提高学生运用统计数据以及进行数据处理的能力。同时,也删减了一些与数据科技时代不相适应的数据处理方法。

本书第六版的修订,具体分工如下:陈珍珍教授承担第一章导论,第二章统计数据的搜集、处理和显示,第五章参数估计的第三节,第十章统计综合评价;罗乐勤教授承担第三章分布数量特征的统计描述,第八章时间序列分析与预测,第九章统计指数与因素分析;黄长全博士、副教授承担第四章统计推断的理论基础,第五章参数估计的第一、二节,第六章假设检验;皇甫秀颜博士、助理教授承担第七章相关分析与回归分析;袁加军博士、高级工程师承担 Excel 在统计中的应用和附表的修订工作。陈珍珍教授负责全书修订的设计、总纂和定稿工作。

我们衷心地感谢厦门大学出版社领导和责任编辑的鼎力支持,使本书能够顺利再版。各位同仁对本书的修订提出了宝贵的意见和建议,在此我们也表示衷心的感谢!

在修订过程中,我们的目标是尽善尽美。但限于编著者的水平与经验,书中难免有不够完善乃至错误之处,恳望同行专家和读者不吝赐教。

陈珍珍

2018 年 1 月于厦门珍珠湾花园

目 录

第一章 导 论	1
第一节 认识统计	1
一、统计的含义	
二、统计学的研究对象及其特点	
三、统计实践的活动过程	
四、统计数据类型	
五、变量及其分类	
第二节 统计学的基本概念	8
一、统计总体和样本	
二、总体单位与标志	
三、统计描述和统计推断	
第三节 统计在大数据与云计算时代的作用	11
一、大数据(big data, mega data)与云计算(cloud computing)	
二、大数据、“云”与统计的互补作用	
三、大数据时代,统计学是大数据分析的灵魂	
第二章 统计数据的搜集、处理和显示	21
第一节 统计数据的搜集	21
一、统计数据搜集的基本理论与方法	
二、统计调查的组织形式和统计调查体系	
三、统计调查方案的设计	
第二节 数据处理	33
一、数据处理的基本理论	
二、统计分组(分类)	
三、统计指标	
第三节 频数分布	45

一、频数分布及其两个要素	
二、频率及其性质	
三、累计频数与累计频率	
四、频数分布的类型	
第四节 数据显示	50
一、以统计表显示	
二、以统计图显示	
第三章 分布数量特征的统计描述	62
第一节 分布的平均水平、集中趋势和位置的度量	62
一、统计平均数的含义与作用	
二、数值平均数	
三、众数、中位数和其他分位数	
第二节 分布离散程度的度量	80
一、变异指标的含义与作用	
二、极差与四分位差	
三、平均差	
四、方差与标准差	
五、变异系数	
六、异众比率	
七、箱线图在统计描述中的运用	
第三节 分布的偏度和峰度	91
一、统计动差	
二、偏度	
三、峰度	
第四章 统计推断的理论基础	99
第一节 概率与概率分布	99
一、随机事件与概率	
二、随机变量与概率分布	
第二节 随机变量的数值特征与随机向量	111
一、随机变量的数值特征	
二、随机向量与独立性	
第三节 大数定律与正态分布定理	118
一、大数定律	

二、中心极限定理与正态逼近	
第四节 抽样分布	121
一、统计推断与简单随机样本	
二、统计量与抽样分布	
第五章 参数估计	129
第一节 参数估计原理与点估计	129
一、参数估计原理	
二、点估计	
第二节 区间估计	136
一、区间估计的意义	
二、总体均值的区间估计	
三、总体成数的区间估计	
四、正态总体方差的区间估计	
第三节 各种抽样设计下的参数估计	143
一、简单随机抽样形式下的参数估计	
二、类型抽样设计下的参数估计	
三、整群抽样设计下的参数估计	
四、等距抽样设计下的参数估计	
五、二阶段抽样设计下的参数估计	
第六章 假设检验	170
第一节 假设检验基本原理	170
一、基本原理与程序	
二、两种类型的错误与检验功效	
第二节 总体参数检验	175
一、单侧检验与双侧检验	
二、总体均值的假设检验	
三、正态总体方差的假设检验	
第三节 非参数检验	183
一、非参数假设检验概述	
二、总体分布的 χ^2 检验	
三、符号检验法	
四、秩和检验	
五、游程检验	

