

应用型高等教育“十二五”经管类规划教材



应用统计学

——基于SPSS运用

张良 主编
徐默莅 副主编

Applied Statistics



- ▶ 理论够用
- ▶ 突出实务
- ▶ 方法新颖
- ▶ 贴近实际
- ▶ 强调应用
- ▶ 富有特色

应用型高等教育“十二五”经管类规划教材

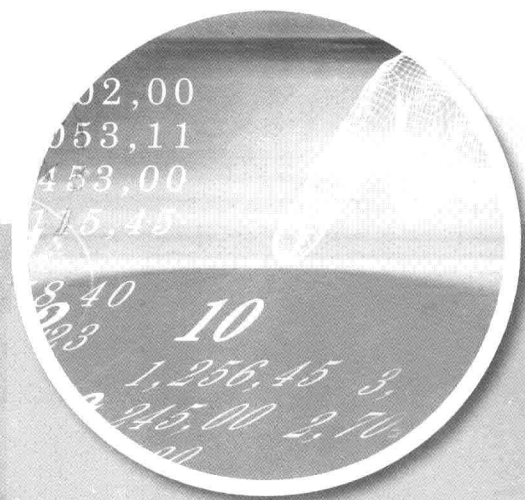


应用统计学

——基于SPSS运用

张良 主编

徐默菴 副主编



- ▶ 理论够用
- ▶ 突出实务
- ▶ 方法新颖
- ▶ 贴近实际
- ▶ 强调应用
- ▶ 富有特色

图书在版编目(CIP)数据

应用统计学:基于 SPSS 运用/张良主编,徐默菴副主编. —上海:
上海财经大学出版社,2013.9

(应用型高等教育“十二五”经管类规划教材)

ISBN 978-7-5642-1742-6/F·1742

I. ①应… II. ①张… ②徐… III. ①统计分析-软件包-高等学
校-教材 IV. ①C819

中国版本图书馆 CIP 数据核字(2013)第 186250 号

- 策 划 何苏湘
- 责任编辑 何苏湘
- 书籍设计 张克瑶
- 责任校对 林佳依 赵 伟

YINGYONG TONGJIXUE

应用统计学

——基于 SPSS 运用

张 良 主 编

徐默菴 副主编

上海财经大学出版社出版发行
(上海市武东路 321 号乙 邮编 200434)

网 址: <http://www.sufep.com>

电子邮箱: webmaster@sufep.com

全国新华书店经销

上海叶大印务发展有限公司印刷装订

2013 年 9 月第 1 版 2013 年 9 月第 1 次印刷

787mm×1092mm 1/16 14.25 印张 355 千字
印数:0 001—4 000 定价:35.00 元

前 言

应用统计学是一门以探索客观事物的内在规律为目标的方法论学科,在社会经济管理领域有着广泛的应用。应用统计学主要是从应用的角度阐述对数据进行搜集、整理和分析的方法和技术,能够指导人们科学地设计统计调查方案和调查问卷,科学地组织抽样调查,依据统计整理与分析技术处理大量数据,从中挖掘信息并作出科学的决策。因此,应用统计成为研究社会经济问题不可缺少的重要工具。

本书是一本“能力为本、知识够用”的新型统计学教材,紧紧围绕“统计职业技能的形成”这一主线,将统计岗位所需要的统计知识和实务相融合,力图体现职业型人才培养的要求,通过课内任务教学和课外实战演练的融合,体现侧重职业技能培养的课程改革思路,体现教师从主导者向引导者角色转变的要求,体现学生参与性学习、自主性学习的思想。

按照统计工作岗位的知识和技能的要求,全书共分为**课程导入、统计数据的搜集整理、统计数据的图表描述、统计数据的度量、统计抽样与参数估计、相关与回归分析、统计数据的动态分析、统计指数分析**八个项目,每个项目都由不同的任务构成。具体有以下特点:

(1)本书体系合理,力求选择在实践中用得较多的统计分析方法,去掉一些过于理论而在实践中运用不多的部分。

(2)本书以通俗易懂的语言讲述统计学中较为深奥的数学知识,通过图表的方式将相对深奥抽象的理论问题简单化,在每个重要的环节设置适量的例题或案例。

(3)教材突出理论联系实际的特点,以实际应用为案例,让学生感受到“学而有用、学而能用、学而会用”,给学生“统计就在我们身边”的亲切感,提高学生的学习兴趣。通过“任务导入”引出要讲授的内容和需要解决的问题。

(4)本书强调实训环节,对重要的统计方法均配有实例进行具体阐述。在每个项目后的实训材料中,结合 SPSS 软件进行实例演示,使复杂的计算过程变得特别的轻松并有意思。

(5)本书易教易学,各项目均按任务导入、原理介绍、技能训练、思考与练习、项目案例、项目实训等形式安排,使读者易学易练,通过练习来加深对统计方法的理解和掌握。本书所需配套教学资源,请向作者索取,Email:chnbzl@163.com。

