

高等院校应用型人才培养“十三五”规划教材

统计学

TongJiXue

■ 主编 王军虎



 武汉理工大学出版社
WUTP Wuhan University of Technology Press

高等院校应用型人才培养“十三五”规划教材

统 计 学

主 编 王军虎
副主编 肖婷婷 张亚峰
鲁 瑜

武汉理工大学出版社

· 武 汉 ·

内 容 提 要

本书共分 10 章,内容包括导论,统计数据的收集,统计数据的整理、显示和对比分析,变量分布特征的统计描述,动态数列分析,统计指数,抽样及其分布,参数估计,假设检验,相关分析与回归分析等。本书体系完整、内容详细,既讲方法,又讲方法的原理和具体应用,还详尽介绍了 Excel、SPSS 和 EViews 软件的数据分析的操作过程和解读方法。

本书可作为高等院校经济管理类专业应用型人才培养的教材,也可作为从事科学研究的技术人员和自学者的参考资料,还可作为统计数据分析方法爱好者的阅读材料。

图书在版编目(CIP)数据

统计学/王军虎主编. —武汉:武汉理工大学出版社,2019.7

ISBN 978-7-5629-6026-3

I. ①统… II. ①王… 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(2019)第 144536 号

项目负责人:崔庆喜(027-87523138)

责任编辑:雷 蕾

责任校对:向玉露

封面设计:芳华时代

出版发行:武汉理工大学出版社

社 址:武汉市洪山区珞狮路 122 号

邮 编:430070

网 址:<http://www.wutp.com.cn>

经 销:各地新华书店

印 刷:武汉华工鑫宏印务有限公司

开 本:787×1092 1/16

印 张:20.5

字 数:525 千字

版 次:2019 年 7 月第 1 版

印 次:2019 年 7 月第 1 次印刷

印 数:2000 册

定 价:48.00 元

凡使用本教材的教师,可通过 E-mail 索取教学参考资料。

E-mail:wutpcqx@163.com 1239864338@qq.com

本社购书热线电话:027-87384729 87664138 87165708(传真)

凡购本书,如有缺页、倒页、脱页等印装质量问题,请向出版社发行部调换。

· 版权所有 盗版必究 ·

前 言

统计学是认识客观世界数量特征和数量规律的方法论学科,是教育部确定的经济管理类专业的核心课程之一。在新科技革命和产业变革奔腾而至的时代,教与学的关系正在发生着前所未有的变化。为了适应新形势下高等教育应用型人才培养对统计学教学的要求,我们在总结多年教学经验,吸收、借鉴国内外优秀成果的基础上编写了这本教材。

本教材具有下列特点:

1. 系统性。本教材按照先描述统计后推断统计的逻辑顺序,先易后难,循序渐进。在描述统计中,先静态描述后动态描述,先收集数据、整理数据后分析数据。在推断统计中,先介绍单一变量的分布规律,然后介绍对它的估计和检验,最后介绍多变量的相关和回归分析。在描述统计中,详细介绍了一般教材忽视的统计调查和统计汇总等内容。在推断统计中,全面介绍了除简单随机抽样以外的类型抽样、系统抽样、整群抽样和多阶段抽样的方法及其统计推断问题,使读者能全面了解推断统计的基本框架。在每一章最后都对所介绍的知识点进行了系统梳理,并列出了主要的计算公式。

2. 应用性。本教材围绕应用型人才培养目标编写,首先保证基本知识的系统和完整,然后详细介绍每一种统计方法的基本思想、应用条件和优缺点,并辅以经济管理中的实际应用,使读者在实践中正确判断有哪些统计方法可用于问题研究,选择什么具体统计方法解决问题。此外,每章最后还提供了大量的思考题、判断题、选择题、计算分析题等,通过练习有助于读者准确掌握和应用所学知识。

3. 实践性。本教材在介绍统计思想和统计方法的同时,特别强调现代信息技术的应用。在统计汇总、统计描述和统计推断中,详细介绍了3种国内流行的数据分析软件(EXCEL、SPSS和Eviews)的具体操作方法和统计结果的详细解读,使读者在准确应用统计方法的同时提高了效率。每章最后还提供了实训题,通过这些实训练习,读者能够亲身经历统计实践过程,用所学的统计知识解决实际统计问题。

