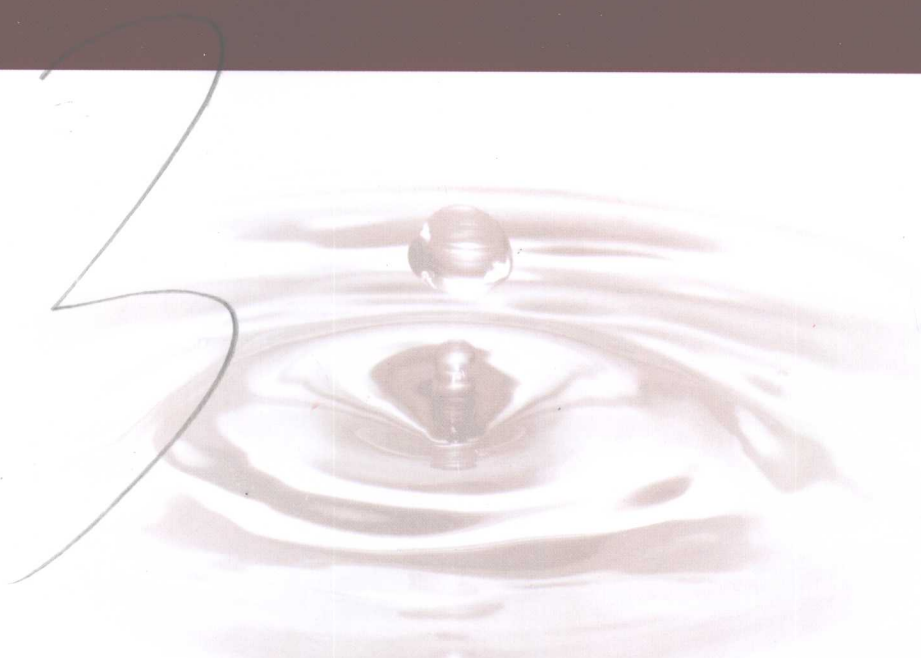


精·品·课·程·立·体·化·教·材·系·列



统 计 学

吴风庆 王艳明 主 编



 科学出版社
www.sciencep.com

精·品·课·程·立·体·化·教·材·系·列

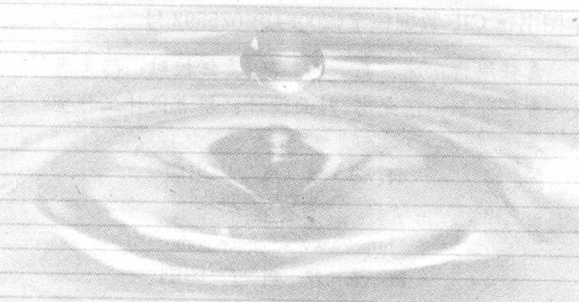
C8/182

2008



统 计 学

吴风庆 王艳明 主 编



科 学 出 版 社

北 京

内 容 简 介

统计学是收集、整理、分析、解释、使用数据并从数据中得出结论的科学。统计离不开数据,数据是统计的核心。作为数据分析的一种有效的科学工具,统计方法与技术已被广泛应用于理论与实践的各领域,是各领域理论研究和实际工作者必备的知识与技能。为此,我们编写了本书。本书主要包括描述统计(统计数据的收集、数据的整理与显示、数据分布特征的测度)、推断统计(抽样与参数估计、假设检验、方差分析)、经济管理常用的统计方法与技术(相关与回归分析、时间序列分析、统计指数、统计综合评价)等三部分内容。本书在借鉴国内外已有统计理论与实践研究成果的基础上,在内容与形式上又做了一些新的探索,力图使本书有一些特色、新意和适用性,从而更加适应新时期经济、管理类专业的统计教学和其他方面的需要。