第七章 相关分析与回归分析	195
第一节 相关分析	195
一、函数关系与相关关系	
二、相关关系的种类	
三、相关图和相关表	
四、相关系数及其计算方法	
五、相关系数的检验	
第二节 简单线性回归分析	201
一、相关分析与回归分析的关系	
二、一元线性回归模型	
三、模型参数的估计	
四、一元线性回归模型的检验	
五、一元线性回归模型预测	
第三节 多元线性相关与回归分析	218
一、标准的多元线性回归模型	
二、多元线性回归模型的参数估计	
三、多元线性回归模型的检验和预测	
四、复相关系数和偏相关系数	
第八章 时间序列分析与预测	230
第一节 时间序列概述	230
一、时间序列的概念	
二、时间序列的种类	
三、时间序列的编制原则	
第二节 时间序列的水平分析	234
一、发展水平与平均发展水平	
二、增长水平与平均增长水平	
三、年距增长量	
第三节 时间序列的速度分析	240
一、发展速度与增长速度	
二、平均发展速度与平均增长速度	
三、年度化增长速度	
四、水平分析与速度分析的结合与应用	
第四节 时间序列的分解分析	246

一、影响时间序列的因素	
二、长期趋势的测定	
三、季节变动的测定	
四、循环变动的测定	
第五节 趋势外推预测	259
一、移动平均法	
二、指数平滑法	
三、加权最小二乘法	
四、预测结果评价	
第九章 统计指数与因素分析	271
第一节 统计指数概述	271
一、统计指数的概念和特点	
二、统计指数的种类	
三、统计指数的作用	
第二节 综合指数的编制与应用	274
一、综合指数的编制原理	
二、综合指数的编制方法	
三、生产指数	
四、其他综合指数的主要应用	
第三节 平均指数的编制与应用	284
一、平均指数的编制原理	
二、平均指数的编制方法	
三、平均指数的主要应用	
第四节 指数体系与因素分析	290
一、指数体系的概念和作用	
二、连锁替代法	
三、平均指标变动的因素分析法	
第十章 统计综合评价	303
第一节 统计综合评价概述	303
一、统计综合评价的概念和特点	
二、统计综合评价的一般程序	
三、综合评价的局限性	
第二节 评价指标选择及其权重的确定方法	307

一、选择评价指标的原则	
二、评价指标的选择方法	
三、权重的含义及其分类	
四、权重的确定方法	
第三节 评价数据的处理与评价结果的综合	319
一、评价指标类型的一致化处理	
二、定性数据的量化处理	
三、定量数据的同度量化处理	
四、评价数据的综合汇总	
第四节 模糊综合评价法	332
一、模糊综合评价的基本理论问题	
二、模糊综合评价的数学模型	
三、模糊综合评价的步骤	
附录一 Excel 在统计中的应用	345
一、Excel 在一般描述统计中的应用	
二、Excel 在推断统计中的应用	
三、统计特有分析方法中 Excel 的应用	
附录二 附表	369
附表 1 二项分布临界值表	
附表 2 正态分布概率表	
附表 3 t 分布临界值表	
附表 4 χ^2 分布临界值表	
附表 5 F 分布临界值表($\alpha=0.05$)	
附表 6 秩和检验表	
附表 7 游程检验 R 临界值表	
主要参考文献	378

第一章 导论

对统计学的一知半解常常造成不必要的上当受骗,对统计学的一概排斥往往造成不必要的愚昧无知。

——[美]C.R 劳

第一节 认识统计

一、统计的含义

在大数据时代,人们面临着海量的、膨胀式的且又是碎片式的信息。人们希望能够有效地驾驭各类信息,使之更好地为人类所用。统计,为人类提供了驾驭信息的有效工具。例如,股民要预测股票价格指数的走势以决定其投资的取向;电商开拓新的市场,要掌握客户的消费特点和需求;药剂师想了解新研制的药物对某种疾病的治愈率及其是否值得推广;城市的管理者要了解不同时段不同区域车辆的流向、流量以科学地解决交通拥挤问题等。要处理这些海量数据,认识现象的特点和规律,就必须借助统计。统计为人们提供了获取相关信息、认识客观现象的工具。统计已渗透到我们的经济管理、科学实验和日常生活等各个领域。