4. 新颖性。本教材在多方面尝试着创新编写:在统计整理中详细介绍了计算机统计汇总的方法;在指数体系的因素分析中系统介绍了单一因素纯影响的分解和因素间交互作用的分析方法;在参数估计中增加了开区间估计和样本容量的确定,以及类型抽样、整群抽样时的必要样本容量;在假设检验中增加了区间估计决策准则,阐述了临界值准则、 p 值准则和区间估计准则的一致性,并增加了区间估计时必要样本容量的确定;为了体现统计方法的科学性,对于一些重要的数学推导,如总方差等于组内方差加组间方差、不重复抽样的标准误差等一般教材不做说明的内容,本教材通过增加附录的形式给出。

5. 灵活性。本教材的体系完整、内容详细,既讲方法,又讲方法的原理和应用,特别适用于自学的读者。在实际教学中,可以把实训题作为学生完成的目标任务,围绕完成目标任务展开教学讨论和讲解,实现以学习者为中心的教学过程。对于学时有限,不能完成全部教材内容教

学的,可以根据专业特点和教学目标要求,选择其中的部分内容进行教学,把教材其他内容作为课外扩展内容,丰富学习者的知识结构。

本教材由王军虎担任主编,负责编写第一章,第三章的第一、二、三节,第四章,第六章的第三节,第七章,第八章,第九章,第十章,以及各章的小结、练习和附录,并负责全书的设计、修改、总纂和统稿工作。参加编写的还有:肖婷婷(第二章)、张亚峰(第三章的第四、五节,第五章)、鲁瑜(第六章的第一、二、四节)。

编者在编写过程中参考了相关研究资料,在编写和出版过程中得到了武汉理工大学出版社的大力支持,在此一并表示衷心的感谢。

由于编者水平有限,教材中的不当和错谬之处在所难免,恳请广大读者批评指正。

王军虎

2019年2月

目 录

第一章 导论	(1)
第一节 统计与统计学	(1)
一、统计的含义	(1)
二、统计学的产生和发展	(2)
三、统计学的研究内容	(3)
第二节 统计数据类型	(4)
一、分类数据和数值数据	(4)
二、截面数据、时间序列数据和面板数据	(5)
三、调查数据和实验数据	(6)
四、原始数据和次级数据	(6)
五、个体数据、总体数据和样本数据	(6)
第三节 统计学的基本概念和基本研究方法	(7)
一、总体、个体和样本	(7)
二、标志、指标、参数、统计量和指标体系	(8)
三、变量	(11)
四、统计学的基本研究方法	(11)
本章小结	(12)
思考练习题	(13)
第二章 统计数据的收集	(17)
第一节 统计数据的收集方式和方法	(17)
一、统计数据收集的基本要求和种类	(17)
二、统计数据的调查方式	(18)
三、统计数据的收集方法	(22)
四、统计数据的间接来源	(23)
第二节 统计调查方案设计	(24)
一、明确统计调查目的	(24)
二、确定被调查对象、被调查单位和报告单位	(25)
三、确定调查项目、设计调查表或调查问卷	(25)
四、确定调查项目所属时间、调查期限和调查地点	(29)
五、确定调查的方式和方法	(29)
六、制订调查的组织实施计划	(30)
第三节 统计调查问卷设计	(30)
一、统计调查问卷的类型和结构	(30)
二、问题的设计	(33)

