



高职高专“十二五”规划教材

统计学概论

TONGJIXUE GAILUN



主编 王 栋 成凯文 汪丽萍

 吉林大学出版社

高职高专“十二五”规划教材

统计学概论

主 编 王 栋 成凯文 汪丽萍

副主编 周恩德 李良玉 王 刚

吉林大学出版社

内容提要

《统计学原理》是经济管理类专业的一门重要的专业基础课。统计数据信息是国家和企业正确制订国民经济发展规划和科学管理的重要依据,如何科学地分析和处理这些数据信息,是统计学所要解决的重要问题。本书共十章,内容包括导论、概率与概率分布、统计数据的收集、统计数据的初步整理、抽样推断与参数估计、假设检验、方差分析、相关与回归、时间数列分析、指数等。本书作为高职高专经济管理类专业的教材,主要考虑了理论的适度,强调统计知识原理的理解和应用。既适用于高职高专学生的学历教育,也适用于成人教育、函授大学教育,以及作为其他经济管理学科相关专业和干部培训的教材。也可作为高等院校有关专业本科、专科教学的参考书。

图书在版编目(CIP)数据

统计学概论/王栋,成凯文,汪丽萍主编. —长春

: 吉林大学出版社, 2010.6

(高职高专“十二五”规划教材)

ISBN 978-7-5601-5829-7

I. ①统… II. ①王… ②成… ③汪… III. ①统计学

—高等学校:技术学校—教材 IV. ①C8

中国版本图书馆 CIP 数据核字 (2010) 第 090977 号

书 名: 高职高专“十二五”规划教材

统计学概论

作 者: 王栋 成凯文 汪丽萍 主编

责任编辑、责任校对: 邵宇彤

吉林大学出版社出版、发行

开本: 787×1092 毫米 1/16

印张: 15.25 字数: 356 千

ISBN 978-7-5601-5829-7

封面设计: 超视觉工作室

北京市彩虹印刷有限责任公司 印刷

2010 年 6 月 第 1 版

2011 年 2 月 第 2 次印刷

定价: 29.00 元

版权所有 翻印必究

社址: 长春市明德路 421 号 邮编: 130021

发行部电话: 0431-88499826

网址: <http://www.jlup.com.cn>

E-mail: jlup@mail.jlu.edu.cn

出版说明

作为高等教育的重要组成部分，高等职业教育是以培养具有一定理论知识和较强实践能力，面向生产、面向服务和管理第一线职业岗位的实用型、技能型专门人才为目的的职业技术教育，是职业技术教育的高等阶段。目前，高等职业教育教学改革已经从专业建设、课程建设延伸到了教材建设层面。根据国家教育部关于要求发展高等职业技术教育，培养职业技术人才的大纲要求，我们组织编写了这套《高职高专“十二五”规划教材》。本系列教材坚持以就业为导向，以能力为本位，以服务学生职业生涯发展为目的的指导思想，以与专业建设、课程建设、人才培养模式同步配套作为编写原则。

从专业建设角度来说，相对于普通高等教育的“学科性专业”，高等职业教育属于“技术性专业”。技术性专业的知识往往由与高新技术工作相关联的那些学科中的有关知识所构成，这种知识必须具有职业技术岗位的有效性、综合性和发展性。本套教材不但追求学科上的完整性、系统性和逻辑性，而且突出知识的实用性、综合性，把职业岗位所需要的知识和实践能力的培养融会于教材之中。

从课程建设角度来说，现有的高等职业教育教材从教育内容上需要改变“重理论轻实践”、“重原理轻案例”，教学方法上则需要改变“重传授轻参与”、“重课堂轻现场”，考核评价上则需改变“重知识的记忆轻能力的掌握”、“重终结性的考试轻形成性考核”的倾向。针对这些情况，本套教材力求在整体教材内容体系以及具体教学方法指导、练习与思考等栏目中融入足够的实训内容，加强实践性教学环节，注重案例教学，注重能力的培养，使职业能力的培养贯穿于教学的全过程。同时，使公共基础类教材突出职业化，强调通用能力、关键能力的培养，以推动学生综合素质的提高。

从人才培养模式角度来说，高等职业教育人才的培养模式的主要形式是产学结合、工学交替。因此，本教材为了满足有学就有练、学完就能练、边学边练的实际要求，纳入新技术应用、生产案例介绍等来满足师生教学需要。同时，为了适应学生将来因为岗位或职业的变动而需要不断学习的情况，教材的编写注重采用新知识、新工艺、新方法、新标准，同时注重对学生创造能力和自我学习能力的培养，力争实现学生毕业与就业上岗的零距离接触。

为了更好地落实指导思想和编写原则，本套教材的编写者既有一定的教学经验、懂得教学规律，又有较强的实践技能。同时，我们还聘请生产一线的技术专家来审稿，保证教材的实用性、先进性、技术性。总之，该套教材是所有参与编写者辛勤劳作和不懈努力的成果，希望本套教材能为职业教育的提高和发展做出贡献。

