



高等院校财经类专业应用型本科系列教材

统 计 学

◎主编 黄 英 刘亚琼

TONGJIXUE



重庆大学出版社
http://www.cqup.com.cn

高等院校财经类专业应用型本科系列教材

统 计 学

TONGJIXUE

◎主 编 黄 英 刘亚琼

◎副主编 胡晓峰 俞良蒂

重庆大学出版社

内容提要

本书是一部关于统计数据的收集、整理、分析和解释的统计学教材。统计学是一门方法论科学,其目的是探索数据的内在数量规律性。本书在体系上涵盖了描述统计和推断统计两大分支,在内容上主要体现:收集统计数据是进行统计分析的基础;统计数据的整理是数据收集与数据分析之间的必要环节;统计数据的分析是统计学的核心内容,是通过统计描述和统计推断,探索数据内在规律的过程,提供重要的方法与手段;数据解释是对分析的结果进行说明。四者构成《统计学》教材的主体内容,形成了统计学严密的方法论体系。本书在内容上强调理论与实践相结合,重统计理论思想传播,注重运用统计学方法采用 Excel 作为分析工具来解决实际问题。

本书既适宜作为高等院校经济管理类专业统计学课程的教材,也可提供给社会经济统计工作者以及学习 Excel 进行统计分析的读者参阅。

图书在版编目(CIP)数据

统计学/黄英,刘亚琼主编. —重庆:重庆大学出版社,
2017.1

高等院校财经类专业应用型本科系列教材
ISBN 978-7-5689-0270-0

I. ①统… II. ①黄…②刘… 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(2016)第 298455 号

高等院校财经类专业应用型本科系列教材

统计学

主 编 黄 英 刘亚琼

副主编 胡晓峰 俞良蒂

策划编辑:范 莹

责任编辑:李定群 版式设计:范 莹

责任校对:邬小梅 责任印制:赵 晟

*

重庆大学出版社出版发行

出版人:易树平

社址:重庆市沙坪坝区大学城西路 21 号

邮编:401331

电话:(023) 88617190 88617185(中小学)

传真:(023) 88617186 88617166

网址:<http://www.cqup.com.cn>

邮箱:fxk@cqup.com.cn (营销中心)

全国新华书店经销

重庆升光电力印务有限公司印刷

*

开本:787mm×1092mm 1/16 印张:18.75 字数:422千

2017 年 2 月第 1 版 2017 年 2 月第 1 次印刷

印数:1—3 000

ISBN 978-7-5689-0270-0 定价:39.00 元

本书如有印刷、装订等质量问题,本社负责调换

版权所有,请勿擅自翻印和用本书

制作各类出版物及配套用书,违者必究

序

随着时代的发展及各类高等院校办学规模的不断扩大,人才培养质量成为衡量各院校办学水平的主要指标,而教材作为实现人才培养目标的载体,对各类高等院校的发展和人才培养质量具有举足轻重的作用。

就许多院校目前使用的教材现状来看,现行教材不太适合院校学生特点、不太符合办学定位的问题较为普遍,许多高职高专院校使用一本、二本的教材,偏理论化、内容死板;而本科院校使用高职高专的教材,过分重视应用和操作,缺少相关学科的广度和深度。如何解决这个问题?必须根据各校办学层次、专业特色和培养目标,进一步加强自编教材的建设工作,做到因材施教、量体裁衣。对于民办院校来讲,应根据这类院校应用型专业人才的培养目标,使教材建设工作做到“面向应用、面向职业、面向能力、面向实践”,重视教材的适用性、实用性、思想性和生动性。教师编写和课堂选用的教材,应符合学生层次,既能调动学生的学习热情,又能切实提高学生动手能力、分析能力和可持续发展能力,满足独立学院培养应用性专业人才的定位要求。武汉工程科技学院经济与管理学院的中青年教师在老教师的指导与带动下,基于多年的教学工作,不断学习,亲历实践,编写了一套适用于应用型本科经济与管理类专业学生学习的专业基础课教材。