认识统计,可以从三维的角度分析:一是统计学,二是统计实践活动,三是统计数据。统计学就是关于数据的搜集、整理、归纳和分析的方法论科学。统计学研究的是如何应用科学的手段向客观实际中搜集数据资料;对搜集到的数据,如何进行加工和整理,又如何从这些复杂纷繁的数据中,探索其数量特征、数量关系以及变动趋势并得出结论,应用统计语言解释这个结论,以达到对客观现象更正确、更深刻的认识。统计学主要包括:描述统计学和推断统计学等。

统计作为一种实践活动已有悠久的历史,其英文为 statistics,与“国家”、“状况”同一词根,可以说,自从有了国家,就有统计实践活动。据历史记载,我国在西周朝代已建立了统计报告制度,到秦建立中央集权的国家时,从地方到中央形成了比较完善的“上计”报告制度。统计被认为是治国创业图强的重要手段,正如春秋战国时的管子所说的“举事必成,不知计数不可”。

最初,统计只是为统治者管理国家搜集资料、提供数量依据。随着社会经济的发展,统计的应用领域越来越广泛,不仅仅在经济管理领域,在军事、医学、生物、物理、化学、审计和司法等领域中,也大量地运用统计方法。统计实践活动,就是人们为认识客观事物,通过实验或调查搜集有关数据,并加以整理、归纳和分析,而后对客观事物规律性的数量表现做出统计上的解释。统计实践活动的过程,一般经过三大环节:数据搜集、数据处理和数据评价与分析。实质上,统计实践活动过程也是人们认识客观世界的过程。

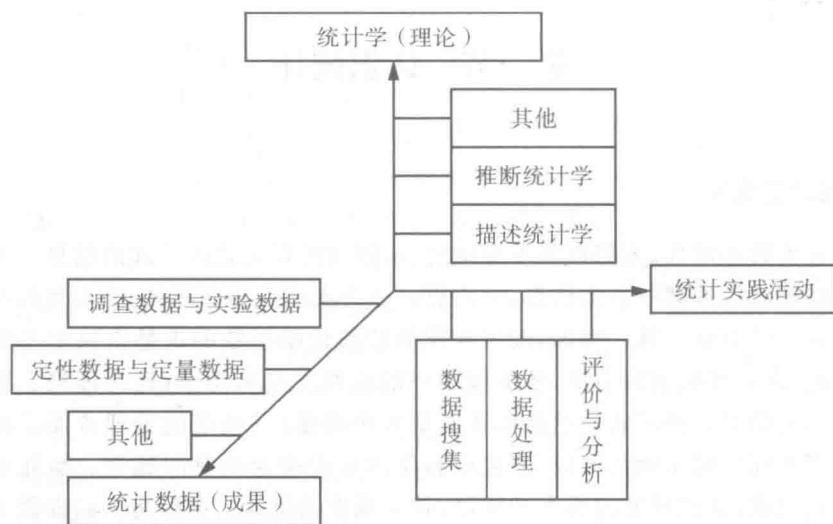


图 1-1 三维角度认识统计

人们通过统计实践活动所得的成果即统计数据,如果按观察现象的属性划分,可以分为调查数据和实验数据;如果按数据的计量尺度划分,可以分为定性数据和定量数据。当然,如果采用不同的划分标准,还有其他的类别。例如,2017年5月对比上年同月我国居民消费价格指数为101.5%,2017年1—5月全国房地产开发投资37 595亿元等,这些都是统计数据。

统计学、统计实践活动、统计数据三者之间的关系:统计学与统计实践活

动的关系是理论与实践的关系,理论源于实践,理论又高于实践,反过来又指导实践。统计实践活动与统计数据的关系是工作过程与工作成果的关系。工作过程的好坏关系到工作成果质量的高低。人们对统计数据的要求是:客观性,即它能反映客观事实而不受任何偏见的影响或任何势力的干扰;准确性,即统计数据的偏差不能超过根据统计研究目的而事先确定的允许误差范围;及时性,即统计数据应及时搜集、及时加工、及时公布。只有在科学理论的指导下,统计工作才能获得更高的效率,从而提高其工作成果——统计数据的质量。