这就是我们编写这套教材的初衷。

前 言

统计数据信息是国家和企业正确制订国民经济发展规划和科学管理的重要依据，如何科学地分析和处理这些数据信息，是统计学所要解决的重要问题。统计学主要通过产生、收集、描述、分析、总结和表达数据来达到发现新知识的目的。同时，它也研究如何对试验和调查进行设计，以便得到观测值的主要特征，特别是变异的种类和大小，试验以及调查数据相依性类型，是研究客观事物数量方面关系的方法论。统计学在自然科学、社会科学、军事科学以及人文科学中都有广泛的应用，特别是随着近年来计算机及软件技术的发展，使很多以前人工无法进行的统计计算成为可能，统计学的应用日趋普及。现在，统计学的应用几乎遍及所有科学领域和国民经济的各部门，统计学的一些基本概念和知识已成为很多社会生活和经济活动的必备常识。

在本书的编写过程中，我们从高职高专学生的特点出发，主要介绍了统计学原理和最基本、最重要的理论和方法，力求做到“重点突出，通俗易懂，结构合理，内容充实，语言精练”，并且每章都附有复习思考和练习题，可以使学生对所学的理论进一步加深认识和理解。

本书由王栋、成凯文、汪丽萍任主编，周恩德、李良玉、王刚任副主编。具体编写分工如下：湖北交通职业技术学院的王栋编写第一章和第七章，湖北汽车工业学院的周恩德编写第五章和第六章，河南专修学院的李良玉编写第二章和第四章，新余学院的成凯文编写第八章，平顶山工业职业技术学院的王刚编写第三章和第十章，九江职业大学的汪丽萍编写第九章。全书由王栋担任主审，由成凯文担任全书的最后统稿和定稿工作。

由于时间仓促，加上编者水平有限，书中难免有疏漏和不足之处，恳请使用本书的广大读者批评指正。

编 者

目 录

第一章 导 论	1
第一节 统计学的概念	1
第二节 统计学的发展历程	2
第三节 统计学的研究对象和研究方法	5
第四节 统计学的分类	8
第五节 统计学的几个常用名词	9
习 题	13
第二章 概率与概率分布	15
第一节 事件与概率	15
第二节 概率的运算法则	17
第三节 几个重要的概率分布	22
习 题	32
第三章 统计数据的收集	35
第一节 统计数据收集的作用及种类	35
第二节 统计数据的类型	36
第三节 统计数据收集的方法简介	38
第四节 统计调查	40
第五节 统计数据的质量	44
习 题	47
第四章 统计数据的初步整理	49
第一节 统计数据整理概述	49
第二节 数据结构的基本特征描述	51
第三节 数据的位置特征描述	60
第四节 数据的离散特征描述	70
第五节 数据的形态特征描述	74
习 题	79
第五章 抽样推断与参数估计	84
第一节 抽样推断概述	84
第二节 抽样误差	89
第三节 抽样方法	91
第四节 参数估计	96
习 题	102

第六章 假设检验	106
第一节 假设检验的基本概念	106
第二节 假设检验的基本步骤	111
第三节 假设检验的应用	112
习 题	120
第七章 方差分析	122
第一节 方差分析的概述	122
第二节 单因素方差分析	124
第三节 双因素方差分析	130
习 题	138
第八章 相关与回归	142
第一节 相关关系的判断	142
第二节 一元线性回归	146
第三节 利用回归方程进行预测	154
第四节 多元线性回归	155
习 题	160
第九章 时间数列分析	164
第一节 预 测	164
第二节 时间数列	167
第三节 时间数列预测法	178
习 题	186
第十章 指 数	190
第一节 指数的概念、类别及作用	190
第二节 指数的编制	192
第三节 指数体系	200
习 题	208
课后习题答案	211
附表 1 标准正态分布表	218
附表 2 正态分布概率表	220
附表 3 Z 分布表	224
附表 4 泊松分布表	225
附表 5 χ^2 分布表	226
附表 6 t 分布表	228
附表 7 F 分布表($\alpha=0.05$)	231
附表 8 F 分布表($\alpha=0.10$)	233
参 考 文 献	234

第一章 导论

学习目标

- 掌握统计学的概念、作用和特殊性。
- 了解统计学的发展历史。
- 理解统计学的研究对象及研究方法。
- 掌握统计学研究的一般步骤。
- 了解统计学的一般分类。
- 理解统计学常用的术语。

第一节 统计学的概念

科学家和艺术家有理解外部世界的共同愿望，他们希望用一些有条理的表达式来表现外部世界，进而减少外部世界明显的复杂性甚至混乱性。科学工作就包含着用有规则的方法表示无规则现象的内容。