本书正是经管类应用型专业人才培养目标要求的体现。通览本书,它无疑是一部内容丰富、生动活泼和具有亲和力的专业基础教材,本书适用于应用型本科经管专业的学生,对于对统计学有兴趣但非统计专业的学生,它也不失为一部适用的优秀教材。本书对统计思想和具体应用着墨甚多,在“统计公式”“数字推导”等统计方法上面,作者努力淡化其复杂和困难程度;本书重视“统计思想”的传输,重视“如何做统计思考”胜于鼓励学生搬硬记统计方法或公式的推导。作者将统计学这门看似属于数理类的学科,努力尝试以文科的方式来表达,这种努力体现在本书中的案例选用、概念解释、统计结论表述以及习题的设计等各个细节,它尽力在改变人们对统计学的传统看法,旨在使初学者产生学习统计学的兴趣,并能从中受益,是本书最大的亮点之一。

本书作者们既有长期从事统计学教学的老师,还有从事统计方法应用的实际工作者,皆有丰富的统计学教学和实践经验,并在各自研究领域有独到见地。他们长期从事在统计学教学的第一线,孜孜不倦地探求以统计学理论和方法解决现实问题的途径,作者们均拥有多年统计学教学经验和企业相关工作经历,这使得本书的统计方法、思想在联系实际方面更具有优势。无论是引文或案例,还是课后习题,均取材丰富并贴合现实,政府政策、社会服务、餐饮、金融、地产、汽车、信息通信和教育等均有所涉及;统计结论及实践适用于社会生活和

经济管理的各个方面,如政策制定、产品宣传、研发、生产经营、财务、人力资源管理、战略策划等;这使得本书内容翔实丰富,并极具实践性。

每章节之前的“学习目标”,最简要地提示本章重点内容;接下来的“开篇案例”给出一个社会、经济或商务实际案例,以引入本章内容。在每章结束之时,安排“案例讨论”与全文相呼应,诠释如何将本章所讲统计方法应用于这个实际案例;“学习指导与小结”按教学安排顺序列出各节要点;“常用术语”列出本章的全部名词术语;“思考与练习”分成不同难度等级,以便适用于不同教学对象。

计算机技术日新月异,它推进了统计学学科的发展,同时为统计方法的应用开辟广阔天地。本书每章根据内容,专门介绍计算机技术与本章内容的紧密关系,并编写相应案例进行详细说明。例如,计算机在数据收集中的应用,Excel 在统计数据输出,以及统计图、结果演示、描述统计和推断统计方法的运用等,这些内容使读者不仅可从中学习到统计学方法和理论,并且可实际上机操作,充分提高学生运用统计方法解决问题的实际动手能力。这是本书的又一大亮点。

综上所述,该部《统计学》定位合理,在作者能力范围之内和读者可接受的限度之下,较为明确地表达出作者对统计学的观点。它基于理论,立足于现实和运用,为广大读者提供了一个良好的选择读本。我们希望各位读者通过阅读和学习本书,对统计学思想有全新的认识,掌握更多的统计学理论和方法,并运用于实际的工作和生活,体会到统计学带给我们的好处和乐趣。感谢参与编写的各位教师为本书所做的努力,希望他们能在统计学教学和研究领域再接再厉,为我国统计科学的教育事业作出新的贡献。

最后,我们期望广大读者在使用该套系列教材的过程中,给我们提出宝贵的意见!