二、统计学的研究对象及其特点

统计学的研究对象亦即根据统计研究的目的所要认识的客体,具体说来,是指客观事物的数量特征和数量关系。统计学的研究对象具有如下特点:

(一)可度量

可度量是统计学研究对象的基本特点。统计数据是客观事物量的反映,通过数据以测度事物的类型、量的顺序、量的大小和量的关系。

(二)总体性

统计研究虽然是从个别入手,对个别单位的具体事实进行观察研究,但其目的是为了达到认识总体数量特征。例如,进行居民家计调查,虽然是对具体的每个调查户入手进行观察,但其目的并非研究个别居民户的家庭基本状况,而是要反映一个部门、一个地区以至一个国家居民的收入、消费、就业和人口等状况。

(三)变异性

统计研究的是同质总体的数量特征,其前提是总体各单位特征表现存在着差异,而这些差异不是由某些特定的原因事先给定的。统计上把总体各单位由于随机因素引起的某一标志表现的差异称为**变异**。例如,一个城市的居民,其收入水平、消费结构、从业情况存在着差异,这才有必要研究其收入、消费和从业等现象的总体水平、结构特点、差异程度及其变动趋势。如果各单位不存在差异,也无须进行统计了。

三、统计实践的活动过程

统计实践的活动过程,一般经过三个环节,如图 1-2 所示。

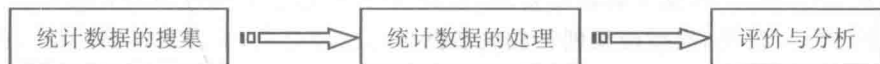


图 1-2 统计实践三个环节

(一) 统计数据的搜集

数据的搜集是通过统计调查而获得,该环节是统计实践的基础环节。有关数据搜集的具体理论与方法,我们将在第二章详细阐明。

(二) 统计数据的处理

调查所获得的数据需要处理,现代的处理手段一般是利用统计软件进行处理。数据处理的具体内容和方法,我们将在第二章介绍。

有关利用统计软件处理统计数据的方法,我们将在本书的附录一“Excel在统计中的应用”里具体加以说明。

(三) 评价与分析

用统计语言解释数据的含义,分析现象的数量特征、数量关系和变动趋势,评价调查数据的质量。

本书的第三章“分布数量特征的统计描述”、第七章“相关分析与回归分析”、第八章“时间序列分析与预测”、第九章“统计指数与因素分析”和第十章“统计综合评价”,简要地介绍了一些常用的统计分析理论与方法。

四、统计数据的类型

(一) 按观察现象的属性分类

客观现象按其属性可以分为社会经济现象和自然科学现象。统计数据按其所观察现象的属性可以分为调查数据和实验数据。

1. 调查数据

人们为获得社会经济现象的相关数据,运用相应的调查方法,向客观实际搜集数字资料的过程就是统计调查。通过统计调查搜集的数据就是调查数据。

2. 实验数据

在自然科学研究领域,人们往往通过实验获得相关数据。在实验中控制一个或多个变量,在一定的控制条件下观测实验对象,从实验过程中获得数据。这种在实验中控制实验对象、观测整个实验过程而搜集到的数据就是实验数据。例如,某种新的疫苗问世,如何测试它对某种疾病的防疫效果呢?实验者常常是进行随机对照实验,将实验对象进行分组,分为一个实验组,一个或多个对照组。根据实验设计方案,实验者观测、记录不同组别的实验对象对疫苗的反应情况,根据搜集到的相关数据进行比较分析,从而证实新的疫苗的防疫效果。