三、问题顺序的设计·····	(37)
本章小结·····	(38)
思考练习题·····	(39)
第三章 统计数据的整理、显示与对比分析 ·····	(41)
第一节 统计数据整理的基本问题 ·····	(41)
一、统计数据整理的内容·····	(41)
二、统计数据的预处理·····	(42)
第二节 统计数据分组和汇总 ·····	(43)
一、统计分组·····	(43)
二、统计汇总的方法·····	(46)
第三节 频数分布及其显示 ·····	(52)
一、频数分布的概念·····	(52)
二、分类变量频数分布及其显示·····	(53)
三、数值变量频数分布及其显示·····	(56)
第四节 统计指标的对比分析 ·····	(65)
一、相对指标及其表现形式·····	(65)
二、计划完成相对指标·····	(66)
三、结构相对指标·····	(67)
四、比例相对指标和比较相对指标·····	(67)
五、强度相对指标·····	(68)
六、动态相对指标·····	(69)
第五节 统计表 ·····	(69)
一、统计表的结构·····	(69)
二、统计表的种类·····	(70)
三、统计表的设计规范·····	(71)
本章小结·····	(72)
思考练习题·····	(73)
第四章 变量分布特征的统计描述 ·····	(77)
第一节 变量集中趋势的描述 ·····	(77)
一、算术平均值·····	(78)
二、几何平均值·····	(84)
三、众数·····	(84)
四、中位数和分位数·····	(85)
第二节 变量离散程度的描述 ·····	(89)
一、全距·····	(90)
二、四分位内距·····	(90)
三、方差·····	(91)

四、标准差	(94)
五、平均差	(95)
六、变异系数	(95)
七、异众数比率	(96)
第三节 变量分布形状的描述	(96)
一、偏态及其测度	(96)
二、峰度及其测度	(98)
第四节 利用计算机软件对变量进行统计描述	(99)
一、用 Excel 进行统计描述	(100)
二、用 EViews 进行统计描述	(101)
三、用 SPSS 进行统计描述	(101)
四、Excel、SPSS 和 EViews 统计描述结果的比较	(103)
本章小结	(104)
思考练习题	(106)
第五章 动态数列分析	(111)
第一节 动态数列及其编制	(111)
一、动态数列的概念	(112)
二、动态数列的种类	(112)
三、动态数列的编制原则	(114)
第二节 动态数列的指标分析	(115)
一、动态水平指标	(115)
二、动态速度指标	(121)
三、动态水平指标与动态速度指标的结合运用	(124)
第三节 动态数列的分解	(125)
一、动态数列的变动类型	(126)
二、长期趋势分析	(127)
三、季节变动分析	(133)
本章小结	(135)
思考练习题	(138)
第六章 统计指数	(142)
第一节 指数的概念、种类和作用	(143)
一、指数的概念	(143)
二、指数的种类	(143)
三、指数的作用	(144)
第二节 指数的编制方法	(144)
一、个体指数的编制方法	(144)
二、总量指标指数的编制方法	(145)

三、平均数指数的编制方法	(149)
第三节 指数体系与因素分析	(151)
一、指数体系的概念和作用	(152)
二、因素分析的要点和步骤	(152)
三、总量指标变动的两因素分析	(153)
四、总量指标变动的多因素分析	(156)
五、平均指标变动的因素分析	(159)
第四节 我国编制的主要统计指数	(162)
一、居民消费价格指数	(162)
二、商品零售价格指数	(165)
三、股票价格指数	(165)
四、工业生产者出厂价格指数	(166)
五、采购经理指数	(166)
本章小结	(167)
思考练习题	(170)
第七章 抽样及其分布	(174)
第一节 抽样推断概述	(174)
一、抽样推断的概念和作用	(174)
二、抽样推断的几个基本概念	(175)
第二节 抽样方式	(177)
一、选择抽样方式的基本原则	(177)
二、简单随机抽样	(178)
三、类型抽样	(179)
四、整群抽样	(179)
五、系统抽样	(180)
六、多阶段抽样	(180)
第三节 抽样分布	(181)
一、几个常用的随机变量分布及定理	(181)
二、简单随机抽样的分布特征	(186)
三、类型抽样的分布特征	(188)
四、整群抽样的分布特征	(190)
五、多阶段抽样的分布特征	(192)
本章小结	(193)
思考练习题	(195)
第八章 参数估计	(200)
第一节 参数估计的基本问题	(200)
一、点估计和区间估计	(200)