武汉工程科技学院

经济与管理类专业基础课教材编写委员会

2017 年 1 月

前言

我们在长期从事统计学的教学工作中,亲身体验并见证了统计学科的发展与成熟。正如马克思早在 100 多年前就指出的那样:“一种科学只有当它达到能够成功地运用数学时,才算真正发展了。”统计学就是这样一门成功地运用数学,特别是概率和抽样技术等数学理论而形成的方法论学科。它的发展和成熟的方法技术现已广泛地运用到各个领域,无论是自然科学还是社会科学,无论是宏观调控、微观管理还是科学研究乃至人们的日常生活等,均离不开统计方法的应用,统计学科在其应用中不断地凸现出它的魅力。统计学科的产生与发展为信息社会的数据处理提供了大量的、科学的技术与手段,特别是统计软件的开发,使人们从烦琐的计算中解放出来,从难以理解的数学方法中得到了实际应用,充分体现了该学科的使用价值。随着时代的变更,信息社会的到来,计算机技术的引入,统计软件的开发等为统计学科更好发展与成熟的同时,又对统计学科的教学工件提出了新的更高的要求。通过学习该门方法论学科,使学生增加基本技能,在今后的职业生涯中受益,所以要求统计学科的教学工作无论是教学思想、教学方式方法还是教学手段上都需要不断地更新,以顺应本学科的发展需要。

教学实践告之,成熟的学科需要成熟的教学方式与手段,随着计算机技术的引入,统计软件的开发与应用,使传统的教学方式与方法面临着挑战,面对当今世界进入信息社会后,很难想象,脱离计算机的统计学教学效果会如何。特别是到了 21 世纪初,我国高等教育为满足社会人才培养的需求,民办高校正蓬勃发展,“统计学”作为经济与管理类专业必修的核心课程,其重要性显而易见,对于应用型本科学生而言,学习统计学的终极目的就是应用,而应用的关键是对统计学思想的认识和理解深化,这样一来,在课堂上过多地进行统计理论公式的推导显然意义不大,侧重于实际应用是应用型本科教学的特点,如何适应这一受众教育群体的需要,编写出理论与实践相结合,重统计理论思想传播、重实际应用的教材,以适应统计教学的需要,这是首当其冲要解决的问题,也是我们编写该书的基本定位。

由此可见,统计学教材只有突出计算机的应用,以及选用一些有趣的实用的案例(事例)进行分析,渗透统计思想与方法的实用性,才能得以实现统计思想的传播,而统计思想的灌输是统计学的精髓。为此,我们倾心编写此书的宗旨,具体体现在 3 个方面:第一,本书的编写应适应独立学院学生特点,所以内容上不能太理论,不能太枯燥,编写教材的目标应是学生学了能用得上,能立即上手;第二,学生读了此书对统计学能产生兴趣,不再畏难,愿意学并且有学好它的信心;第三,在文字表述上尽量生动和亲切,力求通俗易懂。让学生、教师、统计工作者和一般读者均能从中受益。

从编写思想的确立到教学内容的挑选,从编写手段到表述方式的选择等,我们始终遵循此宗旨,力求此教材具有以下特点:

第一,理论性。注重挑选《统计学》最基本的统计理论,充分表现出统计学方法论思想。围绕着统计数据的收集、整理、表述与分析主线条,以“描述统计学”与“推断统计学”两大学科支柱为依托,精选内容,体现统计学科的基本理论方法论体系。

第二,应用性。结合每一章节内容,选择合适的生活中的案例或事例,以 Excel 为计算工具,展示统计方法在实际生活中的应用,让读者在学习统计思想的同时,了解其在实际生活中的应用意义。结合教学内容详细介绍 Excel 的操作步骤,以及输出结果的识别及其统计分析与结论,反过来加固对统计思想方法的认识,重在应用。

第三,趣味性。在应用性的基础上,结合教学内容选择适当的练习、思考与讨论,增强学习者的主动思维和动手能力。

第四,通俗性。在版面设计上,力求生动活泼;在文字表述上,力求简捷、明了,通俗易懂;在写作上,力求淡化统计理论的推导和计算。重点突出统计思想与方法 and 计算机的应用技术,只有真正理解了统计思想,才能在实际中应用。

本书参编者及分工如下:黄英编写第 5,7,9 章,刘亚琼编写第 2,3,4 章,俞良蒂编写第 1,6,8 章,胡晓峰编写第 10 章。

本书的出版得到各方面的支持与帮助,十分感谢武汉工程科技学院经济与管理学院领导的鼓励与帮助,十分感谢重庆大学出版社的大力支持,感谢武汉工程科技学院曾担任统计教学工作的各位老师提出的宝贵意见和建议,才使本书得以顺利面市。