(二)按数据的计量尺度分类

统计数据是对客观现象进行计量的结果,根据对研究对象计量的不同精确程度,将计量尺度由低到高、由粗略到精确分为两大层次:定性数据和定量数据。

1.定性数据

定性数据常用文字表述,其计量结果表现为类别。定性数据又可以分为定类数据和定序数据。

定类数据亦称为列名数据,它是按照客观现象的某种属性对其进行平行的分类,此时,若用数字表示,该数字仅作为各类的代码,度量各类之间的类别差,不反映各类的优劣、量的大小或顺序。例如,人口按性别分为男女,用“1”表示男性,用“0”表示女性。定类尺度的主要数学特征可以用“=”或“ $\neq$ ”表示。在统计处理中,虽然可以计算单位数,但它不能表明第一类的一个单位可以相当于第二类的几个单位。

定序数据亦称为顺序数据,它是对客观现象各类之间的等级差或顺序差进行测度的数据,比定类数据的计量尺度更高一级。定序数据不仅可以将研究对象分成不同的类别,而且还可以反映各类的优劣、量的大小或顺序。例如,学生成绩可以分为优、良、中、及格和不及格等五类,在这里,定序数据虽然无法表明一个优等于几个良,但却能确切地表明优高于良,良又高于中……定序尺度的主要数学特征可以用“ $<$ ”或“ $>$ ”表示。在统计的变量数列中可以确定其中位数、分位数等。

2.定量数据

定量数据比定序数据的计量尺度又更高一级。定量数据是对客观现象进行计量的结果,表现为具体的数值。它不但可以用数表示客观现象各类别的不同和顺序大小的差异,而且可以用确切的数值反映现象之间在量方面的差异。对于定量数据,一般可以运用数学的运算方法处理。在统计实践中,我们所处理的大多数是定量数据,定量数据在统计中占据重要的地位。例如,各类总量指标、相对指标和平均指标都是定量数据。

(三)按对客观现象的时空状态分类

在统计研究中,根据对客观现象观察的时间、空间状态不同,统计数据可分为横截面数据、时间序列数据和面板数据。

1.横截面数据

横截面数据又称为静态数据,也有人称之为空间数据,它是在同一时间对不同横截面个体的观测值的集合。例如,2017年7月我国各省(直辖市)的消

费价格指数就是横截面数据。

2. 时间序列数据

时间序列数据又称为动态数据,它是在不同时间对同一个体(或同一现象)的观测值的集合。例如,2007年7月—2017年7月福建省的消费价格指数就是时间序列数据。关于时间序列,将在本书第七章中进行详细介绍。

3. 面板数据

横截面数据和时间序列数据的组合形成面板数据,它是指不同的横截面个体在不同时间的观测值的集合。从横向看,它包括某一具体时间上不同横截面的数据;从纵向看,它包括每一个横截面个体的时间序列数据。例如,2007年7月—2017年7月我国各省(直辖市)的消费价格指数是面板数据。

(四) 按数据的表现形式分类

统计数据通常表现为绝对数、相对数和平均数。

1. 绝对数

现象的规模、水平一般以绝对数形式表现,例如,国内生产总值、产品总产量、人口数、进出口额和股票的交易额等。绝对数的计量单位一般为实物单位或价值单位,有时也采用复合单位。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,粮食产量用吨计量,耕地面积用公顷计量等,对于一些化工产品,常常折合为标准实物单位。价值单位是以货币形式进行计量,如国内生产总值、进出口总额是以价值单位为计量单位。复合计量单位是由两种或两种以上计量单位复合而成的,如以“吨公里”为货物周转量的计量单位,以“千瓦时”为用电量的计量单位。

绝对数按其反映的时间状态不同,分为时期数据和时点数据。时期数据是反映现象在一段期间内发展过程的总量,它具有连续统计和可加性的特点,其数值大小与所属的时间长短有直接关系,如国内生产总值、进出口总额等。时点数据是反映现象在某一特定时点所处的状态,它是采用间断登记方式取得资料的,不具有可加性,其数值大小与时点间隔长短没直接关系,如期末人口数、期末在建工程投资额等。

绝对数按其反映对象的内容不同,分为总体单位数和总体标志总量,本书将在第三章对这方面的内容作详细的介绍。

2. 相对数

相对数是由两个存在着相互联系的绝对数对比而得的,是由绝对数派生的。常用的相对数有:结构相对数、动态相对数、比较相对数、比例相对数、强度相对数、利用程度相对数和计划完成相对数等。相对数的计量单位大部分