科学可以分为实质性科学和方法论科学。如经济学、社会学、化学、物理学、生物学、天文学等属于实质性科学，它们都以社会现象和自然现象的内在联系和规律作为自己的研究对象，而哲学、数学等属于方法论科学，它们以如何认识社会现象和自然现象的内在联系和规律作为自己的研究对象。

人们在日常生活中，习惯凭借着先天的理解力识别出事物之间的关系和区别，但是这种直觉力有时很难识别出现存事物之间的关系和区别，有时却凭空想象出并不存在的关系或分布。科学家们则通过另外一种渠道来分析和认识这个世界，那就是当他们发现某些新的现象、相关性、趋势或其他效果的时候，通过提出假设，检验假设，判断所观测到的效果是否是在随机条件下产生的来得出研究结论。至于说如何来检验这些现象是否具备普遍性和规律性，就需要借助统计学知识来解决。事实上，统计学已成为近代科学方法的典型特征。

统计学是应用数学的一个分支，主要通过构建数学模型，收集所观察系统的数据，进行量化的分析、总结，并进而进行推断和预测，为相关决策提供依据和参考。因此，我们可以认为统计学是数据的科学和艺术：通过产生、收集、描述、分析、总结和表达数据达到发现新知识的目的。它研究如何对试验和调查进行设计，以便得到观测值的主要特征，特别是变异的种类和大小，试验以及调查数据相依性类型。借助统计学知识，我们可以对

一些问题做出答复,考察一些陈述的有效性。例如,现在公司的经营运行情况是否正常?每周锻炼身体两个小时对自己的心脏功能和循环系统有益吗?哪一种肥皂可有效防止细菌增生?产品的质量跟工人的工作环境改进有无关系?新的治疗方法对患者治疗的有效性如何?化学肥料 A_1 、 A_2 、 A_3 对粮食的产量的增加效果如何?等等。



小资料

1847 年, Semmelweis 不顾同事们的反对,在维也纳妇产科医院采用了一种治疗方法。虽然他并不知道产褥热的细菌本质,同时也不能直接证明他的试验是成功的,因为即使采用了他的方法,在他的医院里,仍有不少妇女死于产褥热,然而 Semmelweis 根据对许多产妇的计算发现,产妇的死亡率从 1840—1846 年的 10.7% 下降到 1847 年的 5.2%,又下降到 1848 年的 1.3%,因此得出这样的结论,他的治疗方法应当继续使用。

第二节 统计学的发展历程

统计学是一门很古老的科学,它起源于社会经济问题的研究,其学理研究可以追溯到 2300 多年前的古希腊。本书认为从统计学开始出现至今,经历了四个比较重要的发展阶段:统计学萌芽阶段(又被称为城邦政情阶段)、统计学创立阶段(又被称为政治算术阶段)、统计学发展阶段(又被称为数理统计学阶段)、现代统计学阶段。

一、统计学萌芽阶段

统计学萌芽阶段的代表人物是古希腊的亚里士多德,2300 多年前,他撰写了《城邦政情》(或译为《城邦纪要》),在这本书里亚里士多德一共撰写了 150 余种纪要,其内容包括各城邦的历史、行政、科学、艺术、人口、资源和财富等社会和经济情况的比较与分析,具有社会科学特点。“城邦政情”式的统计研究延续了一两千年,直至 17 世纪中叶才逐渐被“政治算术”这个名词所替代,并且很快被演化为“统计学”(Statistics)。由此可见,统计学的产生和发展是和国家的紧密结合的,至今统计学依然保留了城邦(State)这个词根。



小资料

统计语源最早出现于中世纪拉丁语的 Status,意指各种现象的状态和状况。由这一词根组成意大利语 Stato,表示“国家”的概念,也含有国家结构和国情知识的意思。根据这一词根,最早作为学名使用的“统计”,是在 18 世纪德国政治学教授亨瓦尔(G. Achenwall)在 1749 年所著《近代欧洲各国国家学纲要》一书的绪言中,把国家学名定为“Statistika”(统计)这个词。原意是指“国家显著事项的比较和记述”或“国势学”,认为统计是关于国家应注意事项的学问。此后,各国相继沿用“统计”这个词,并把这个词译成各国的文字,法国译为 Statistique,意大利译为 Statistica,英国译为 Statis-

tics, 日本最初译为“政表”、“政算”、“国势”、“形势”等, 直到 1880 年在太政官中设立了统计院, 才确定以“统计”二字正名。1903 年(清光绪廿九年)由钮永建、林卓南等翻译了四本横山雅南所著的《统计讲义录》一书, 把“统计”这个词从日本传到我国。1907 年(清光绪卅三年)彭祖植编写的《统计学》在日本出版, 同时在国内发行, 这是我国最早的一本统计学书籍。“统计”一词就成了记述国家和社会状况的数量关系的总称。