尽管我们努力和尝试编写此书,但由于统计学科的迅速发展,编者水平所限,编写中的不当和疏漏之处在所难免,敬请同行和读者不吝赐教、批评指正。

编者

2017 年 1 月

目录

第1章 导论	1
1.1 统计及其应用领域	3
1.2 统计数据类型	5
1.3 统计学中常用基本概念	7
1.4 Excel 在统计中的应用常识	10
第2章 数据收集	17
2.1 统计数据的来源	19
2.2 调查方式与技术	21
2.3 数据收集方法	25
2.4 调查设计	28
2.5 数据质量	34
2.6 计算机在数据收集中的应用	35
第3章 统计数据的整理与显示	42
3.1 统计数据的预处理	44
3.2 品质数据的整理与图示	49
3.3 数值型数据的整理与图示	55
3.4 统计表设计	68
3.5 Excel 在数据整理和图示中的应用	69
第4章 数据分布特征的测度	81
4.1 集中趋势的测度	82
4.2 离散程度的测度	89
4.3 偏态与峰态的测度	98
4.4 Excel 在数据测度中的应用	101
第5章 抽样分布与参数估计	108
5.1 一个总体参数估计时样本统计量的抽样分布	109
5.2 参数估计的一般问题	118
5.3 一个总体参数的区间估计	122

5.4 样本容量的确定	128
第6章 假设检验	135
6.1 假设检验的基本原理	137
6.2 一个总体参数的假设检验	144
第7章 方差分析	162
7.1 方差分析基本原理	163
7.2 单因素方差分析	168
7.3 双因素方差分析	176
第8章 相关与回归分析	187
8.1 变量间关系的度量	189
8.2 一元线性回归	198
8.3 利用回归方程进行估计和预测	211
第9章 时间序列分析与预测	223
9.1 时间序列及其分解	225
9.2 时间序列的描述性分析	227
9.3 平稳序列的分析和预测	235
9.4 有趋势序列的分析和预测	242
9.5 复合型序列的分解	247
第10章 统计指数	259
10.1 统计指数的概念	260
10.2 常用加权指数	262
10.3 总量指数与指数体系	266
10.4 几种常用的价格指数	269
10.5 多指标综合评价指数	272
附录 用 Excel 生成概率分布表	279
附表1 标准正态分布表	279
附表2 标准正态分布分位数表	281
附表3 t 分布临界值表	284
附表4 χ^2 分布临界值表	286
附表5 F 分布临界值表	289
参考文献	292

第1章 导论

【名言采撷】

学者不能离开统计而研究,政治家不能离开统计而施政,事业家不能离开统计而执业。^①

——马寅初

【学习目标】

本章的学习目标是通过统计的基本理论及常用的统计概念的学习,让读者有一个先入为主的概貌,知道什么是统计学?什么是统计?统计的历史功绩、统计学家等文化背景;对统计数据有一个完整的认识,对该学科常用的基本概念有一个清楚的认知,并在理解的基础上,抓住关键,举一反三,融会贯通。

【知识点浏览】

- 1.统计和统计学的概念。
- 2.统计数据的类型。
- 3.总体和样本。
- 4.参数和统计量。
- 5.变量。

【开篇案例】

第29届夏季北京奥运会奖牌榜^②

2001年7月13日,在莫斯科举行的国际奥委会第112次全会上,国际奥委会投票选定

① 李心愉.应用经济统计学[M].北京:北京大学出版社,2000.