本书既可作为高等院校经济管理类各专业本科生教材,也可作为广大实际工作者的参考书。

图书在版编目(CIP)数据

统计学/吴风庆,王艳明 主编. —北京:科学出版社,2008

精品课程立体化教材

ISBN 978-7-03-022128-5

I. 统… II. ①吴…②张… III. 统计学-高等学校-教材 IV. C8

中国版本图书馆 CIP 数据核字(2008)第 075592 号

责任编辑:陈 亮 马 跃/责任校对:陈玉凤

责任印制:张克忠/封面设计:耕者设计工作室

科 学 出 版 社 出 版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

骏 杰 印 刷 厂 印 刷

科学出版社发行 各地新华书店经销

*

2008 年 7 月第 一 版 开本:B5 (720×1000)

2008 年 7 月第一次印刷 印张:19 1/2

印数:1—4 500 字数:371 000

定价:30.00 元

(如有印装质量问题,我社负责调换〈环伟〉)



前言

《统计学》多年来一直是教育部确定的高等学校经济与管理类各专业的核心课程之一，原因如下：一是当今社会，数字化的信息无处不在，当你面对大量的形形色色的信息，作为一个信息使用者，我们需要具备阅读数据图表并理解这些数字信息的能力；二是统计学方法被广泛应用于生产、生活的各种决策，其应用甚至影响着个人生活的质量；三是统计学方法既可以帮助我们了解为什么会做出某一决策，而且清楚知道这些决策将会产生怎样的影响等。统计学如此重要，那么编写一本适用的教材就变得非常必要而有意义。

统计学是收集、整理、分析、解释、使用数据并从数据中得出结论的科学。统计离不开数据，数据是统计的核心。统计认识的对象主要是客观世界的数量方面，统计活动的目的主要是以数据来表述客观世界的数量规律与特征，统计分析的目标主要是依靠数据对客观事物的性质做出定性与定量的判断。离开了统计数据，统计方法也就失去了用武之地。如何取得较为准确的统计数据是统计学研究的内容之一。统计数据的整理是数据收集与数据分析之间的一个必要环节，是通过对统计数据的加工处理使其系统化、条理化，符合统计分析的需要。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。基于这一认识，鉴于财经院校经济管理类专业统计教学的实际需要，本教材的内容体系包括描述统计（统计数据的收集、数据的整理与显示、数据分布特征的测度）、推断统计（抽样与参数估计、假设检验、方差分析）、经济管理常用的统计方法（相关与回归分析、时间序列分析、统计指数、统计综合评价）等三部分内容。

在本书的编写过程中，我们一方面借鉴国内外已有的成果，另一方面也做了

一些探索,力图使本书有一些特色和新意,从而更加适应新时期经济、管理类专业的统计教学需要。

第一,通过本课程的学习,学生具备基本的统计思想,掌握基本的统计方法,培养运用统计方法分析和解决实际问题的能力。因此,本书注重对统计方法的阐述,结合大量的实际数据和实例说明统计方法的特点、应用条件和适用场合。既介绍具有通用方法论性质的一般统计理论与方法,又讨论社会经济领域特有的一些统计方法问题。

第二,在写法上,力求简明易懂。不仅精简了描述统计学的内容,而且对推断统计学的内容做了适当的取舍。尽可能做到结构合理,概念明确,条理分明,深入浅出。除十分必要外,本书一般不做过多的数学推导与证明,着重通过实例讲述统计思想,培养和提高学生应用统计方法的能力。

第三,加强计算机应用,提高学生学习的兴趣和学习效率。一是结合各章内容给出了用 Excel 进行统计计算和分析的步骤,可以使学生通过上机实现统计计算和分析任务,减轻了学生的计算负担;二是将统计方法与 Excel 软件有机结合,有利于培养学生应用统计方法解决实际问题的能力。

第四,在教材体例上,各章开篇有内容要点和教学要求提示,章末附有本章小结,对有关教学内容作扼要总结。各章后的“思考与练习”通过一些带有启发性的题目来帮助学生掌握有关概念和计算方法。