本书具有很强的针对性、应用性和实践性,符合高校培养高技能应用型人才的目标要求,可以作为高等院校财经管理类专业的教材,也可供实际工作者和社会科学研究者学习参考。

本书由张良担任主编,并负责对全书修改和编纂,徐默莅担任副主编。在编写过程中,编者参考并吸取了许多统计教材和统计学研究成果的精华,在此特向著作者深表感谢。

由于作者水平有限,加之成书仓促,书中难免有错误或不妥之处,恳请读者及同行专家不吝赐教。

编 者

2013年6月

目 录

前言	1
课程导入	1
学习任务一 统计的含义与应用	3
学习任务二 统计学研究过程与研究方法	6
学习任务三 统计学的基本概念	8
技能训练 SPSS 软件安装及相关操作	10
知识回顾	24
思考与练习	24
项目一 统计数据的搜集与整理	27
学习任务一 获取统计数据的方式	27
学习任务二 获取统计数据的方法	30
学习任务三 调查方案的设计	31
学习任务四 统计调查工具	33
学习任务五 数据的整理	39
技能训练 用 SPSS 建立调查问卷的数据文件	40
知识回顾	44
思考与练习	45
项目二 统计数据的图表描述	59
学习任务一 用图表展示定性数据	61
学习任务二 用图表展示定量数据	67
学习任务三 用图表展示相关联数据	74
知识回顾	75
思考与练习	76
项目三 统计数据的度量	80
学习任务一 数据的相对度量	82
学习任务二 数据集中趋势的度量	87
学习任务三 离中趋势的度量	92
学习任务四 数据偏度与峰度的度量	95
技能训练 用 SPSS 对数据分布特征的描述	98
知识回顾	101
思考与练习	102



项目四 统计抽样与参数估计	105
学习任务一 统计抽样概述.....	105
学习任务二 抽样误差.....	109
学习任务三 参数估计的方法.....	116
学习任务四 抽样设计与组织方式.....	119
技能训练 用 SPSS 统计软件进行区间估计	123
知识回顾.....	127
思考与练习.....	127
项目五 相关与回归分析	132
学习任务一 相关分析.....	133
学习任务二 回归分析.....	139
技能训练一 用 SPSS 统计软件计算相关系数	147
技能训练二 用 SPSS 建立简单线性回归方程并进行回归分析和预测	150
知识回顾.....	154
思考与练习.....	154
项目六 统计数据的动态分析	158
学习任务一 动态数列概述.....	159
学习任务二 动态数列的水平分析.....	161
学习任务三 动态数列的速度分析.....	168
学习任务四 动态数列的趋势分析.....	171
学习任务五 动态数列的季节变动趋势.....	180
技能训练一 用 SPSS 软件进行长期趋势测定	184
技能训练二 用 SPSS 软件进行季节变动测定	190
知识回顾.....	193
思考与练习.....	194
项目七 统计指数分析	198
学习任务一 统计指数概述.....	199
学习任务二 综合指数的编制.....	201
学习任务三 平均指数的编制.....	205
学习任务四 指数体系和因素分析法.....	207
学习任务五 几种常用的价格指数.....	212
技能训练 用 SPSS 软件进行指数的计算与分析	216
知识回顾.....	218
思考与练习.....	218
参考文献	221



课程导入

知识目标:

1. 了解统计与统计应用
2. 理解和掌握统计学的基本概念
3. 了解 SPSS 软件

技能目标:

1. 能够根据统计研究的目的正确确定统计总体、总体单位、标志、指标
2. 学会安装 SPSS 软件及有关操作

重点难点:

统计学中的基本概念

任务导入:

在学习统计学之前,我们先来看一些统计学应用的实例。

实例一

2011 年国民经济和社会发展统计公报(节选)

截至 2011 年底,全国社会消费品零售总额 183 919 亿元,比上年增长 17.1%,扣除价格因素,实际增长 11.6%。按经营地统计,城镇消费品零售额 159 552 亿元,增长 17.2%;乡村消费品零售额 24 367 亿元,增长 16.7%。按消费形态统计,商品零售额 163 284 亿元,增长 17.2%;餐饮收入额 20 635 亿元,增长 16.9%。

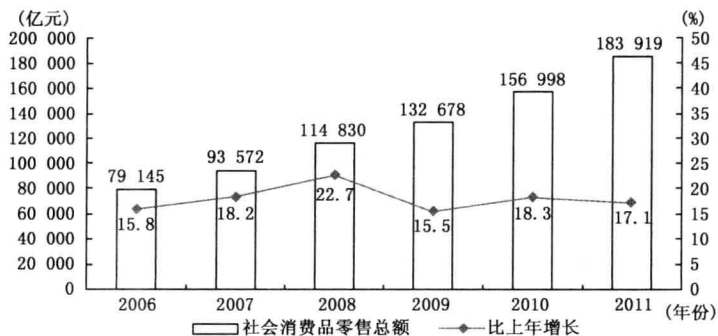


图 0-0 2006~2011 年社会消费品零售总额及其增长速度

在限额以上企业商品零售额中,汽车类零售额比上年增长 14.6%,粮油类增长 29.1%,肉禽蛋类增长 27.6%,服装类增长 25.1%,日用品类增长 24.1%,文化办公用品类增长 27.6%,通信器材类增长 27.5%,化妆品类增长 18.7%,金银珠宝类增长 42.1%,中西药品类增长 21.5%,家用电器和音像器材类增长 21.6%,家具类增长 32.8%,建筑及装潢材料类增长 30.1%。