二、统计学创立阶段

统计学创立阶段的代表学派主要有两个: 政治算术学派和国势学派。

政治算术产生于 19 世纪中叶的英国, 创始人是威廉·配弟(W. Petty, 1623—1687), 代表作是《政治算术》一书。《政治算术》的特点是统计方法与数学计算和推理方法开始结合, 分析社会经济问题的方式更加注重运用定量分析方法。在这部著作中, 他利用实际资料, 运用数字、质量和尺度来说话的方法对英国、法国和荷兰三国的国情、国力作了系统的数量对比分析, 从而为统计学的形成和发展奠定了方法论基础。威廉·配弟用数字、质量和尺度将社会经济现象数量化的方法是近代统计学的重要特征。因此, 威廉·配弟的《政治算术》被后来的学者评价为近代统计学的来源, 威廉·配弟本人也被称为近代统计学之父。



小资料

“政治算术”的另一个代表人物是约翰·格朗特(J. Graunt, 1620—1674)。17 世纪上半叶, 英国多次发生严重的瘟疫, 政府定期公布有关人口出生和死亡的数字。约翰·格朗特利用这些资料研究并发表了《关于死亡表的自然观察和政治观察》的论著, 首次提出通过大量观察, 可以发现新生儿性别比例具有稳定性和不同死因的比例等人口规律; 并且第一次编制了初具规模的“生命表”, 对年龄、性别、死亡率与人口寿命作了分析, 从而引起普遍的关注, 他清楚地说明了统计学作为国家管理工具的重要作用。

国势学派又称记述学派, 产生于 17 世纪的德国。由于该学派主要以文字记述国家的显著事项, 故称记述学派。其主要代表人物为海尔曼·康令(H. Conring, 1606—1681)和阿亨华尔(G. Achenwall, 1714—1772)。康令第一个在德国黑尔姆斯太特大学以“国势学”为题, 讲授政治活动家应具备的知识。阿亨华尔在哥廷根大学开设“国家学”课程, 其主要著作是《近代欧洲各国国势学纲要》, 书中讲述“一国或多数国家的显著事项”, 主要用对比分析的方法研究了解有关国家组织、领土、人口、资源财富等国情国力, 比较了各国实力的强弱, 为德国的君主政体服务。因在外文中“国势”与“统计”词义相通, 后来正式命名为“统计学”。该学派在进行国势比较分析中, 偏重事物性质的解释, 而不注重数量对比和数量计算, 但为统计学的发展奠定了社会经济理论基础, 属于实质性社会科学。因此, 这个学派又被统计史学家称为有其名而无其实。

三、统计学发展阶段

统计学发展阶段的代表学派主要有两个: 数理统计学派和社会统计学派。

数理统计学派产生于 19 世纪中叶, 它是在概率论已有相当发展的基础上, 把概率论

引进统计学而形成的。其奠基人是比利时的阿道夫·凯特勒 (A. Quetelet, 1796—1874), 其著作主要有《论人类》、《概率论书简》、《社会制度》和《社会物理学》等。他深信统计学是可以用于研究任何科学的一般研究方法, 并正式把古典概率论引进统计学, 使统计学进入一个新的阶段。他最先用大数定律论证了社会生活现象纷繁复杂变化的偶然性中存在着一定的规律性, 并提出了误差理论, 在方法论方面解决了统计观测的准确性问题。他把概率论引入统计学, 使统计学在“政治算术”所建立的算术方法的基础上, 在准确化道路上大大跨进了一步, 为数理统计学的形成与发展奠定了基础。直到 1867 年, 才有人把这一门既是数学, 又是统计学的新生学科命名为数理统计学。



小资料

现代统计学的理论基础概率论始于研究赌博的机遇问题, 大约开始于 1477 年。数学家为了解释支配机遇的一般法则进行了长期的研究, 逐渐形成了概率论理论框架。在概率论进一步发展的基础上, 到 19 世纪初, 数学家们逐渐建立了观察误差理论、正态分布理论和最小平方法则。

社会统计学派兴起于 19 世纪后半叶, 该学派以德国为中心, 由德国经济学家、统计学家克尼斯 (K. G. A. Knies, 1821—1898) 创立, 主要代表人物有恩格尔 (C. L. E. Engel, 1821—1896)、梅尔 (G. V. Magr, 1841—1925) 等。社会统计学派认为统计学是一门社会科学, 是研究社会现象变动原因和规律性的实质性科学, 以此同数理统计学派通用方法论相对立。社会统计学派在研究对象上认为: 统计学是研究社会总体而不是个别的社会现象; 由于社会现象的复杂性和整体性, 必须对总体进行大量观察和分析, 研究其内在联系, 才能揭示现象内在规律。社会经济的发展, 要求统计学提供更多的统计方法; 社会科学本身不断地向细分化和定量化发展, 也要求统计学能提供更有的调查整理、分析资料的方法。因此, 社会统计学派也日益重视方法论的研究, 出现了从实质性科学向方法论转化的趋势。社会统计学派仍然强调在统计研究中必须以事物的质为前提和认识事物的重要性, 这同数理统计学的计量不计质的方法论从性质上说是本质区别的。