二、估计量的评价标准	(201)
第二节 一个总体参数的区间估计	(202)
一、一个总体均值的区间估计	(202)
二、一个总体比例的区间估计	(204)
三、一个总体方差的区间估计	(205)
四、一个总体参数的开区间估计	(206)
第三节 两个总体参数差异的区间估计	(207)
一、两个总体均值之差的区间估计	(207)
二、两个总体比例之差的区间估计	(210)
三、两个正态总体方差之比的区间估计	(211)
四、两个总体参数差异的开区间估计	(212)
第四节 区间估计时必要的样本容量	(213)
一、必要样本容量	(213)
二、估计总体均值置信区间时的必要样本容量	(213)
三、估计总体比例置信区间时的必要样本容量	(216)
本章小结	(217)
思考练习题	(219)
第九章 假设检验	(224)
第一节 假设检验的原理和步骤	(224)
一、假设检验的原理	(224)
二、假设检验的步骤	(225)
第二节 总体参数的假设检验举例	(232)
一、一个总体均值的假设检验举例	(232)
二、一个总体比例的假设检验举例	(235)
三、一个总体方差的假设检验举例	(236)
四、两个总体方差差异的假设检验举例	(236)
五、两个总体均值差异的假设检验举例	(237)
六、配对样本均值的假设检验举例	(238)
七、两个总体比例差异的假设检验举例	(238)
第三节 假设检验的两类错误和样本容量的确定	(239)
一、假设检验的两类错误	(239)
二、检验功效和假设检验时的必要样本容量	(240)
第四节 方差分析——多个总体均值差异的假设检验	(242)
一、方差分析的基本思想	(242)
二、水平间方差和水平内方差平均值的估计	(243)
三、水平间方差估计量与水平内方差平均值估计量比值的分布	(244)
四、方差分析的检验决策	(245)
五、多重比较法	(245)

六、两因素方差分析	(246)
第五节 用计算机软件进行假设检验	(249)
一、用计算机软件进行假设检验的特点	(249)
二、用 Excel 进行参数的假设检验	(250)
三、用 EViews 进行假设检验	(253)
四、用 SPSS 进行假设检验	(259)
五、用 Excel 和 SPSS 进行方差分析	(262)
本章小结	(265)
思考练习题	(266)
第十章 相关分析与回归分析	(271)
第一节 相关分析	(272)
一、变量间的关系	(272)
二、相关关系的种类	(273)
三、相关关系的描述与测度	(274)
四、皮尔逊线性相关系数	(275)
五、斯皮尔曼等级相关系数	(278)
第二节 一元线性回归分析	(279)
一、回归分析的概念和种类	(279)
二、一元线性回归模型的设定	(280)
三、一元线性回归模型的估计	(282)
四、一元线性回归模型的拟合优度	(285)
五、一元线性回归模型回归参数的显著性检验	(285)
六、一元线性回归模型的预测	(286)
第三节 多元线性回归分析	(288)
一、多元线性回归模型	(288)
二、多元线性回归模型的估计	(289)
三、多元线性回归模型的拟合优度、复相关系数和偏相关系数	(291)
四、多元线性回归模型的显著性检验	(294)
五、可化为线性的非线性回归模型	(295)
第四节 用计算机软件进行相关和回归分析	(296)
一、用 Excel 进行相关和回归分析	(296)
二、用 EViews 进行相关和回归分析	(298)
三、用 SPSS 进行相关和回归分析	(299)
本章小结	(300)
思考练习题	(302)
附表	(307)
参考文献	(315)

第一章 导 论

名人论统计

白衣天使南丁格尔说：“若想了解上帝在想什么，我们就必须学统计，因为统计学就是在测量他的旨意。”

我国著名经济学家马寅初指出：“学者不能离开统计而治学，政治家不能离开统计而施政，事业家不能离开统计而执业。”

英国科幻作家威尔斯则说：“对于追求效率的公民而言，统计思维总有一天会和读写能力一样重要。”

从上面的观点我们可以得出一个结论：统计无处不在，它已经渗透到我们工作和生活的方方面面。那么什么是统计？统计学研究哪些内容？我们在学习统计学的时候首先要掌握哪些基本的概念和术语？本章我们将对这些问题进行解答。

第一节 统计与统计学

一、统计的含义

统计是人们在认识客观事物数量的过程中形成的概念。人类已经进入信息社会，以数量为主要特征的信息收集、加工分析和利用，是互联网信息时代参与社会竞争的重要手段，各种统计正在深刻地影响着我们工作、学习和生活的方方面面。概括起来，人们对“统计”一词的使用有以下三个方面的含义：