本书为中国科学院规划教材与精品课程立体化教材,由山东工商学院统计学院(山东省级重点学科与特色专业)的骨干专业教师完成。本书执笔人是:王艳明(第1、11章)、吴风庆(第2、4章)、苏永明(第3、11章)、王忠辉(第5章)、李丽(第6章)、张首芳(第7、10章)、何琳(第8章)、李志强(第9章)。吴风庆、王艳明任主编,负责全书的设计、修改、总纂和定稿工作。

本书的编写与出版得到科学出版社的大力支持,马跃同志为本书的组稿、编辑做了大量工作,我校许启发博士为有关图表的制作提供了很多帮助。在此,表示衷心的感谢。

本书的编写得到了山东工商学院统计学院全体教师的支持和帮助。从大纲拟定到教材内容编写,参加编写的老师及全体教师进行了充分讨论。大家提出了许多宝贵意见,并进行了多次修改,在此表示感谢。应当指出,尽管我们为提高教材的质量做了很多努力,但由于水平有限,本书难免有疏漏或错误之处,恳请同行专家和读者不吝赐教,以便今后进一步修改与完善。

编 者

2008年3月



录

前言

第1章

导论	1
1.1 统计基本问题	1
1.2 统计数据类型	7
1.3 统计学中的几个基本概念	8
附录 1.1 统计学的再认识	11
本章小结	13
思考与练习	14

第2章

数据的收集	19
2.1 数据的来源	19
2.2 调查方案设计	26
2.3 调查问卷设计	29
附录 2.1 第五次全国人口普查表短表	35

附录 2.2 第五次全国人口普查办法	37
附录 2.3 调查问卷	43
本章小结	44
思考与练习	44

第 3 章

数据整理与显示	48
3.1 数据的审核、筛选与排序	48
3.2 统计分类与频数分布	53
3.3 统计表	60
3.4 统计图	63
本章小结	69
思考与练习	70

第 4 章

数据分布特征的测度	75
4.1 分布的集中趋势	76
4.2 分布的离散程度	86
4.3 偏度与峰度	93
案例	97
本章小结	100
思考与练习	101

第 5 章

抽样与参数估计	109
5.1 抽样与抽样分布	110
5.2 一个总体参数的区间估计	123
5.3 两个总体参数的区间估计	128
5.4 样本量的确定	132

附录 5.1 有关证明	134
本章小结	134
思考与练习	135

第 6 章

假设检验	140
6.1 假设检验的基本问题	140
6.2 一个总体参数的检验	150
6.3 两个总体参数的检验	157
本章小结	170
思考与练习	171

第 7 章

方差分析	175
7.1 方差分析概述	175
7.2 单因素方差分析	180
7.3 双因素方差分析	186
附录 7.1 Excel 操作介绍	190
本章小结	193
思考与练习	193

第 8 章

相关与回归分析	200
8.1 变量间关系的度量	200
8.2 一元线性回归分析	207
8.3 多元线性回归分析	222
本章小结	227
思考与练习	228

第9章

时间序列分析和预测	232
9.1 时间序列的概念和构成	232
9.2 时间序列的平滑法	235
9.3 长期趋势分析	244
9.4 季节变动分析	248
附录 9.1 本章相关 Excel 操作介绍	252
本章小结	256
思考与练习	256

第10章

统计指数	261
10.1 指数的概念、种类	261
10.2 总指数的编制	263
10.3 指数体系及因素分析	267
10.4 几种常见的经济指数	268
本章小结	272
思考与练习	273

第11章

统计综合评价	277
11.1 统计综合评价的基本思路	277
11.2 常用统计综合评价方法	283
本章小结	289
思考与练习	290

主要参考文献	292
附表 1 标准正态分布表	294
附表 2 t 分布上侧分位数表	295
附表 3 χ^2 分布临界值表	296
附表 4 F 分布临界值表	298