实例二

2012 年 1~8 月份全国房地产统计资料(视频截图)



资料来源:<http://sz.house.sina.com.cn/video/2012-09-10/095382088.shtml>。

实例三

2012 年国家公务员考试试题(资料分析部分)

(一)根据以下资料,回答 116~120 题

2010 年,某省广电实际总收入为 145.83 亿元,同比增长 32.07%。其中,广告收入为 67.08 亿元,同比增长 25.88%;有线网络收入为 45.38 亿元,同比增长 26.35%;其他收入为 33.37 亿元,同比增长 57.3%。

2010 年,该省广电收入中,省级收入为 65.32 亿元,比上年增加 15.5 亿元;地市级收入为 41.61 亿元,比上年增加 13.39 亿元;县级收入为 38.90 亿元,比上年增加 6.52 亿元。

2010 年该省各市、县广电收入的区域分布如下:

东部地区 50.06 亿元,同比增长 32.48%,占市县收入份额的 62.18%,该地区的市均收入为 10.01 亿元,上年同期为 7.56 亿元。中部地区 17.78 亿元,同比增长 40.70%,该地区的市均收入为 5.93 亿元,上年同期为 4.21 亿元。西部地区 12.67 亿元,同比增长 80.86%,该地区的市均收入为 2.53 亿元,上年同期为 1.4 亿元。

截至 2010 年底,该省有线电视用户数为 1 885.88 万户,比上年末净增 161.7 万户。其中有有线数字电视用户为 1 007.8 万户,比上年末净增 277.58 万户。

116. 2009 年,该省的有线网络收入约为()亿元。

- A. 21 B. 36 C. 57 D. 110

117. 2009年,该省广告收入占广电总收入的比重约为()。
- A. 23% B. 26% C. 31% D. 48%
118. 2010年,该省地级市广电收入的同比增速约为()。
- A. 15% B. 20% C. 32% D. 47%
119. 该省有线电视用户平均每月的有线网络费用约为()元。
- A. 20 B. 36 C. 180 D. 240
120. 关于该广电收入的情况,下列说法正确的是()。
- A. 2010年,省级广电收入同比增长金额低于县级
- B. 2010年,东部地区广电收入超过中部地区的3倍
- C. 2009年,中部地区市均广电收入约是西部地区的3倍
- D. 2010年,有线电视用户中有线数字电视用户的比重不足一半

看了上面的资料,我们不禁要问:这些数据是统计数据吗?数据是如何得到的?这些数据能够给我们什么启示?另外,“明天是否会下雨”,“体育彩票你中了吗”,“子女为什么像父母”,“其相似度有多大”,“美国的民意测验是如何进行的”,“影响中国2012年‘双节日’长假高速公路拥堵的因素到底有哪些”,“中国的市场调查的可信度有多大”,等等问题,哪些是概率问题?哪些是统计问题?要回答好这些问题,就需要我们掌握一些统计学的知识。

同时,我们还能体会到统计已是人们在社会经济生活中必不可少的工具,是人们认识世界、探索现象数据差异的本质及其规律的有效方法,是人们进行明智决策的一门艺术。随着人类社会进入信息时代,统计作为一种方法和工具变得越来越重要。

学习任务一 统计的含义与应用

一、统计的含义

统计与人类社会活动密切相关。在人类的最初认识中,“统计”就是“计数”。小至家庭、个人,大至企业、国家都有计数的任务。一个月的收入、一年的利润都是我们经常关心的问题,这些数据就是统计的结果。世界各国都有各自的官方统计部门负责对人口、资源、环境和社会经济活动等各方面进行“计数”,并将这些数据资料以公共产品的方式定期公布,往往命名为“统计年鉴”。

在日常生活中,人们对于“统计”常常有不同的用法。例如,每年的高考结束后要“统计”考生的总分,这是将其作为一种工作来看待。了解股票的交易状况要看有关成交额和股票指数“统计”,这又是将其作为数据来运用。我们正在学习的“统计”,则是指一门学科,即统计学。

总之,统计是人们对客观世界总体数量变动关系和变动规律的活动的总称。它包含以下三种含义。

1. 统计工作

统计工作即统计实践,是对社会经济现象客观存在的现实数量方面进行搜集、整理和分析预测等活动的总称。一个完整的统计工作一般包括统计设计、统计调查、统计整理、统计分析等环节。