② 根据第29届夏季北京奥运会网站公布的数据整理。

北京获得2008年奥运会主办权。2008年北京奥运会即第29届夏季奥林匹克运动会于2008年8月8日20时开幕,2008年8月24日闭幕。本届奥运会口号为“同一个世界,同一个梦想(One World, One Dream)”,主办城市是中国北京。参赛国家及地区204个,参赛运动员11438人,设302项(28种运动)比赛项目。



2008年8月8日是值得中国人永远铭刻在心中的日子。这天在我国首都北京,成功地主办了第29届奥运会。提起第29届夏季北京奥运会,人们仍能清晰地记得中国人在世人面前充分展示出的我国以“科技奥运、绿色奥运、人文奥运”开幕式的成果。在开幕式上以“有朋自远方来,不亦乐乎”的人文情怀,展示了高科技与现代艺术的完美结合;奥运会期间的环境保护工作、社会文明和精神文明达到了空前水准,得到了全世界人民的喝彩,那激动人心的场景,相信每个中国人都引以骄傲与自豪,我们更不会忘记我国运动员在这届奥运会上取得的骄人的成绩。获得奖牌居前10位的国家及奖牌数见表1.1。

表 1.1

排 名	国家/地区	金 牌	银 牌	铜 牌	总 数
1	中国	51	21	28	100
2	美国	36	38	36	110
3	俄罗斯	23	21	28	72
4	英国	19	13	15	47
5	德国	16	10	15	41
6	澳大利亚	14	15	17	46
7	韩国	13	10	8	31
8	日本	9	6	10	25
9	意大利	8	10	10	28
10	法国	7	16	17	40
合计		302	303	353	958
中国占比/%		16.9	6.9	7.9	10.4

我国以金牌总数居首,名列第一,占全部金牌总数的16.9%,比名列第二的美国所获金牌多15块;所获得的奖牌总数为100块,奖牌总数占全部奖牌总数的10.4%。第29届夏季北京奥运会在我国的成功主办,不仅展示了中国人的风采,体现了中国人的实力,也使一批年轻的选手脱颖而出。

类似这种用统计数据对某个事件的情况和结果进行报道,用统计方法对宏观和微观经济现象进行描述、分析、推算和预测、决策的事例,在我们的实际生活中比比皆是,这些都是我们最常见的统计形式与结果。

【思考与讨论】

1.列举生活中的统计方法或思想应用的实例。谈谈你对统计学的直观感受。

2.为什么说学习统计学对每个人都是必要的?你有何体会?

类似第29届奥运会的统计数据的公布报道与分析,在日常生活中比比皆是,那么,什么是统计?什么是统计学?统计学有哪些科学方法和技术?统计有哪些思想?用怎样的现代化计算工具完成数据的处理?诸如此类问题,我们将在本书中一一为读者释然。本章就统计学中的基本理论与概念、统计的应用领域和统计数据分类等问题进行介绍。

1.1 统计及其应用领域

提起“统计”一词大家并不陌生,人们经常在报纸杂志、电视广播里看到或听到“统计”字眼,在日常生活中经常用到统计的方法,如用统计图表描述武汉市历年天气气象的变化趋势;用统计指标“年递增速度”表述国民经济的发展状况;用统计预测的方法估计“能源的生产量与消费量”情况;用经济模型找出事物发展变化的规律;用假设检验的思想检验生产厂家对某产品质量的承诺的可靠性,等等。这些都是统计的思想体现,统计方法应用的结果。

1.1.1 统计学的概念

统计是泛指统计资料、统计工作的统计科学的总称。而统计学是对统计数据的收集、整理、加工、汇总、展示全工作过程中,起指导意义的一门方法论学科,是理论与方法论的集合。统计通常指统计实际工作,而统计学是一门方法论学科。如不列颠百科全书的定义:

◎定义 1.1:统计学(statistics)是收集、分析、表述和解释数据的科学。

statistics: the science of collecting, analyzing, presenting, and interpreting data.

数据收集也就是取得数据,数据整理就是对取得的数据进行加工处理,用统计图表等形式展示出来,数据分析则是通过统计方法研究数据及其特征,数据的解释是对分析的结果进行说明。从这一定义中我们可以清楚地知道,统计学是一门关于数据处理技巧的方法论学科。

1.1.2 统计学分科

统计学是一门方法论学科,按应用的方法不同可分为描述统计方法和推断统计方法,因而形成了统计学的两大分科,即描述统计学和推断统计学。

◎定义 1.2:描述统计学(descriptive statistics)是研究数据收集、整理和描述的统计学分支。

描述统计的内容包括取得研究需要的数据,以图表形式进行加工处理与显示、概括与分析,得出所研究现象的一般性规律或特征。如某班的上课出勤率为 97%,就是对学生到课情况的一般性的描述。