四、现代统计学阶段

现代统计学发展实际上还是数理统计学发展的延伸, 即其基础来源于数理统计。不同的是, 现代统计学阶段在数理统计的基础上建立了多种新的统计方法和数据处理技术, 与其他学科的融合也已经到了一个新的时期, 这个时期以推断统计的产生和发展为起点标志。所谓推断统计, 即通过随机样本来推断总体数量特征的方法。这种方法起源于英国数学家哥塞特 (W. S. Gosset, 1876—1937) 的小样本 t 分布理论。其后由费希尔 (R. A. Fisher, 1890—1962) 加以充实, 并由波兰统计学家尼曼 (J. Neyman, 1894—?) 以及 E. S. 皮尔生 (Person, 1895—?) 等人进一步发展, 建立了统计假设理论。后来美国统计学家瓦尔德 (A. Wald, 1902—1950) 又将统计学中的估计和假设理论加以归纳, 创立了“决策理论”; 美国的威尔克斯 (S. S. Wilks, 1906—1964)、英国的威沙特 (J. Wishart, 1898—1956) 等统计学家对样本分布理论又加以充实和发展; 美国的科克伦 (W. G. Cochran, 1909—1980) 等在 1957 年又提出了实验设计的理论和方法, 进一步拓宽了统计学的研究范围。

20 世纪 60 年代以后, 统计学与数学方法的结合越来越紧密, 随着社会、经济和科学技术的发展, 统计的范畴已覆盖了社会生活的一切领域, 几乎无所不包, 成为通用的方法论科学。它被广泛用于社会和自然界的各个方面的研究, 并发展成为有着许多分支学科的科学。同时, 随着计算机技术和网络技术的不断发展, 统计数据的搜集、处理、分析、存储、传递、印制等过程日益现代化, 提高了统计工作的效能。计算机技术的发展, 日益扩大了传统的和先进的统计技术的应用领域, 促使统计科学和工作发生了革命性的变化。如今, 计算机科学已经成为统计科学不可分割的组成部分。随着科学技术的发展, 统计理论和实践在深度及广度方面也不断发展。英国统计学家哈斯利特说: “统计方法的应用是这样普遍, 在我们的生活和习惯中, 统计的影响是这样巨大, 以致统计的重要性无论怎样强调也不过分。”甚至有的科学家还把我们的时代叫做“统计时代”。



小资料

20 世纪初至今, 以社会现象为研究对象的社会统计学也在积极地发展, 其基本趋势是由实质性科学向方法论转变。如二次大战后社会统计学派重要人物德国法兰克福大学教授弗拉斯卡姆波 (P. Flaskamper, 1886—?) 吸收了英国数理统计学派的通用方法论, 把自然科学中的方法应用于社会现象的研究。但是, 总的来说, 社会统计学的发展在 20 世纪 80 年代以前比较缓慢, 主要原因是社会现象本身非常复杂, 而且随着国际化趋势的加剧和科学技术的迅猛发展, 各种新的社会现象不断涌现, 使得社会统计科学的研究进展相对缓慢, 直到 20 世纪 80 年代中后期, 社会统计学研究借鉴心理学研究成果, 将各种社会现象之间的关系进行量表式开发, 同时借助计算机及其软件技术的发展, 才使社会统计学迎来新的快速发展时期。

第三节 统计学的研究对象和研究方法

统计学的研究对象和研究方法, 实际上是在不断发展变化的, 学术界在这两个问题上争议也较大。本书所讲的统计学研究对象和研究方法, 只是基于统计学本身所具有的主要特征来进行说明的, 目的是让大家更清楚、简明地理解统计学的研究对象和方法。

一、统计学研究对象

由统计学发展的历程, 我们可以知道统计学研究的对象是在不断地发展变化的, 研究的内容也是越来越广泛, 不能简单地用统计学的概念来推断其研究的对象。在这里我们只能用统计学研究的主要对象来进行阐述, 基于统计学研究的基础和对象离不开各种各样的数据, 本书认为, 统计学研究的主要对象为各种现象所表现出来的数据, 即统计学主要的研究对象为数据。统计学通过研究各种数据之间的关系来发现或解释数据背后隐含的一般性知识, 为决策提供依据或者说进行准备工作。事实上, 人们在利用统计学分析数据的时候, 感兴趣的不是数据本身, 而是产生这些数据的原始总体。例如: 掷骰子, 出现 6 的概率是人们所关心的。

二、统计学研究方法

统计学的研究方法很多,如果从表层来看研究方法的话,可以说现在所有已知的研究方法都可以被统计学所使用,甚至都可以成为统计学研究方法,因为统计学涉及的研究内容太广泛了,而且内容还在不断地被添加。本书不想罗列那么多的研究方法,只想探讨统计学最基本最常用的原理和方法,以使学生们理解这门学科,避免被庞杂的具体方法所困扰。