例如,银行的计划统计科每月编制项目报表,这个过程就是统计工作。又如,我国进行人口普查时要经过方案设计、入户登记、数据汇总、分析总结和资料公布等一系列过程,这些都是统计工作。在我国,各级政府机构基本上都有统计部门,如统计局,它们的职能主要就是从事统计数据的搜集、整理和分析工作。



2. 统计资料

统计资料(统计信息),是统计工作过程中所取得的各项数字资料和与之相关的其他实际资料的总称。例如:

(1)2011年末全国就业人员 76 420 万人,其中城镇就业人员 35 914 万人。年末城镇登记失业人数为 922 万人,城镇登记失业率为 4.1%。

(2)2011年,全国房地产开发投资 61 740 亿元,比上年增长 27.9%,增速比上年回落 5.3 个百分点,比 1~11 月回落 2 个百分点。其中,住宅投资 44 308 亿元,增长 30.2%,增速比 1~11 月回落 2.6 个百分点,占房地产开发投资的比重为 71.8%。

这些由文字和数字共同组成的数字化信息就是统计资料,是统计提供数据信息的基本表现形式,是统计工作的直接结果。

统计资料包括原始资料和整理后的资料(即次级资料)。例如,企业各车间的统计台账、人口普查时初次登记的资料都是原始资料,而统计公报、统计分析报告的部分现实和历史资料就是次级资料。统计资料的表现形式有统计表、统计图、统计分析报告、统计公报和统计年鉴等。

随着信息技术的发展与网络的普及,统计资料的公布不再仅仅是纸质资料了,大量的电子版的数据可以方便地从各国网站上获得,大部分都是免费的。我国统计资料的发布途径越来越规范,官方的统计数据通过“中国统计年鉴”、“中华人民共和国统计局网站”以及各省、市、地区的统计年鉴和官方统计局网站发布,一般都会同时提供纸质和电子版两种形式。

3. 统计学

统计学是系统论述统计理论和方法的科学,是长期统计工作实践的经验总结和理论概括。其中,应用纯逻辑推理方法研究抽象的随机现象的数量规律性的科学称为理论统计学,而应用统计方法研究各领域客观现象的数量规律性的科学称为应用统计学。社会经济统计学则是关于国民经济和社会发展数量方面的调查、整理和分析的原理、原则和方式方法的科学,按性质它属于应用统计学。

统计的三种含义之间具有密切的联系。

首先,统计工作和统计资料是统计活动与统计成果的关系。一方面,统计资料的需求支配着统计工作的局面;另一方面,统计工作的好坏又直接影响着统计资料的数量和质量。

其次,统计工作与统计学是统计实践与统计理论的关系。一方面,统计学来源于统计实践,只有当统计工作发展到一定程度,才可能形成独立的统计学;另一方面,统计工作的发展又需要统计理论的指导,统计科学研究就大大促进了统计工作水平的提高,统计工作的现代化和统计科学的进步是分不开的。

总之,三者中最基本的是统计工作,没有统计工作就不会有统计资料,没有丰富的统计实践经验就不会产生统计科学。

二、统计的应用

统计是适用于所有学科领域的通用数据分析方法和语言。现代人类的生活与统计活动密不可分,统计信息正成为人们了解世界变化的重要来源。购房者根据房产数据的变化来决定出手的时机;投资者根据换手率或指数来决定是否继续投资;政府机构根据国民经济统计数据(如 CPI)来决定是否干预市场;企业管理层根据财务数据和销售数据等调整企业经营决策;自然科学家通过各种观测数据来揭示自然现象……统计的应用领域可以用图 0-1 简单列示。

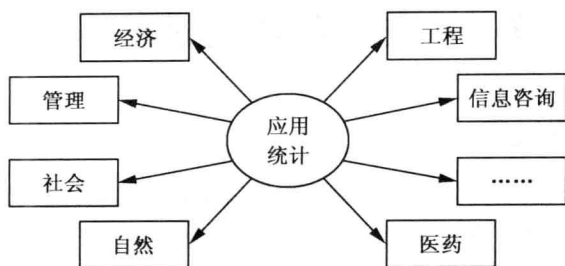


图 0-1 统计的应用领域

(一)统计在市场营销中的应用

对企业而言,新产品能成功推向市场的几率是比较低的,这涉及所设计产品是否被广大消费者所接受、新产品定价是否在市场可接受的范围内、新产品的宣传推广是否做到位等一系列问题。所以很多企业在推广新产品前都会进行市场调查,然后对调查结果进行分析,以决定新产品策略。例如,国际上广泛流行的结合统计学的顾客满意度指数可以较好进行顾客满意度分析;市场营销中用因子分析与聚类分析方法去进行市场细分,用主成分的统计特性去寻找企业产品与竞争厂商相似产品价格的近似线性关系,并以此作为定价依据。