1. 统计数据

统计数据，又称为统计资料或统计信息，是反映经济、社会、科技和自然等客观对象诸方面的数字特征及其相互联系的文字和图表材料。统计数据广泛存在于各种信息传播媒体中，如国家和各地方政府编纂出版的统计年鉴、统计资料汇编、统计图表，以及被引用在书籍、报纸、杂志和互联网络及其他媒介中的数据。

2. 统计工作

统计工作，也叫统计活动，是人们为了某种需要，对特定客观对象收集、加工、整理数据和使用数据进行统计分析与解释等方面的工作过程。统计数据是统计工作的主要成果。统计活动的主体可以是个人，也可以是企业、事业单位等社会组织，还可以是政府机关。在我国，各级政府机构基本上都设有专事统计工作的统计局，大多数的企业、事业单位都有专职或兼职的统

计人员,负责收集和加工整理企业日常活动中的各种数据用于内部管理,也有专门从事外包服务的统计调查咨询公司。

3. 统计科学

统计科学,简称统计学,是研究如何收集、加工整理统计数据,分析和解释统计数据的方法和技术的科学。在我国,许多大学设置了统计(学)专业,也有许多从事统计理论和应用研究的科技工作者。

统计的上述三个含义是密切联系的。首先,统计数据是统计工作的成果,统计工作是取得统计数据的工作过程。其次,统计学是统计工作的科学总结,是系统化了的理论知识体系。统计工作是在一定的统计理论指导下,采用科学的方法和技术收集、处理统计数据的一系列调查研究过程。

二、统计学的产生和发展

统计活动是人们认识客观世界的一项重要社会实践活动。自从人类结绳记事开始就有了统计活动。随着社会生产力的发展,阶级和国家的产生为统计活动的发展提供了客观需求。由于赋税征缴、征兵和徭役的需要,在古代奴隶制国家建立起了人口、土地和粮食等统计。但是在漫长的奴隶和封建社会中,缓慢的生产力发展对统计的数量和质量要求都比较低。随着资本主义生产力的迅猛发展,人们对各种生产资源和市场等方面的数据需求迅速增加,促使统计活动异常活跃,人口、工业、农业、商业、海关和物价等统计先后形成规模,相应的统计机构和研究组织先后建立了起来,促成了统计学的产生。

古典统计学产生于17世纪中叶至18世纪中末叶时期。从1660年开始,德国希尔姆斯特大学的康令(1606—1672)教授以“国势学”为题,通过记述和比较德国、法国、荷兰等国家的土地、人口、财政、军事、政治、宗教等方面的情况,为学生讲授政治活动家应当具备的知识。康令的主要后继者阿亨华尔(1719—1772)教授在德国的哥丁根大学开设了“国势学”,通过收集大量实际资料,分门别类地记述了关于国情国力的系统知识,并创造出了“statistik”即“统计学”(英文为 statistics)的学科名称。1662年,英国学者约翰·格朗特(1620—1674)出版了《关于死亡率公报的自然观察和政治观察》一书,收集和整理了1603年以来伦敦教会公布的死亡表,分析了60年来伦敦居民的死亡原因和人口变动关系,并用统计数据进行间接地推算,相互印证。他首先注意到,充分大量观察可使事物发展中的非本质的偶然因素相互抵消或削弱,从而显现出现象的稳定特征,这是被后来称为大数定理的统计学基石之一。1690年,英国学者威廉·佩第(1623—1687)出版了《政治算术》一书。书中运用大量统计数据,对英国、法国和荷兰三国的国情国力进行了系统的对比分析,阐明了英国的国际地位和发展方向,开创了国际比较的先河。约翰·格朗特和威廉·佩第等人的有关统计研究,创立了政治算术学派,为后来的经济统计,人口统计,医疗、卫生和保险统计等领域的研究提供了基础。与国势学相比较,佩第的政治算术更接近于现代的国民经济统计或统计核算,格朗特的“死亡表”更接近现代推断统计的萌芽。人们一般认为政治算术学派开创了真正统计学的先河。