统计方法涉及从周围环境得来的数据及这些数据的收集和加工,即描述、估计、解释,目的在于准备做出决策。统计学是一门应用性非常强的学科,统计学的最主要目的也是在于应用实际,所以说统计学怎么才能更有效地应用于实际才是我们探讨统计学研究方法的关键。具体到统计学方法,我们先要明白一个统计学里经常用到的词语——模型,所谓的模型在统计学里面主要是指一个理论的正式表达方式或可由此生成观测到的数据。我们可以利用各种模型对现在所遇到的问题进行分析、解决,模型一旦建立,就是一个理论工具,会指导我们怎么去做类似的工作。

统计学的研究方法和典型的科学方法从本质上讲没有什么区别,即都是通过设计出一套有助于发现一般规律的策略,它们都需要借助用于检验和在拒绝时所做出的推断来实现,这些规律可以发展成一套有数学逻辑的结构理论,使我们能够获得可探知真实性的近似描述。它们两者脉络发展都必须经历:推断→(想法)步骤→观察→分析→结果→新的推测(新想法),如此不断地循环,在循环过程中相抵触的和矛盾的东西都要去掉,模型和理论因此得到改进,能使我们做出更多的结论、更好的预测,产生更好的理论。因此,我们可以把统计学研究简单地看做是,制作新模型或者使用已有模型来解决现实问题。统计学的研究难在如何科学地设计、使用或修订模型,这就需要各种的检测技术和统计数据处理方法来进行研究了,图1-1是统计学研究步骤的一个简略图。

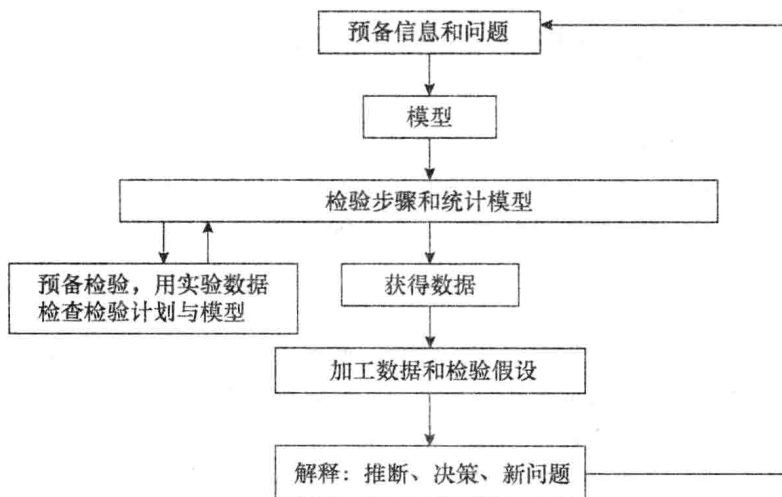


图 1-1 统计学研究步骤简图

从图中我们可以看到统计学研究的整个过程,具体到统计学研究方法,主要是指在整理数据和统计检验步骤时,统计学经常用到的研究方法问题,从某种意义上讲,这些方法

是统计学的特色方法，也是统计学最基本的方法，本书所要重点讲的也是这些具体的方法。在这里我们简要地介绍一下这几种常用的统计方法：

1. 调查或观察法

它是对大量性质相同的现象中每个观察单位的特点进行逐个观察、登记和研究，用以反映现象总体的数量表现，如计报表、普查、重点调查、抽样调查等。

2. 统计分组法

就是整理大量统计资料，按照社会现象总体单位的质的差别划分不同类型的组。统计分组是统计整理的关键。统计分组具有两方面的含义：对总体而言是“分”，即将总体中的各个个体按照它们的差异性区分为若干部分；对个体而言是“合”，即将性质相同的个体归并在一起。

3. 综合指标法

综合指标法是指运用各种综合统计指标，从具体数量方面对现实社会经济总体的规模及特征所进行的概括和分析的方法。在大量观察和分组基础上计算的综合指标，基本排除了总体中个别偶然因素的影响，反映出普遍的、决定性条件的作用结果。

4. 时间数列分析法

时间数列法是通过分析时间数列的组成要素来研究其变化形态，把过去的发展趋势延续下去和外推未来的预测方法。它的主要方法有移动平均法、加权移动平均法、指数平滑法、最小平方法，等等。

5. 指数分析法

指数分析法主要是指，利用指数体系对现象的综合变动从数量上分析其受各因素影响的方向、程度及绝对数量。现象的总体是复杂的，其发展变动受其构成要素变动的影响，但这些构成要素往往不可以直接相加，很难进行直接的观察比较，因此需要逐个因素进行分析，分析它们的变化对总体变动的影响程度和影响方向。