(二)统计在金融中的应用

金融中很多数字的核算以及模型的建立需要统计学基础。在投资领域,分析师们要利用各种各样的统计信息进行投资指导。例如,证券分析师在推荐“买入”或“卖出”某只股票前,会对这只股票做多方面的评估。他们收集该上市公司过去的销售额数据,并估计未来的盈利。在推荐前,还需要考虑其他一些因素,比如,市场对该公司的产品具体需求、竞争力以及管理合同的效力等。

(三)统计在生产中的应用

生产型企业都会实行产品质量控制,以确保产品质量达到标准,这就需要借助统计技术。例如自动钻并不是钻出直径 1.3 厘米的洞(这是由于钻的磨损、机器的震动和其他因素)。细小的误差是允许的,但如果洞太小或太大,生产就是有缺陷的,该产品就不能被使用。质量保证部门承担着使用统计抽样技术来不断监控生产的任务。

(四)统计在会计中的应用

会计师事务所对其客户进行审计时要使用统计抽样程序。例如,假设一个事务所想确定列示在客户资产负债表上的应收账款金额是否真实地反映了应收账款的实际金额。通常应收账款的数量是如此之大,以至查看和验证每一账户将花费大量的时间和费用。在这种情况下,一般的做法是:审计人员从账本中选择一个子集作为样本,在查看样本账号的准确性后,得出有关列示在客户资产负债表上的应收账款金额是否可以接受的结论。

(五)统计在宏观经济中的应用

人们经常要求政府发布对目前的经济状况的统计结果并预测未来的经济走向。通常政府会进行大量的调查以了解消费者信心,并展望未来 12 个月与生产和销售有关的管理情况,每月编制诸如消费者价格指数等指数用于评估通货膨胀。百货商店销售额、住房开工货币周转额、工业生产量等信息仅仅是构成推测基础的上百项因素中的几项而已。这些评估被银行用来确定最佳放款利率,并用于确定货币储备的控制界限。

学习任务二 统计学的研究过程与研究方法

一、统计学的研究过程

统计学的研究过程就是从搜集个体现象的数据入手,在得到大量原始数据资料后,按照研究问题的需要对数据进行必要的整理,据此再对总体现象的数据特征进行分析和解释。可见,统计是一项高度集体性的工作,围绕预定的统计目的,需要每一环节紧密衔接。某一单位或个人只从事某一环节的工作,但却对全过程有至关重要的影响。统计学的研究过程如图 0-2 所示。

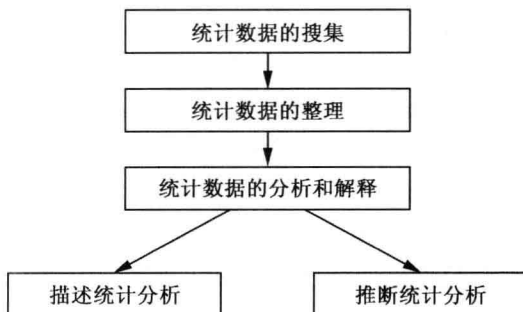


图 0-2 统计学的研究过程

一般而言,统计研究主要围绕以下内容展开。

(一)统计数据的搜集

统计数据的搜集是指对现象总体中全部或足够多的个体进行调查,搜集大量的以数字为主的信息资料,借以反映总体的数量特征。统计数据的搜集是统计工作的开始阶段,是进行统计整理和统计分析的基础。只有有了相应的统计资料,才能通过一定的统计方法和技术对数据进行整理、显示和分析,从而为决策提供依据。因此,统计数据搜集的质量如何,直接影响统计工作成果的质量。统计数据的搜集一般通过统计调查方式实现,如普查、抽样调查等。

(二)统计数据的整理

通过调查搜集的统计数据只能反映总体中各个个体的具体情况,还不能显示出总体的综合信息。统计数据的整理就是对这些分散的、不够系统的统计数据进行系统化的加工处理过程,使数据更加符合统计分析的需要。统计数据整理既是数据搜集环节的继续和深入,又是数据分析环节的前提和基础,在统计研究过程中起着承前启后的作用。统计数据整理的主要步骤包括:数据的审核,即查找数据中的错误,保证应用的数据准确可靠;数据的编码,即设计一套规则,并按照规则把文字形式的数据转化成数码符号形式的数据的全过程;数据的录入,即将编码后的资料输入到计算机内存储起来,以便计算机进行分类汇总。

(三)统计数据的分析和解释

统计数据的分析和解释是统计研究的核心内容和决定性环节,它是通过各种统计方法描述现有数量的状况、推断未知总体的数量特征,以达到探索数据内在规律的目的。常用的统计方法包括指标分析、趋势分析、回归分析等,这将在以后项目中详细说明。



二、统计学的研究方法

统计学作为一门方法论学科,具有自己完善的方法体系。统计研究的具体方法有很多,这将在后续课程中进行介绍,而从大的方面看,其基本研究方法有以下几种。

(一)大量观察法

大量观察法是统计活动过程中搜集数据资料阶段(即统计调查阶段)的基本方法,即要对所研究现象总体中的足够多数的个体进行观察和研究,以期认识具有规律性的总体数量特征。大量观察法的数理依据是大数定律,大数定律是指虽然每个个体受偶然因素的影响作用不同而在数量上存有差异,但对总体而言可以相互抵消而呈现出稳定的规律性,因此只有对足够多数的个体进行观察,观察值的综合结果才会趋向稳定,建立在大量观察法基础上的数据资料才会给出一般的结论。统计学的各种调查方法都属于大量观察法。