18世纪末至19世纪末,概率论日渐成熟并与统计学的研究相融合,进入了近代统计学时期。比利时学者阿道夫·凯特勒(1796—1874)通过大量的观察和计算,对天文、气象、物理,尤其是社会现象进行了规律性的研究,把统计学建立在概率论的基础上,使德国的国势学、英国的政治算术、法国的概率论融合统一,提出了统计学是一门独立的方法论学科,它的研究对象

既可以是自然现象也可以是社会现象。在凯特勒工作的基础上,后经高尔登(1822—1911)、皮尔逊(1857—1936)等人的发展,形成了数理统计学派。此外,德国社会统计学家克尼斯(1821—1898)认为统计学是对社会现象进行大量观察和分析以揭示其内在联系和规律的科学。社会统计学派的主要研究领域包括统计指标的设定和计算、指数的编制、统计调查的组织和实施、社会经济评价与预测等,主要代表成果有德国统计学家恩斯特·恩格尔(1821—1896)提出的恩格尔系数、美国统计学家库兹涅茨和英国统计学家斯通提出的国民收入核算体系等。

从20世纪初开始,进入了现代统计学时期。在这一时期,随机抽样技术不断成熟,格塞特(1876—1937)提出了用于小样本推断的 t 分布,并在统计学家费歇尔(1890—1962)、耐曼(1894—1981)和皮尔逊(1895—1980)的推动下,建立了包括统计估计、统计假设检验和实验设计等内容的统计推断理论。同时,统计决策、贝叶斯统计、非参数统计、多元统计分析、时间序列分析和预测等新研究内容不断充实和完善。随着现代计算机技术和移动互联网技术的飞速发展和广泛普及,实现了数据收集加工和分析利用的革命性变革,统计数据的社会化、网络化和智能化趋势日益突出,并产生了数据分析、数据挖掘、大数据等新的学科分支。

三、统计学的研究内容

统计学是关于数据的学科。《不列颠百科全书》把统计学定义为:统计学是收集、分析、表述和解释数据的学科。可见,统计学是关于如何收集数据、整理分析数据和解释数据的方法论学科,它围绕统计工作的内容展开研究,为有效开展统计工作提供一系列的理论和方法。统计学的研究内容主要包括数据收集、数据整理、数据分析和数据解释四个部分。

1. 数据收集

数据收集是采集统计数据的统计工作过程,它是统计工作和统计研究的基础,没有统计数据,统计数据的收集方法和技术就失去了意义。统计数据收集的基本方法可以分为两大类:第一类是调查法;第二类是实验法。调查法是通过统计对象进行计数、测量和问询等收集数据;实验法是在实验中控制实验对象和环境等因素而收集数据。

2. 数据整理

数据整理是对统计调查或实验所得到的数据进行加工整理的过程。通过整理,使统计数据系统化、条理化,消除或淡化某些个别因素的偶然影响,反映总体数据的一般特征,为数据分析做准备。数据的整理内容包括数据的审核和筛选、数据的分组和分类,以及用统计表或统计图等表达方式显示整理结果。

3. 数据分析

数据分析是运用统计研究方法和技术探索数据之间的内在联系和规律的统计工作过程。数据分析是统计研究的核心内容。数据分析的方法包括统计描述法和统计推断法两种类型。统计描述法是运用总量指标、平均指标、相对指标和动态指标等研究总体现象的一般数量特征和数量联系,或依据一定的理论和假定建立统计模型,模拟数据的相互依存关系。统计推断法是依据掌握的部分数据的特征科学地估计总体数据的特征,或对总体数据特征估计的真实可靠性进行假设检验。

4. 数据解释

数据解释是对统计分析结果进行说明,它要回答的是统计结果为什么是这样的,数据结果背后隐藏的主要影响因素是什么,数据规律对现实工作有什么决策指导意义等。对数据的解

释过程需要用到统计研究对象所涉及的实质性学科的专业知识。例如,统计数据显示,企业生产量与单位生产成本大致呈反比例关系,从而得出扩大企业生产规模有助于降低单位生产成本的统计结论。然而,为什么会是这样的,还需要用经济学和管理学等相关学科理论加以解释。再如,利用统计方法对吸烟和不吸烟的肺癌患者数量进行研究,得到吸烟是导致肺癌的原因之一,但是吸烟诱发肺癌的机理需要医学知识提供解释。