6. 相关分析法

相关分析法是在分析某个问题或指标时，将与该问题或指标相关的其他问题或指标进行对比，分析其相互关系或相关程度的一种分析方法。社会现象之间存在着大量的相互联系、相互依赖、相互制约的数量关系。这种关系可分为两种类型：一类是函数关系，它反映着现象之间严格的依存关系，也称确定性的依存关系。在这种关系中，对于变量的每一个数值，都有一个或几个确定的值与之对应。另一类是相关关系，指事物之间存在的一种不严格的依存关系，即对于某一现象的每一数值，有另一现象的一个或若干个数值与之相对应，所对应的数值在一定范围内变动。

7. 回归分析法

回归分析法是处理变量与变量之间关系的一种数学方法，它侧重于考察变量之间的数量伴随关系，并通过一定的数学表达式将这种关系描述出来，进而确定一个或几个变量（自变量）的变化对另一个特定变量（因变量）的影响程度。回归分析中，当研究的因果关系只涉及因变量和一个自变量时，叫做一元回归分析；当研究的因果关系涉及因变量和两个或两个以上自变量时，叫做多元回归分析。此外，回归分析中，又依据描述自变量与因变量之间因果关系的函数表达式是线性的还是非线性的，分为线性回归分析和非线性回

归分析。通常线性回归分析法是最基本的分析方法,遇到非线性回归问题可以借助数学手段转化为线性回归问题处理。

8. 抽样推断法

抽样推断法是在遵守随机原则的条件下,从总体中抽选样本,并且以样本指标推断总体指标的一种统计分析方法。这种方法主要用于难以进行全面调查或不宜进行全面调查的场合。当然在可以进行全面调查或其他非全面调查的场合,抽样调查仍然具有独到特点。抽样推断一般分为重复抽样和不重复抽样。抽样推断的组织形式主要包括简单随机抽样、分层随机抽样、等距随机抽样(又称系统或机械随机抽样)、整群随机抽样。

第四节 统计学的分类

从统计学产生和发展的历史可知,统计学最初是作为一门实质性科学建立起来的,它是从数量上研究具体的社会经济现象和自然现象发展规律的,即研究社会现象与自然现象中的数量特征和数量关系。随着统计学的研究范围的不断扩大和统计方法在各个领域的有效应用,使得统计学的研究对象也发生了变化,统计学逐渐从实质性科学中分离出来成为方法论科学。当然我们必须认识到统计学与其相联系的实质性科学又是密不可分的。首先,统计方法的使用取决于所研究具体对象的性质,不同领域现象的内在联系和规律是不同的,就要使用不同的统计方法;其次,对于统计研究结果的解释要依据所研究的实质性科学理论;最后,统计研究的成果使人们对实质性科学又有了新的认识,促进了实质性科学的不断发展和完善。

随着统计学被日益广泛地应用到社会科学和自然科学的各个领域,统计学也逐渐发展成为若干个分支学科组成的学科体系。根据人们的认知规律,统计学可以分为描述统计学与推断统计学。给定一组数据,统计学可以摘要并且描述这份数据,这种用法称作为描述统计学。另外,观察者以数据的形态建立出一个用以解释其随机性和不确定性的数学模型,以之来推论研究中的步骤及母体,这种用法被称做推断统计学。

一、描述统计学 (Descriptive Statistics)

描述统计学是研究如何取得所反映现象的数据,并通过统计图或统计表的形式对所搜集的数据进行加工和整理,进而利用一些综合性的指标来描述所研究现象的数量关系和数量特征。例如,证券交易每天都在公布各种股票的交易和涨跌信息,但是普通读者面对这样详细的数据,很难对股市的走势进行合理的判断。而在股票一段时间内价格变动资料的基础上所整理和计算出来的“上证指数”、“深圳指数”,以及绘制出的“K线图”,则能给读者提供一个清晰概要的股票变动情况。描述统计学里面的表现方式,主要包括图示方法和数量方法两大类。用统计图来表示一组数据的特征称为图示法,如柱形图、直方图、线形图、饼图等。用数量方法来反映一组数据的特征,是在所搜集数据基础之上计算一些综合性的指标,如总量指标、平均指标、相对指标、变异指标等。

二、推断统计学 (Inferential Statistics)

推断统计学是研究如何根据样本数据去推断总体数量特征的方法，它是在对样本数据进行描述的基础上，对统计总体的未知数量特征或未知的总体分布形式做出具有一定把握程度的推断。在现实条件下，由于各种因素的限制不可能搜集到有关总体的数据，或者有些研究推断不能使用总体数据，在这样的情况下就需要使用推断统计学的方法。例如，要评价某种酒，我们只能从一桶酒中取出少量部分进行品尝，然后根据取出的样本来对酒的整体品质作推断。

推断统计学的核心内容是统计推断问题，实质是以归纳的方法研究随机变量的一般规律。例如统计分布理论、参数估计和假设检验理论、相关与回归分析、时间数列分析等，这些方法普遍使用于社会现象和自然现象中随机变量的研究。