(二)统计分组法

由于所研究对象本身的复杂性、差异性及多层次性,需要我们对所研究对象进行分组或分类研究,以期在同质的基础上探求不同组或类之间的差异性。统计分组在整个统计活动中占有重要地位,在统计调查阶段可通过统计分组来搜集不同类的资料,并可使抽样调查的样本代表性得以提高(即分层抽样方式);在统计整理阶段可以通过统计分组法使各种数据资料得以分门别类地加工处理和储存,并为编制分布数量提供基础;在统计分析阶段可以通过统计分组法来划分现象类型、研究总体内在结构、比较不同类或组之间的差异(显著性检验)和分析不同变量之间的相关关系。统计学中的统计分组法有传统分组法、判别分析法和聚类分析法等。

(三)综合指标法

统计研究现象在数量方面的特征是通过统计综合指标来反映的。所谓综合指标,是指从总体上反映所研究现象数量特征和数量关系的范畴及其数值,常见的有总量指标、相对指标、平均指标和标志变异指标等。综合指标法在统计学、尤其是社会统计学中占有十分重要的地位,是描述统计学的核心内容。如何最真实客观地记录、描述和反映所研究现象的数量特征和数量关系,是统计指标理论研究的一大课题。

(四)统计模型法

在以统计指标来反映所研究现象的数量特征的同时,我们还经常需要对相关现象之间的数量变动关系定量研究,以了解某一(些)现象数量变动与另一(些)现象数量变动之间的关系及变动的影晌程度。在研究这种数量变动关系时,需要根据具体的研究对象和一定的假定条件,用合适的数学方法来进行模拟,这种方法叫作统计模型法。

(五)统计推断法

在统计认识活动中,我们所观察的往往只是所研究现象总体中的一部分单位,掌握的只是具有随机性的样本观察数据,而认识总体数量特征是统计研究的目的,这就需要根据概率论和样本分布理论,运用参数估计或假设检验的方法,由样本观测数据来推测总体数量特征。这种由本来推断总体的方法就称为统计推断法。统计推断法已在研究的许多领域中得到应用,除了最常见的总体指标推断法外,统计模型参数的估计和检验、统计预测中时间序列的估计和检验等,也都属于统计推断的范畴,都存在着误差和置信度的问题。在实践中这是一种有效又经济的方法,其应用范围广泛,发展迅速。统计推断法已成为现代统计学的基本方法。

学习任务三 统计学的基本概念

一、总体与样本

(一)总体

所谓总体就是根据一定的目的确定的所要研究的事物的全体,即某一特定研究中所有单元的集合,通常由客观存在的、具有相同性质的许多单元构成,而构成总体的每个单元被称为个体。个体是我们借以搜集数据的对象,可以由人、行政、企事业单位或物及事件充当。例如我们要研究全国城镇居民的收支情况,就要以全国城镇居民作为一个总体,其中每一户城镇居民家庭就是一个个体。成千上万不同的城镇居民家庭可以结合在一起构成总体,这是因为它们具有共同的性质,即他们都是我国的城镇居民,都有一定的收入和支出,都要消费一定的商品和服务。有了这个总体,我们就可以研究全国城镇居民的各种数量特征,例如人均收入、人均消费等。

根据构成总体的单位数是否有限,总体分为有限总体和无限总体。有限总体是指构成总体的个体数量是有限的、可数的。例如,某一时点某企业的在册人数或成品库存量是有限总体。再如,全国人口普查,尽管其包含的单位数量很大,但仍然是有限总体。无限总体是指构成总体的个体数量是无限的、不可数的。例如,要检验某种新工艺是否真正能够改善由流水线生产的产品性能的问题,由于新工艺有可能一直延续下去,利用该工艺制造的产品包括已经生产和将要生产的产品,其数量难以具体确定,因此属于无限总体。

(二)样本

样本是指从总体中抽取的部分个体组成的集合。统计研究的目的是认识总体的数量特征,但有时构成总体的个体数量非常大,实际工作中不可能或不必要对每个个体的数据特征逐一调查,通常是以某种方式从总体中抽取一部分个体代表总体加以研究。例如,某外贸公司从供货方提供的一批即将出口的茶叶中随机抽取了120袋,检验并推断该茶叶的质量,这120袋茶叶就构成一个样本。由此可见,样本是总体的代表,从总体中抽取样本的目的就是通过对样本数量特征的分析去推断总体的数量特征。