第 章

导 论

掌握统计学，你就把握了数字的规律；
把握了数字的规律，你就把握了一切！

——佚名

本章从整体上对统计学的基本内容作简要介绍，为以后各章的学习奠定基础。具体要求：①正确理解与认识统计基本问题；②掌握统计数据的类型及用法；③掌握统计学的主要基本概念：统计总体、样本、参数、统计量、变量等。

在我们日常工作、生活和学习中，经常需要大量的信息，也会常常接触到“统计”（statistics）这一专业术语。例如，据 2007 年统计公报有关信息，我国国内生产总值为 24.66 万亿元，比 2006 年增长 11.4%。三种产业比重分别为 11.7%，49.2%，39.1%。居民消费价格比 2006 年上涨 4.8%。其中，食品价格上涨 12.3%。普通高等教育招生 566 万人，在校生 1885 万人，毕业生 448 万人。这些统计信息如何获得，如何正确地理解与运用统计信息，如何利用所获得的信息做出正确的决策，如何利用历史信息进行预测，如何依据样本信息得出有关总体的结论……这些问题的解决都离不开统计学。

1.1 统计基本问题

本节主要从统计本质论、统计任务论、统计方法论的角度，探讨统计是什么、统计干什么和统计怎么干等基本问题。

一、统计是什么

提出这个问题，主要是力图阐述统计的本质。“统计是什么”主要可以从统

计活动和统计学两方面来阐述。

所谓统计活动,是关于数据的收集、整理、分析、发布、使用的全过程的一种活动的总称。统计作为一项活动,早在原始社会,人们为了记录和比较均匀地分配其劳动成果,就产生了最初的统计需求和统计方法,即结绳记事。后来由于采集对象不同、大小不同,就在结的大小和形状上进行了区别,这就形成了最早的统计分类思想。可以说,统计分类是统计的基础。例如,某国家或地区的人口统计,只有在反映人口总量的同时也反映分类,即男性人口、女性人口;老年人口、中青年人口、儿童人口等分组统计的人口,人口统计才全面、才有意义,才能真正为国家或地区制定与人口有关的各项方针政策提供依据。统计作为一项管理活动,中外历代统治者都不同程度地利用统计来治理国家,在此就不一一列举。可以说,资本主义社会之前的统计活动,仅处于原始调查、登记和简单计数的初级阶段,没有涉及统计理论的探求。到16世纪,欧洲进入工场手工业时代,工业、商业、交通、航海业等都得到了空前的发展,统计活动开始从国家管理领域扩大到社会经济及管理的许多方面。

随着统计活动的不断发展和统计实践经验的日益丰富,作为一门科学的统计学也应运而生。关于统计学的阐述,从已出版的国内外统计学教科书来看,没有本质的差异,只是在表述上略有不同,都主要是围绕数据的搜集、整理、分析等来描述的。关于统计学的阐述主要有以下几种代表性的解释:

(1) 统计学是收集、分析、表述和解释数据的科学(《大不列颠百科全书》)。

(2) 统计学是收集、分析、表述和解释数据的艺术和科学(David R. Anderson, et al. 2006. 01. 商务与经济统计. 北京:机械工业出版社)。

(3) 统计学是收集、处理、分析、解释数据并从数据中得出结论的科学(贾俊平. 2006. 09. 统计学. 北京:中国人民大学出版社)。

(4) 统计学是对数据收集、组织、展示、分析和解释,从而帮助做出更为有效的科学的决策。(Douglas A. Lind, et al. 2005. 09. 商务与经济统计技术. 易丹辉等译. 北京:中国人民大学出版社)

(5) 统计学是有关如何测定、收集、整理、归纳、分析反映客观现象总体数量的数据,以便给出正确认识的方法论科学。(曾五一,肖红叶. 2006. 统计学导论. 北京:科学出版社)

(6) 统计学是用以收集数据,分析数据和由数据得出结论的一组概念、原则和方法。(吴喜之. 2004. 统计学:从数据到结论. 北京:中国统计出版社)