从运用统计方法研究问题的过程来看，描述统计学是整个统计学的基础，如果没有描述统计提供可靠的样本信息，即使再科学的推断统计方法也无法得到总体数量特征的准确结论。

第五节 统计学的几个常用名词

一、总体 (Population)

在统计学中，总体又称“调查总体”，简称“总体”，是指客观存在的、在同一性质基础上结合起来的许多个别单位的整体。构成总体的这些个别单位称为总体单位。例如，所有的工业企业就是一个总体，这是因为在性质上每个工业企业的经济职能是相同的，即都是从事工业生产活动的基本单位，这就是说，它们是同性质的。这些工业企业的集合就构成了统计总体。对于该总体来说，每一个工业企业就是一个总体单位。

总体根据其所包括的总体单位数是否有限，分为有限总体与无限总体。一个总体中包括的总体单位数是有限的，称为有限总体，例如，人口总数、企业个数、农田数量等。若总体是由无穷多个总体单位构成的，则称为无限总体，例如，某种产品的使用寿命。在实际工作中，为了便于问题的研究，有时也把一些有限总体近似地看成无限总体，例如，构成森林的总体单位——树木的数量是无法计数的，所以森林总体可以近似地看成是一个无限总体。

总体和总体单位是相对而言的，这是因为总体单位也有自己的构成。例如，要了解某市化工行业的生产经营情况，则该市所有化工企业构成一个总体，每个化工企业是总体单位。若把研究范围扩大到该市所有工业行业，那么除了化工行业外，还有机械行业、冶金行业、纺织行业等其他行业。这是，所有的工业行业构成了一个总体，化工行业就变成了这个新总体的一个总体单位。

二、样本 (Sample)

样本就是从总体中抽取的部分单位所构成的集合，其中的每一个单位称作样本单位。

例如,从100件产品中随机抽取20件,从上万名学生中随机抽取100人,等等。样本具有以下几个特点:样本单位必须来自总体内部,总体以外的单位不能参加样本抽取;从一个总体中可以抽取多个样本;样本必须具有代表性;样本必须具有客观性。

样本按照来源方式不同可以分为随机样本和非随机样本。随机样本又被称为概率样本,即按照随机原则从总体中抽取的样本,总体中每一个个体被抽到的机会均等,例如,在调查产品合格率时,应该采用随机抽样得来样本,以保证结果的代表性。而非随机样本,主要是指人们根据他们对被研究总体的了解有意识地抽取部分单位,或者人们只想对总体中某类个体进行研究时,就需要排除在其他类型样本个体被抽到的可能性,例如,我们要调查某大学艺术专业女生上网情况,就需要排除其他专业的女生样本进入本样本。需要指出的是,即便是非随机抽样样本,也应该遵循“随机”原则。

三、标志与标志表现

总体单位有许多不同的属性和特征,统计研究就要从这些属性和特征入手。总体单位的属性或特征称为标志,或者说标志是表明总体单位的名称,按性质不同可分为品质标志与数量标志。品质标志是表明总体单位属性的特征,是不能用数量来衡量的,例如,人的性别、民族、文化程度、职业等。数量标志是表明总体单位量的特征的,是可以用来衡量的,例如,人的年龄、身高,工业企业的产值、利润等。标志在各单位的具体表现称为标志表现,简称标志值。根据各单位的统一标志的标志表现是否相同,标志又可分为可变标志和不可变标志。不可变标志,就是所有总体单位同一标志的标志表现是相同的,可变标志是每个总体单位同一标志的标志表现各不相同。总体单位的不可变标志正是构成总体的必要条件,即“同质性”。例如,对国有企业总体进行研究,“经济类型”就是一个不变标志,而企业“年末资产总额”、“年利润率”等都是可变标志。

四、统计指标和指标体系

统计指标是十分重要的统计学基本范畴。对统计指标通常有两种理解和使用方法:第一种理解是用来反映总体现象总体数量状况的基本概念。这种理解的统计指标包括三个构成要素:(1)指标名称;(2)计算方法;(3)计量单位。例如,年末全国人口总数、全年国内生产总值、国内生产总值年度增长率等。第二种理解是指反映现象总体数量状况的概念和数值。这种理解的统计指标除包括上述三个构成要素,还包括三个要素:(1)时间限制;(2)空间限制;(3)指标数值。如,2008年年末我国全国就业人员77 480万人,国家外汇储备19 460亿美元,全年税收收入57 862亿元(不包括关税、耕地占用税和契税),等等。

统计指标主要有三个特点:(1)可量性;(2)综合性;(3)具体性。

统计指标按其作用功能不同可分为:描述指标、评价指标、预警指标。

统计指标按其所反映的数量特点不同可分为:数量指标和质量指标。

统计指标按照其反映的内容或其数值表现形式,可以分为总量指标、相对指标和平均指标三种。总量指标是反映现象总体规模的统计指标,通常以绝对数的形式来表现,因此又称为绝对数。例如土地面积、国内生产总值、财政收入等。总量指标按其反映的时间状