此外,在调查具体统计对象的数据之前,还必须根据研究目的进行统计设计。统计设计是对统计数据收集、统计整理、统计分析的要求、方法、技术等内容和联系,以及各阶段工作的衔接配合,工作组织和工作条件安排等所做的通盘规划和统筹部署。进行统计设计,除了要具备扎实的统计理论和实践技能外,同样需要依托大量的实质性学科的专业理论和实践经验才能顺利完成。

第二节 统计数据的类型

统计数据是统计工作的成果,是由数字、文字等组成的用于表述和解释客观事物数量特征的事实依据或图表。统计数据按照所采用的计量尺度不同可分为分类数据和数值数据;按照被描述的对象与时间的关系不同可分为截面数据、时间序列数据和面板数据;按照收集方法的不同可分为调查数据和实验数据;按照数据的加工程度不同分为原始数据和次级数据;按照描述的对象不同分为个体数据、总体数据和样本数据等。

一、分类数据和数值数据

(一) 分类数据

分类数据是指用最简单的计量尺度把事物区分成不同的类别或属性。在统计研究中,分类数据主要用于对事物进行分类研究。分类数据还可以进一步细分为定类数据和定序数据。

1. 定类数据

定类数据也称为类别数据、名义数据,它仅仅是把事物按照某种属性进行分类,标以各种名称,表明其类别。例如,人口按照性别分为“男”“女”,大学生按照所学的专业分为“统计”“会计”“计算机科学”等都属于定类数据。定类数据的特点是它只是事物的类别或属性的代码,各种定类数据的地位是相等的,不能比较和排序,被称为无序分类数据。从数学特征上看,定类数据只能进行“=”和“≠”运算,即“是”和“不是”的判断。

2. 定序数据

定序数据也叫等级数据、顺序数据,是测度事物之间等级或顺序差别的数据。例如,把产品按照质量从优到差的顺序依次分为一等品、二等品、三等品和次品,学生的操行评定等级可以依次评定为优秀、良好、中等、及格和不及格等。定序数据比定类数据包含更多的信息量,不仅包含类别的信息,还包含各类别先后次序的信息,被称为有序分类数据。从数学特征上看,定类数据不但能进行“=”和“≠”运算,即“是”和“不是”的判断,还可以进行“>”和“<”运算,即“大于”和“小于”的判断。但是,不同定序数据的加总和是模糊的,不能精确地描述它们的总和及其之间的差别大小。

(二) 数值数据

数值数据是使用数值尺度测定事物数量特征的统计数据,因而数值数据之间的差别能够

精确地得到计量描述。统计研究中的数据主要是数值数据,它还可以分为定距数据和定比数据。

1. 定距数据

定距数据是用于测度事物类别或次序之间距离的数据。定距数据通常使用度量衡单位作为计量尺度,如温度用“摄氏度”度量、企业净利润用“万元”测度、海拔高度用“米”等计量,因此,定距数据的测度结果表现为数值。由于这一尺度的每一间隔都相等,只要给出一个度量单位,就可以精确地指出两个计数的总和及其之间的差值,如甲地区的温度 30 摄氏度与乙地区的温度 10 摄氏度之间相差 20 摄氏度,或者说乙地区的 10 摄氏度再升高 20 摄氏度可以达到甲地区的 30 摄氏度,但不能说甲地区的温度是乙地区的 3 倍,也不能说 2 个 10 摄氏度的地区的温度之和等于一个 20 摄氏度的地区温度。从数学特征上看,定距数据除了具备分类数据的全部性质,可以进行“=”“ \neq ”“ $>$ ”“ $<$ ”运算外,还可以进行“+”“-”运算。