总结上述几种关于统计学的描述,可以将统计学的解释概括如下:统计学是收集、整理、分析、解释、使用数据并从数据中得出结论的科学。

统计离不开数据,数据是统计的核心。统计认识的对象主要是客观世界的数量方面,统计活动的目的主要是以数据来表述客观世界的数量关系,统计分析的

目标主要是依靠数据对客观事物的性质做出定性与定量的判断。统计数据的准确、及时、全面、方便,是统计活动的基本要求。统计研究客观现象的数量方面,主要是研究现象的数量多少,即现象的规模、大小、水平等(如某国家或地区人口数、工业企业规模、原煤总产量等);现象之间的数量关系,即研究现象的内部结构、比例关系、相关关系等(如性别比例、三次产业结构、收入与消费的关系等);质与量间的关系,即研究现象质与量互变的界限,研究质与量的统一。例如,完成计划与未完成计划具有质的差异,计划完成程度100%就是质与量互变的界限,对于一项正指标,如劳动生产率,当计划完成程度等于或大于100%即为完成或超额完成计划,低于100%则为未完成计划,至于完成110%还是120%对完成计划来说只是量的差异,没有质的区别。质与量的统一即要正确认识事物的量,必须以认识事物的质为前提,否则会得出错误的结论。例如要客观地认识我国居民收入状况,必须要正确认识我国的国情及社会发展的阶段。

二、统计干什么

“统计干什么”是从统计任务论的角度提出的问题。统计的主要任务可以概述为:通过科学准确地反映客观实际,以认识和把握客观现象的规律与特征为管理与决策服务。实践证明,统计学是一门应用性很强的学科,统计只有与其他实质性学科结合才能发挥其应有的作用。已故的陈希孺院士在就“统计学发展前景展望”答记者问时曾提到:“统计学研究应努力与其他实用学科结合而形成交叉或边缘学科。”著名统计学家C R Rao也提出:“统计学基本上是寄生的:靠研究其他领域内的工作而生存。”统计学与各实质性学科结合,并根据各自领域的特点和问题,开发和研究新的统计方法,是统计学学科发展的主流与研究方向。

目前,统计方法已被应用到自然科学和社会科学的众多领域,统计学科也已发展成为由若干分支学科组成的学科体系。可以说统计方法是适用于所有学科领域的通用数据分析方法,只要有数据的地方就会用到统计方法。下面,我们以统计学在某些领域的应用加以说明。

1. 经济分析

在当今的经济生活中,判断和认识经济形势及未来的走向,是政府、企业和家庭进行决策的基础和前提。对经济形势的分析判断,需要掌握有关统计信息与统计方法。例如,微观经济领域的专家需要使用统计方法为决策者提供生产、消费和定价方面的预测;宏观经济领域的专家既可以利用统计方法描述居民家庭收入分布状况,也可以对经济变量(如通货膨胀率、国内生产总值、失业率等指标)的未来水平进行分析预测,还可以研究变量(如消费和可支配收入)之间的关系。

2. 财务会计

财务会计主要以财务报表的形式来传达信息。上市公司必须有独立的外部审计机构来审计财务报表,以评价这些报表的合法性。统计抽样对选择合适的样本单位(如账目)进行审计有很重要的作用。例如,假设一事务所想确定上市公司的资产负债表上的应收账款金额是否真实。通常应收账款的数量很大,查看和验证每一账户要花费太多的时间和费用。这种情况下一般的做法是审计人员从所有账户中按照统计抽样技术,选择部分账户作为样本,在查看样本账户的准确性后,得出资产负债表上的应收账款金额是否是可以接受的结论。

3. 财务管理

财务管理是企业的一种职能,主要处理企业在提供产品和服务时的财务决策问题。在企业,财务部门常常运用统计学方法在不同的投资组合中做出决策,并且追踪财务指标随时间的变化趋势。例如,在股票市场中,财务顾问综合市盈率和股息等方面的财务统计数据,通过对比单只股票和股票市场平均价格信息,就可以得出某只股票其价值高估还是低估的结论。