◎定义 1.3:推断统计学(inferential statistics)是研究如何利用样本数据来推断总体数量特征的统计学分支。

统计学研究的是总体现象的规律,而不是个体的数量特征,当被研究的总体很大或是无限总体时,无法收集到所有的数据,只能采集样本信息,通过样本数据来推断总体特征。如要研究全国独立学院的办学条件,由于范围太大,数据难以收集,这时就可采用统计抽样技术,抽取适量样本,通过样本提供的信息来推断全国独立学院的教学投入情况。这种由部分推断总体的方法就是应用了科学的抽样技术来实现的。

描述统计学和推断统计学形成了统计学科完整的方法论体系,两者既有联系又有区别。联系表现在:对数据处理的起点和终点是一致的,它们的起点都是收集反映客观现象的数据,终点都是研究总体内在的数量规律性,区别在于:当收集的数据来源于总体,直接运用描述统计学的方法寻找总体内在的数量规律性,若收集的数据来源于样本,则通过描述统计学的方法对样本信息进行描述的基础上,应用概率论(含分布理论、大数定律和中心极限定理)对总体的数量特征进行估计和检验等方法,从而得出总体内在的数量规律性。

1.1.3 统计的应用领域

随着统计学的发展,统计方法被广泛地应用到众多领域,如经济学、管理学、会计学、医学、地质学、社会学、教育学、工程学、气象学等;所有的行业如工业、农业、商业、信息业、金融业、军事、教育等都十分重视统计思想与方法的应用。表 1.2 列出了统计的一些应用领域,目的是让我们通过简单的浏览,充分认识到统计学的应用十分广泛。

当然,统计不是万能的,统计能帮助我们进行数据分析,并从分析中得出某些结论,但对统计结论的进一步解释,则需要专业知识。如在医学上,我们可用统计学中的大量观察法,通过临床观察找到病因及其规律,但是如何加以控制与治疗,却要由医务工作者根据医学专业知识来探索与解决。

统计学科的发展成就了许多统计学家,其中最具有影响力的著名统计学家有 Jacob Bernoulli (伯努利)(1654—1705)、Edmond Halley (哈雷)(1656—1742)、De Moivre (棣莫弗)(1667—1754)、Thomas Bayes (贝叶斯)(1702—1761)、Leonhard Euler (欧拉)(1707—1783)、Pierre Simon Laplace (拉普拉斯)(1749—1827)、Adrien Marie Legendre (勒让德)(1752—1833)、Thomas Robert Malthus (马尔萨斯)(1766—1834)、Friedrich Gauss (高斯)(1777—1855)、Johann Gregor Mendel (孟德尔)(1822—1884)、Karl Pearson (皮尔

逊) (1857—1936)、Ronald Aylmer Fisher (费希尔) (1890—1962)、Jerzy Neyman (奈曼) (1894—1981)、Egon Sharpe Pearson (皮尔逊) (1895—1980)、William Feller (费勒) (1906—1970)等。

表 1.2 统计的应用领域

actuarial work (精算)	industry (工业)
archaeology (考古学)	literature (文学)
agriculture (农业)	management science (管理科学)
auditing (审计学)	marketing (市场营销学)
anthropology (人类学)	medical diagnosis (医学诊断)
demography (人口统计学)	military science (军事科学)
dentistry (牙医学)	meteorology (气象学)
econometrics (经济计量学)	political science (政治学)
engineering (工程)	psychology (心理学)
ecology (生态学)	psychophysics (心理物理学)
education (教育学)	quality control (质量控制)
epidemiology (流行病学)	religious studies (宗教研究)
finance (金融)	sociology (社会学)
genetics (遗传学)	survey sampling (调查抽样)
geology (地质学)	taxonomy (分类学)