2. 定比数据

定比数据是测度两个测度值之间比值的数据,它的测度结果同定距数据一样表现为数值。定比数据取值可以用度量衡单位表示,例如职工月工资收入 3500 元;也可以用实物量单位表示,例如企业产量 120 万件;还可以是其他单位,例如公司资产负债率 65%、劳动生产率 15 万元/人等。定比数据与定距数据的区别在于定比数据规定有一个固定的“零点”,而定距数据则没有。定距数据中的“0”并不是表示“不存在”,它仅仅表示一个测度值;而在定比数据中的“0”才是真正地表示没有,一个非 0 定比数值与一个 0 定比数值的比值等于无穷大。例如,对温度的测度,0 摄氏度只是一个普通的温度,并不表示没有温度,因此它属于定距变量;而体重属于定比数据,因为不存在体重为零的情况。定比数据是测量尺度的最高水平,它具有定类数据、定序数据和定距数据的全部特征,同时用数学规则求得两个定比数据的比值都具有现实意义,因此它可以满足数学上的一切运算规则的处理要求。由于定距数据和定比数据在绝大多数的统计分析中没有明显的差别,通常情况下并不需要严格地区分它们。

二、截面数据、时间序列数据和面板数据

1. 截面数据

截面数据描述的是某一相同的时间上不同空间范围事物即个体的数据,因此也叫静态数据,它是在同一时间或近似同一时间上收集到的个体数据的集合。例如,在第六次人口普查中得到的我国 2010 年 12 月 31 日 24 时的各地区人口数据就是截面数据。又如,2017 年全国各轿车生产厂家的年产量和年销售量都是截面数据。

2. 时间序列数据

时间序列数据描述的是同一现象的数据随时间变化而变化的情况,因此也叫动态数据,它是在同一空间上的不同时间收集的数据集。例如,2000—2016 年某公司各个月度销售收入的数据便构成了时间序列数据。

3. 面板数据

面板数据也叫平行数据或长期固定观察数据,它是截面数据和时间序列数据的组合。从横向看,面板数据具有截面数据的特征;从纵向看,面板数据又具有时间序列数据的特征。例如,在对我国城市居民家庭调查的几年内,每月对许多固定家庭收支进行连续登记而收集到的数据集合就构成了面板数据。如果面板数据的时间项数较多而个体数较少,称为长面板;如果

面板数据的时间项数较少而个体数较多,称为短面板。

三、调查数据和实验数据

1. 调查数据

调查数据是直接通过访问或观测而收集到的数据。例如,有关居民的性别、年龄、家庭收入、生活费支出等的数据都属于调查数据。调查数据是自然产生的数据,它是在不受人为控制的条件下产生的,因此具有客观性。社会经济研究中的统计数据基本上都属于调查数据。

2. 实验数据

实验数据是在实验中通过控制实验对象和条件,对实验结果进行观测记录而收集到的数据。例如,关于某种医疗方法的疗效数据,对某一小麦新品种的试种数据就属于实验数据。实验数据的产生一般受到实验者对实验结果的某种预期的支配,因而带有一定的主观选择性。自然科学研究领域中的数据多为实验数据。

四、原始数据和次级数据

1. 原始数据

原始数据是指通过统计调查或实验得到的反映研究对象个体特征,待整理的统计数据,也称为第一手统计数据。原始数据包含了最初收集到的全部数据信息,通过对原始数据的加工整理能够得到丰富的研究结果,因此它是统计数据收集的主体。

2. 次级数据

次级数据是指已经被加工整理过的统计数据,也称为二手数据或加工数据。次级数据既包括直接对原始数据进行整理的汇总数据,也包括根据已有数据用各种方法推算得到的统计数据。使用次级数据的成本低,能提高统计数据的利用率,是普通研究人员常用的统计数据类型。

五、个体数据、总体数据和样本数据

1. 个体数据

个体数据是指对某个特定个体进行调查或者实验从而得到的统计数据。个体数据也称为个体标志值,它是统计数据的基础。

2. 总体数据

总体数据是指对研究事物的全部个体进行调查或者实验,从而得到包含研究事物的全部信息的统计数据。总体数据既可以是研究事物的全部个体数据,也可以是用全部个体数据整理计算而成的,说明研究事物整体特征的统计数据。前者是总体的个体标志值集合,后者是总体的统计指标。用总体数据能够对研究的事物总体特征进行准确的分析研究。

3. 样本数据

样本数据是指对研究事物的一部分个体进行调查或者实验,从而得到关于研究事物的部分信息的统计数据。在统计研究中,得到总体数据往往会受到各种条件的限制,通常情况下是通过收集样本数据并加以分析,进而推断事物的总体特征。