4. 市场调研

营销部门运用统计学方法来估计偏爱某种产品的顾客比例,以及他们为什么喜欢这种产品,并由此做出决策;采取何种广告战略能够最有效地扩大产品销量。目前许多超市结账柜台的电子扫描盘正用于收集各种市场调研的数据。例如,数据供应商从零售商购买扫描数据,经过加工处理,将分析结果出售给制造商;制造商通过分析这些资料,掌握促销活动和销售额之间的关系等。这类分析对制定各类产品未来市场营销战略大有裨益。

5. 质量管理

生产部门改进产品质量或服务质量常常运用统计学方法来进行,质量控制是统计学在生产中的一项重要应用。在质量管理体系中,过程控制、数据分析、纠正与预防措施等很多要求均与统计技术应用有关,无论在质量管理哪一个阶段,统计技术作为一种有效的质量控制和质量改进的工具,都发挥了巨大的作用。例如,质量管理中的直方图主要用于显示质量波动的形态,直观地传递有关过程质量状况的信息,确定在什么地方进行质量改进工作等。

另外,统计方法也被大量应用到其他领域。例如,美国总统大选往往需要统计调查来获知选民对于各种问题及候选人的看法,获悉将是谁以多大的优势赢得选举;生物医学领域也大量用到统计学方法进行遗传学研究、药品研发、病理研究等;古往今来的战争中用统计方法设计作战方案,效果也十分明显。二战期间的“理智避开德军潜艇”的案例就给我们很多的启示。1943年以前,大西洋上的英美运输船队常常受到德国潜艇的袭击。当时,英美两国由于实力有限,又无力增派更多的护航舰艇。一时间,德国的“潜艇战”搞得盟军焦头烂额。为此,

一位美国海军将领专门请教了几位统计学家，他们经过统计计算分析发现，舰队与敌潜艇相遇具有一定的规律：一定数量的船编队规模越小，编次就越多；编次越多，与敌人相遇的可能性就大。美军海军接受了统计学家的建议，命令舰队在指定海域集合，再集体通过危险海域，然后各自驶向预定港口，结果盟军舰队遭袭被击沉的比例由原来的25%下降为1%，大大减少了损失。

以上所介绍的只是统计学在部分领域的个别应用，但是一叶落而知天下秋，由此我们可以看出，统计无处不在，统计大有可为。

三、统计怎么干

作为人类认识客观世界的一种活动，“统计怎么干”可以从其研究客观事物的全过程即基本环节加以认识。

1. 统计设计

统计设计是整个统计研究的前期工程，其完成质量直接关系到整个统计研究的质量。统计设计一般包括统计指标体系设计、统计调查方案设计、统计报表设计等。搞好统计设计不仅要有统计学的一般理论和方法为指导，而且还要求设计者对所研究的问题本身具有深刻的认识和相关学科的知识。例如，要设计一套反映国民经济运行情况的统计体系与方案，仅有一般的统计知识是不够的，还必须具备经济学的知识和理论素养，要知道国民经济运行的含义、特征、环节等内容。

2. 统计数据收集

经过统计设计，形成统计指标、方案或报表之后，就可以开始收集统计资料，收集数据也就是取得统计数据。统计数据的质量如何，直接关系到能否得出客观、正确的结论。数据的收集有两种基本方法，即实验法与调查法。实验法主要应用于自然科学和工程技术研究。统计学中的实验设计就是研究如何科学地设计实验方案，从而使得通过实验采集的数据能够符合分析的目的和要求。而对于社会经济现象来说，一般无法进行重复实验，要取得有关数据就必须到社会总体中去选取足够多的单位进行调查，并加以综合研究。如居民消费行为调查、电视收视行为调查等。如何科学地进行调查是统计学研究的重要内容。详细内容将在第二章讲授。