1.2 统计数据类型

统计数据是对客观现象进行测量的结果。例如,“2009 年全年国内生产总值 335 353 亿元,比上年增长 8.7%。年末全国总人口为 133 474 万人,全年出生人口 1 615 万人,出生率为 12.13‰;死亡率为 7.08‰;自然增长率为 5.05‰。出生人口性别比为 119.45。”^①这些统计数据就是对客观现象进行测量的结果。

统计数据在测量时,按使用的测量尺度不同,可分为分类数据、顺序数据和数值型数据。按数据收集的方法不同,可分为观测数据和实验数据。按被描述的对象与时间的关系,可分为截面数据和时间序列数据。

^① 《中华人民共和国 2009 年国民经济和社会发展统计公报》中华人民共和国国家统计局,2010 年 2 月 25 日。

1.2.1 分类数据、顺序数据和数值型数据

分类数据是对事物进行分类的结果,数据表现为类别,是用文字来表述的。它是由分类尺度计量形成的。例如,人口按照“性别”分为男、女两类;企业按经济性质分为国有、集体、私营、合资、独资等,这些均是按“类别尺度”进行测量的。为便于统计处理,对于分类数据通常可用数字代码表示。如用“1”表示“女性”,用“0”表示“男性”;用“1”表示“国有企业”,用“2”表示“集体企业”,用“3”表示“私营企业”等。

◎定义 1.4:只能归于某一类别的非数值型数据,称为**分类数据**(categorical data)。

顺序数据同样也是对事物进行分类的结果,但这些类别是有顺序的,它是由顺序尺度计量形成的。例如,按受教育程度由低到高,可分为小学、初中、高中、大学及以上;按对某事物是否赞成的态度,可分为非常同意、同意、保持中立、不同意、非常不同意等。同样,对顺序数据的测量也可用数字代码表示,如用“1”表示“非常同意”,用“2”表示“同意”,用“3”表示“保持中立”,用“4”表示“不同意”,用“5”表示“非常不同意”。

◎定义 1.5:只能归于某一有序类别的非数值型数据,称为**顺序数据**(rank data)。

数值型数据是使用自然或度量衡单位对事物进行测量的结果,其结果表现为具体的数值。例如,国民生产总值、人口数、产品产量等,表明事物具体的数量特征;又如,月工资为 1 880 元、体重 55 kg 等。

◎定义 1.6:按数量特征尺度测量的具体观察值,称为**数值型数据**(metric data)。

分类数据和顺序数据说明事物的品质特征,通常是用文字来表示的,其结果均属于事物的类别,故也可称为品质数据或定性数据。数值型数据说明现象的数量特征,通常是用具体的数值来表现的,故也可称为数量数据或定量数据。

小贴士

分类数据、顺序数据和数值型数据,是根据数据测量的层次进行分类的,此分类将数据由低级到高级排列,分类数据是低层次的,数值型数据是高层次的。高层次的数据满足低层次的计算要求,而低层次的数据不能进行高层次数据的计算。

由于分类数据和顺序数据均是表明事物质的属性,而数值型数据表明事物的量的特征,因此,从描述事物的特征来看,可将分类数据和顺序数据统称为品质数据,而数值型数据可称为数量数据,如图 1.1 所示。

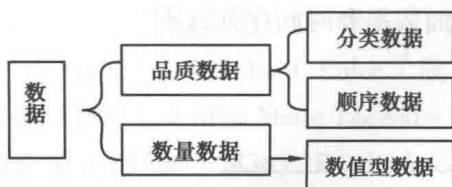


图 1.1 数据按计量层次与特征分类

1.2.2 观测数据和实验数据

观测数据是在对事物没有人为控制条件下而得到的,有关社会经济现象的统计数据几乎都是观测数据。如对人的社会公德水准的评价,可通过长期观察得出结论。

◎定义 1.7:通过调查或观测而收集到的数据,称为**观测数据**(observational data)。

实验数据是对事物进行人为控制实验而收集到的数据,自然科学领域的大多数数据都为实验数据。如对农作物新品种在实验田一定的湿度、温度等条件,通过做实验收集到的数据。

◎定义 1.8:在实验中控制实验对象而收集到的数据,称为**实验数据**(experimental data)。

1.2.3 截面数据和时间序列数据

截面数据所描述的是现象在某一时刻的变化情况。例如,某校某专业 2009 届毕业生人数为 429 人,这就是截面数据,表明 2009 年这一时间点上的毕业人数数据。