二、变量、数据与分类

(一)变量与数据

每天晚上收看天气预报,会发现今天的气温与明天的气温不同,今天是晴天,明天可能就是多云转阴;观察股票市场,会发现上证股指天天在变化;每个在职工作的人员从事的职业不同,月收入不相同;再观察成年人,每个人所受的教育程度也不相同。这里的“天气温度”、“天气形势”、“上证股指”、“职业”、“月收入”以及“受教育程度”等就是变量,它们的特点是从一次观察到下一次观察会出现不同的结果,把观察到的结果记录下来就是数据。

(二)数据的分类

1. 定性变量(数据)与定量变量(数据)

“天气温度”、“天气形势”、“上证股指”、“职业”、“月收入”以及“受教育程度”这些变量反映现象的特点不同。“天气形势”、“职业”和“受教育程度”是从现象的属性来表明现象的特征,如“天气晴”和“多云转阴”就是反映两种天气状况,“生产工人”和“公务员”就是两种不同的职业,文化



程度“小学”与“大学”就反映了两种完全不同的受教育程度。这样的变量称为定性变量,定性变量的观察结果称为定性数据。这类数据的最大特点是只能反映现象的属性特征,而不能说明具体量的大小和差异。如“天气晴”和“多云转阴”谁大谁小?“生产工人”和“公务员”谁好谁坏?这里没有量的特征,只有分类特征。这种只能反映现象分类特征的变量又称为分类变量,分类变量的观察结果就是分类数据。如果类别具有一定的顺序,如“受教育程度”,中学的教育程度就比小学高,大学又比中学高,这样的变量称为顺序变量,相应的观察结果就是顺序数据。

与定性变量不同的是,“天气温度”、“上证股指”、“月收入”这些变量可以用数值表示其观察结果,而且这些数值具有明确的数值含义,不仅能分类而且还能测量出具体大小和差异。这些变量就是定量变量,也称为数值变量,定量变量的观察结果称为定量数据。如职工的月收入可以表现为1 500元、1 820元、2 638元、3 100元,等等。

分类变量没有数值特征,所以不能对其数据进行数学运算。即使有时人们用若干个数值来代表不同的类别,这些数值本身也不具有数量差异的意义。例如,人们用编号1代表性别男,用编号2代表性别女,这些编号的主要作用在于使文字性概念成为计算机可识别的代码,但其本身并不具有计算意义上的数量差异。由此可见,分类数据只能用来区分事物,而不能用来表明事物之间的大小、优劣关系。

顺序变量比分类变量向前进了一步,它不仅能用来区分客观现象的不同类别,而且还可以表明现象之间的大小、高低、优劣关系。如某一门课程的考核成绩的表现可以是优、良、中、及格、不及格,人们对某种商品售后服务的态度可以表现为非常满意、满意、一般、不满意、很不满意等。显然,顺序数据的功能比分类数据要强一些,对事物的划分也更精细一些。需要注意的是,顺序数据经常会以数值的形式出现,如产品质量可以分为1、2、3级品,这里的1、2、3虽然是以数值的形式出现的,但仍然是用来反映产品之间在质量上的性质差异。因此,从本质上说,顺序数据仍然是定性数据的一种。

定量变量的特征在于它们都是以数值形式出现的。有些定量数据只可以计算数据之间的绝对差,而有些定量数据不仅可以计算数据之间的绝对差,还可以计算数据之间的相对差。显然,定量数据的计量功能要远远大于前面介绍的两种定性数据,其计量精度也远远高于定性数据。

就统计分析而言,区分数据的类型是非常重要的。因为根据不同类型的数据,我们将采用不同的统计整理技术和分析方法。例如,对于定性数据中的分类数据我们主要是进行分类整理、汇总,计算出各数据的频数和频率,分析方法主要是观察各类数据占总体的比重,以研究社会经济现象总体内部的结构,或用于比较不同组别的水平差异等;对于定性数据中的顺序数据我们不仅可以进行分类整理、汇总,计算出各类数据的频数或频率、观察各类数据占总体的比重,还可以进一步观察顺序排列的累计观察值、中位数、分位数等;而对于定量数据我们通常按数值顺序进行分组整理、汇总,计算出各组数据的频数或频率,分析方法主要是对观察到的数据做各种统计运算处理,进行各种统计描述或统计推断等。

2. 观测数据与实验数据

根据获得数据的方法不同,可以将数据分为观测数据和实验数据。观测数据可能是全面数据也可能是样本数据;实验数据一般为样本数据。

(1) 观测数据

观测数据是对客观现象进行实地观测所取得的数据,在数据取得的过程中一般没有人为的控制和条件约束。在社会经济问题研究中,观测是取得数据最主要的方法。很多社会经济问题不适合应用实验的方法,只能通过对客观实际做调查得到数据,用各种调查方法得到的数

据属于观测数据。

(2) 实验数据

实验数据一般是在科学实验环境下取得的数据。实验环境是受到严格控制的,数据的产生一定是某一约束条件下的结果。在自然科学研究中所用的数据多为实验数据,在市场研究中,也经常采用实验法来取得实验数据。例如,某服装销售公司为了验证改变服装产品的包装是否会得到消费者的认同而采用的实验法所得到的实验数据,如表 0-1 所示。