3. 统计整理与分析

原始统计数据收集上来之后，还必须经过整理、加工和分析才能发挥其作用。统计整理在统计活动过程中，处于承上启下的位置。一方面它是收集资料工作的继续，另一方面又是统计分析的前提，能够为统计分析与研究准备系统的资料。统计分析则是通过统计方法研究数据，其所用的方法可分为描述统计和推断统计。

描述统计学 (descriptive statistics) 是研究如何取得反映客观现象的数据, 并通过图表形式对所搜集的数据进行加工处理和显示, 进而通过综合、概括与分析得出反映客观现象的规律性数量特征。描述统计学的内容包括统计数据的搜集方法、数据的加工处理与显示方法、数据分布特征的概括与分析方法等。统计描述也是对客观事物认识的不断深化过程, 描述的过程也就是认识的过程。应该说统计对客观问题描述越全面, 认识就越深刻。

推断统计学 (inferential statistics) 是在对样本数据进行描述的基础上, 利用一定的方法根据样本数据去估计或推断总体的数量特征。在进行统计研究时, 常常存在这种情况: 由于各种原因, 我们所掌握的数据只是部分单位的数据或有限单位的数据, 而我们所关心的却是总体的数量特征。例如, 市场分析员需要了解不同广告对某种产品销售量的影响, 某种电子元件的寿命等等。这种情况需要利用统计推断的方法来解决。

描述统计学与推断统计学的划分, 还反映了统计方法发展的前后两个阶段和使用统计方法探索客观事物数量规律性的不同过程。统计研究过程的起点是统计数据, 终点是探索出客观现象内在的数量规律性。在这一过程中, 如果搜集到的是总体数据 (如普查数据), 那么运用描述统计就可以达到认识总体数量规律性的目的; 如果获得的只是研究总体的一部分数据 (样本数据), 那么要找到总体的数量规律性, 就要运用概率论的理论并根据样本信息, 对总体进行科学的推断。显然, 描述统计和推断统计是统计方法的两个组成部分。描述统计是整个统计学的基础, 推断统计在现代统计学中的地位和作用越来越重要, 已成为统计学的核心内容, 这是因为在对现实问题的研究中, 所获得的数据主要是样本数据。但这并不等于说描述统计不重要。如果没有描述统计收集可靠的统计数据并提供有效的样本信息, 再科学的统计推断方法也难以得出切合实际的结论。从描述统计学发展到推断统计学, 既反映了统计学发展的巨大成就, 也是统计学发展成熟的重要标志。

4. 统计资料的积累、开发与应用

统计的目的在于认识客观世界的规律。对于已经公布的统计资料需要加以积累, 同时还可以进行进一步的加工, 结合相关的学科 (经济学、管理学、医学等) 的理论知识去进行分析和利用。如何更好地将统计数据和统计方法应用于各自的研究领域是应用统计学研究的一个重要方面。

应该说, 统计的大量应用及发展很大程度上得益于信息技术的进步, 信息技术的进步使人们拥有了早期统计学家无法想象的计算能力。从 20 世纪六、七十年代开始, “统计软件包” 出现, 帮助用户自动完成统计分析中的计算部分。如 SAS、SPSS、Minitab 等统计分析软件和 Microsoft Excel 电子表格软件, 这些软件的应用大大地缩短了统计资料处理时间, 使统计工作者能够把更多的精力

集中于选择合适的统计方法以及对结果进行解释与分析上,很好地发挥了统计功能与效用。

1.2 统计数据的数据类型

获得有用的信息对于我们从事任何活动来说都是至关重要的。数据包含了在一个特定情况下做出更明确决策所需要的信息。在很多情况下我们都需要数据,如需要用数据来评价某种产品质量是否符合标准;需要用数据来帮助人们在决策过程中形成可供选择的行动方案等等。熟悉统计数据的类别性质,对我们在统计研究中有针对性地确定合适的统计分析方法很有帮助。