◎定义 1.9:在相同或近似相同的时间点上收集的数据,称为**截面数据**(cross-sectional data)。

时间序列数据所描述的是现象随时间而变化的情况。例如,我国第“十一五”期间国民生产总值各期环比发展速度,就是从一段时间上描述其随时间变化的情况。

◎定义 1.10:在不同时间上收集到的数据,称为**时间序列数据**(time series data)。

1.3 统计学中常用基本概念

1.3.1 总体和样本

总体是统计研究的主体,是由多个个体单位构成的集合,组成总体的元素称为个体。如要研究武汉市工业企业经营状况,则武汉市所有从事工业生产的企业组成的集合称为总体,组成这一集合的每一个工业企业则称为个体。

◎定义 1.11:包含所研究的全部元素(数据)的集合,称为**总体**(population),通常以 N 代表总体全部单位的数目。

根据构成总体的全部元素是否确定,总体可划分为有限总体和无限总体。可确定总体元素的总体称为有限总体。例如,汽车公司一年的汽车产量是可以统计出来的,每一辆汽车是个体,该公司一年的汽车产量就是有限总体。无法确定总体元素的总体称为无限总体,如要研究某一新型洗衣机的销售量,要想知道哪些消费者喜欢,哪些消费者不喜欢,这些个体单位数目是无法确定的,由此构成的总体就是无限总体。

总体之所以分为有限总体与无限总体,主要是为了判别在抽样中每次抽取是否独立。对于无限总体,每次抽取一个单位,并不影响下一次的抽取结果,因此,每次抽取可看作独立的,对于有限总体而言,抽取一个单位,总体元素就会少一个,前一次的抽样结果往往会影响到第二次的抽样结果,所以每次抽取是不独立的。另外,对总体的概念还有一点需说明的是,总体的元素可以是实体单位的集合,也可以是研究某一现象的集合。如检验一批灯泡的寿命,这一批灯泡构成的集合就是总体,但是我们关心的或者说研究的是灯泡的“寿命”,所以这一批灯泡的寿命的集合也可作为总体。

◎定义 1.12:从总体中抽取的一部分元素的集合,称为**样本**(sample)。

◎定义 1.13:构成一个样本的元素数目,称为**样本容量**(sample size),或称为样本量。通常以 n 代表一个样本的数目。

如从武汉市全部工业企业中抽取 100 家企业进行调查,随机抽取的 100 家企业称为样本,而抽取的 100 家企业称为样本量。从总体抽取一部分元素作为样本,目的是要根据样本提供的有关信息去推断总体的特征。例如,我们从这 100 家企业的 2009 年平均工资中,可推算武汉市全部工业企业的 2009 年平均工资。

1.3.2 参数与统计量

◎定义 1.14:用来描述总体特征的概括性数字度量,称为**参数**(parameter)。

参数是研究者想要了解的总体的某种特征值。我们所关心的参数通常有总体平均数、总体标准差、总体比例等。在统计学中,总体参数通常用希腊字母表示。通常用 μ (mu) 表示总体均值,用 π (pai) 表示总体比例,用 σ (sigma) 表示总体标准差。

由于总体数据通常是不知道的,故参数是未知的。例如,我们不知道一个地区所有人口的平均年龄,不知道一批产品的合格率,不知道一个城市所有家庭的收入差异等。正因为如此,我们才进行抽样,根据样本数据计算统计量来估计总体未知参数。

◎定义 1.15:用来描述样本特征的概括性数字度量,称为**统计量**(statistic)。

统计量是根据样本数据计算出来的,它是样本的函数。通常我们所关心的样本统计量有样本平均数、样本比例和样本标准差等。样本统计量通常用小写英文字母表示,通常用 $\bar{x}$