表 0-1 控制组与试验组销售额资料

	网点 1	网点 2	网点 3	网点 4	网点 5	网点 6
控制组销售额(万元)	23	21	24			
试验组销售额(万元)				35	32	34

三、参数与统计量

(一) 参数

参数是用来描述总体数量特征的概括性值,通常用大写字母表示,如总体均值(\bar{X})、总体标准差(σ)、总体比例(P)等。

(二) 统计量

统计量是用来描述样本数量特征的概括性值。通常用小写字母表示,如样本均值(\bar{x})、样本标准差(s)、样本比例(p)等。

参数与统计量是统计推断中的重要概念。在进行统计推断时,总体数据通常是不完整的,所以参数是一个未知的也是唯一的常数,但样本是经过抽样所确定下来的,所以统计量总是可以计算出来的,而且是一个随机变量,其数值随着构成样本的个体不同而不同。统计推断的目的就是要根据已知的样本统计量去估计未知的总体参数。例如,根据样本家庭人均消费水平去推断总体家庭人均消费水平,根据样本每百户家庭汽车拥有量去推断总体每百户家庭汽车拥有量等。

技能训练 SPSS 软件安装及相关操作

一、SPSS 软件简介

SPSS 原意是“Statistical Package for Social Science”,即社会科学统计分析软件包。20 世纪 60 年代末,美国斯坦福大学的三位研究生研制开发了最早的统计分析软件 SPSS,并于 1979 年在芝加哥组建成立了 SPSS 公司。

SPSS 是当今世界上公认和流行的综合统计分析软件包。国际学术界规定,在国际学术交流中,凡是用 SPSS 软件完成的计算和统计分析,可以不必说明算法。它界面友好、功能强大、易学、易用,包含了几乎全部尖端的统计分析方法,具备完善的数据定义、操作管理和开放的数据接口以及灵活而美观的统计图表制作。SPSS 赢得了各个领域中广大数据分析人员的喜爱,并得到了广泛的应用。



二、SPSS 软件运行方式

SPSS 主要有三种运行方式：

1. 批处理方式

批处理方式是已编写好的程序存储为一个文件，然后在 SPSS 的 Production Facility 程序中打开并运行。

2. 完全窗口菜单运行方式

完全窗口菜单运行方式主要通过选择窗口菜单和对话框完成各种操作。用户无须学会编程。这种方式比批处理方式简单易懂。

3. 程序运行方式

程序运行方式是在语句(Syntax)窗口中直接运行编写好的程序或者在脚本(script)窗口中运行脚本程序的一种运行方式。它与批处理方式相同，都要求读者掌握 SPSS 的语句或脚本语言。

本书旨在为初学者提供入门试验教程，采用“完全窗口菜单运行方式”。

三、SPSS 的安装和启动

在启动 SPSS 软件之前，需要先在计算机上进行安装。其安装方法主要有两种：一是直接使用 SPSS 安装光盘进行安装；二是通过网络下载 SPSS 安装程序进行安装。本小节详细介绍使用第二种方法安装 SPSS 的步骤(以 IBM SPSS 19.0 为例)。

1. 打开计算机，找到已经下载到计算机上的 SPSS 安装程序。
2. 双击该程序，打开“安装向导”界面，如图 0-3 所示。

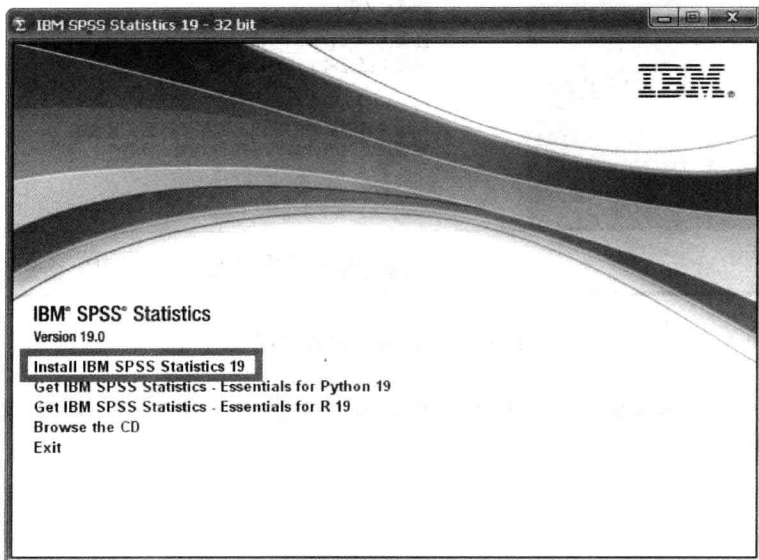


图 0-3 安装向导界面

3. 单击“下一步”，则弹出对话框；个人用户选择第一个“单个用户许可证”，如图 0-4 所示。