一、定性数据和定量数据

统计数据大体上可以分为两种类型:定性数据(qualitative data)和定量数据(quantitative data)。定性数据(亦称品质数据)是说明事物质的特征的具体表现,不能用数值表示。定性数据可分为分类数据(categorical data)和顺序数据(rank data)。分类数据和顺序数据都是反映现象类别特征的。分类数据之间,没有好坏、大小、优劣之分。如人口按性别划分,高校在校生按专业划分等;而顺序数据之间则有好坏、大小、优劣之分,如产品的质量等级、考试五级分等等。定量数据(亦称数量数据或数值型数据),是说明现象数量特征表现的,用数值来表示。如人口按年龄划分,企业规模按职工人数、固定资产规模或产品生产能力的划分等。

不同类型的数据,可采用不同的统计方法来处理和分析。比如,对定性数据的研究,我们感兴趣的通常是有多少或多大比例落入每一类别,即通过计算各组的频数或频率,来反映所分析现象的分布特征及规律。如某高校有多大比例的学生是女生,某门课程考试“优秀”学生所占比例,某地区工业企业中大、中、小企业的数量及比例等。定性数据通常以图表形式表示。而对定量数据的研究,我们感兴趣的除了关心有多少或多大比例落入每一类别外,还关心有关平均指标、现象之间的差异程度等。如根据某地区某年城镇居民收入资料,我们除了可以计算不同收入段的人数与比重外,还可以计算城镇居民人均收入以及收入的差异程度等,同时还可以通过收入与支出的数量关系,构建数量模型进行统计预测等。

二、观测数据和实验数据

按数据的收集方法分类,可将统计数据分为观测数据(observational data)

和实验数据 (experimental data)。观测数据是通过调查而收集到的数据,这类数据是在没有对事物进行人为控制的条件下而得到的,社会经济现象的统计数据几乎都是观测数据,如观察某路口每分钟的车流量,电视收视行为调查得到的数据等。实验数据是在实验中控制实验对象而收集到的数据。如医药研究实验数据、农作物杂交品种实验数据等等。自然科学领域的大多数数据都是实验数据。

三、截面数据和时间序列数据

根据对客观现象观察的角度不同,统计数据可分为截面数据 (cross-sectional data) 和时间序列数据 (time series data)。截面数据又称为静态数据,它是指在同一时间对不同单位的数量进行调查而获得的数据。时间序列数据又称为动态数据,它是指在不同时间对同一总体的数量进行调查而获得的数据。例如,2000年第五次全国人口普查人口地区分布数据(北京市 1382 万人,天津市 1001 万人,河北省 6744 万人,山西省 3297 万人,内蒙古自治区 2376 万人,上海市 1674 万人,江苏省 7438 万人,山东省 9079 万人等)就属于截面数据。“十五”期间我国历年的国内生产总值(2005 年 18.23 万亿元,2004 年 13.65 万亿元,2003 年 11.67 万亿元,2002 年 10.24 万亿元,2001 年 9.59 亿元)、我国 2004~2007 年普通高等教育招生人数(2004 年 447.3 万人,2005 年 505 万人,2006 年 540 万人,2007 年 566 万人)等就属于时间序列数据。

1.3 统计学中的几个基本概念

一、总体和样本

(一) 总体

统计学是研究客观现象总体的数量特征和数量关系。因此,首先对统计总体要有一个明确的认识。所谓统计总体 (population) 就是根据一定目的确定的所要研究的事物的全体。组成总体的各个个体称为总体单位 (简称个体)。例如,我们要研究全国城镇居民的收支情况,就以全国城镇居民作为一个总体,每户城镇居民为一个总体单位。有了总体与总体单位资料,我们就可以研究全国城镇居民的各种数量特征,如人均收入、人均消费、消费构成及特征等。

(二) 样本

统计研究的目的是要确定总体的数量特征。但是,当总体单位数量很多甚至无限时,不必要或不可能对构成总体的所有单位都进行调查。这时,需要采用一定的方式,从总体中,抽取一部分单位,作为总体的代表加以研究。这种由总体的部分单位组成的集合称为样本 (sample)。样本也是由一定数量的单位构成的,