

免费提供

精品教学资料包

服务热线: 400-615-1233
www.huatengzy.com

B_{asic} Statistics

统计学基础

统计学基础



定价: 39.80元



西南财经大学出版社
Southwestern University of Finance & Economics Press



“十二五”高职高专教育规划教材

B_{asic} Statistics

统计学基础

主编 肖战峰 主审 李智沛



西南财经大学出版社
Southwestern University of Finance & Economics Press

“十二五”高职高专教育规划教材

统计学基础

主 编 肖战峰
副主编 蒋智霞 欧阳菲 蒋彦青
王 蕾 肖培耻
主 审 李智沛

西南财经大学出版社

图书在版编目(CIP)数据

统计学基础/肖战峰主编. —成都:西南财经大学出版社,2008.5
(2017.7 重印)

ISBN 978-7-81088-981-0

I. 统… II. ①肖… 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 050608 号

统计学基础

主编:肖战峰

责任编辑:李雪

装帧设计:蒋宏工作室

责任印制:封俊川

出版发行:	西南财经大学出版社(四川省成都市光华村街 55 号)
网 址:	http://www.bookcj.com
电子邮件:	bookcj@foxmail.com
邮政编码:	610074
印 刷:	大厂回族自治县聚鑫印刷有限责任公司
成品尺寸:	185mm×260mm
印 张:	17.25
字 数:	420 千字
版 次:	2017 年 7 月第 1 版第 7 次印刷
书 号:	ISBN 978-7-81088-981-0
定 价:	39.80 元

1. 如有印刷、装订等差错,可向本社营销部调换。
2. 版权所有,翻印必究。

编审委员会

顾 问 宋远方 中国人民大学

主 任 刘淑琴 山西财政税务专科学校

副 主 任 井 颖 山东商业职业技术学院

王曦东 大连职业技术学院

陈头喜 江西工业职业技术学院

梁建民 黄河水利职业技术学院

委 员(以姓氏笔画为序)

王卫国 王水清 王秀芳 王金玲 王晓东

王栓军 王雪琳 付秀彬 史海霞 申海波

石春玲 刘国防 刘国峰 张 萍 张松灿

张莉艳 李军峰 李香者 杨应杰 杨录强

姚进才 段迎君 郭 平 郭 鹏 郭现芳

高 静 黄爱玲 惠献波 蒋春艳

出版说明

高职高专教育作为我国高等教育的重要组成部分,承担着培养高素质技术、技能型人才的重任。近年来,在国家和社会的支持下,我国的高职高专教育取得了不小的成就,但随着我国经济的腾飞,高技能人才的缺乏越来越成为影响我国经济进一步快速健康发展的瓶颈。这一现状对于我国高职高专教育的改革和发展而言,既是挑战,更是机遇。

要加快高职高专教育改革的步伐,就必须对课程体系和教学模式等问题进行探索。在这个过程中,教材的建设与改革无疑起着至关重要的基础性作用,高质量的教材是培养高素质人才的保证。高职高专教材作为体现高职高专教育特色的知识载体和教学的基本工具,直接关系到高职高专教育能否为社会培养并输送符合要求的高技能人才。

为促进高职高专教育的发展,加强教材建设,教育部在《关于全面提高高等职业教育教学质量的若干意见》中,提出了“重点建设好3000种左右国家规划教材”的建议和要求,并对高职高专教材的修订提出了一定的标准。为了顺应当前我国高职高专教育的发展潮流,推动高职高专教材的建设,我们精心组织了一批具有丰富教学和科研经验的人员成立了编审委员会。

编审委员会依据教育部制定的《高职高专教育基础课程教学基本要求》和《高职高专教育专业人才培养目标及规格》,调研了百余所具有代表性的高等职业技术学院和高等专科学校,广泛而深入地了解高职高专的专业和课程设置,系统地研究了课程的体系结构,同时充分汲取各院校在探索培养应用型人才方面取得的成功经验,并在教材出版的各个环节设置专业的审定人员进行严格审查,从而确保了整套教材“突出行业需求,突出职业的核心能力”的特色。

本套教材的编写遵循以下原则:

(1)成立教材编审委员会,由编审委员会进行教材的规划与评审。

(2)按照人才培养方案以及教学大纲的需要,严格遵循高职高专院校各学科的专业规范,同时最大程度地体现高职高专教育的特点及时代发展的要求。因此,本套教材非常注重培养学生的实践技能,力避传统教材“全而深”的教学模式,将“教、学、做”有机地融为一体,在教给学生知识的同时,强化了对学生实际操作能力的培养。

(3)教材的定位更加强调“以就业为导向”,因此也更为科学。教育部对我国的高职高专教育提出了“以应用为目的,以必需、够用为度”的原则。根据这一原则,本套教材在编写过程中,力求从实际应用的需要出发,尽量减少枯燥、实用性不强的理论灌输,充分体现“以行业为导向,以能力为本,以学生为中心”的风格,从而使本套教材更具实用性和前瞻性,与就业市场结合也更为紧密。

(4)采用“以案例导入教学”的编写模式。本套教材力图突破陈旧的教育理念,在讲

解的过程中,援引大量鲜明实用的案例进行分析,紧密结合实际,以达到编写实训教材的目标。这些精心设计的案例不但可以方便教师授课,同时又可以启发学生思考,加快对学生实践能力的培养,改革人才的培养模式。

本套教材涵盖了公共基础课系列、财经管理系列、物流管理系列、电子商务系列、计算机系列、电子信息系列、机械系列、汽车系列和化学化工系列的主要课程。

我们希望有更多经验丰富的教师加入到我们的行列当中,编写出更多符合高职高专教学需要的高质量教材,为我国的高职高专教育作出积极的贡献。

编审委员会

序

随着我国经济的持续快速健康发展,各行各业对高技能专业型人才的需求量迅速增加,对人才素质的要求也越来越高。高职高专教育作为我国高等教育的重要组成部分,在加快培养高技能专业型人才方面发挥着重要的作用。

与国外相比,我国高职高专教育起步时间较短,这种状况与我国经济发展对人才大量需求的现状是很不协调的。因此,必须加快高职高专教育的发展步伐,提高应用型人才的培养水平。

根据教育部关于加强高职高专人才培养工作的意见,“课程和教学体系的改革是高职高专教育教材建设规划的重点和难点,要切实做好高职高专教育教材的建设规划,加强文字教材、实物教材、电子网络教材的建设和发行工作”。

近年来,在各级教育主管部门、学校和有关出版社的共同努力下,各地已出版了一批高职高专教育教材。但从整体上看,具有高职高专教育特色的教材极其匮乏,教材建设仍落后于高职高专教育的发展需要。因此,加强对教材工作的建设力度,是当下高职高专教育教学工作的重点之一。

在此背景下,我们组织了一批具有丰富理论知识和实践经验的专家、一线教师,编写了本系列专门面向高职高专的财经类教材。

本系列教材力求以就业为导向,注重对学生能力的培养,使学生能够学以致用。在内容介绍上,本系列教材将理论知识与实践特点相结合,使学生能够边学习,边吸收,边掌握;在结构安排上,本系列教材特意设置了“学习目标”、“本章小结”、“思考练习”和“案例分析”等板块,以方便学生自主学习。另外,教材还配有教学资料光盘和专业网站,以方便教师教学。

本系列教材一定能够很好地适应现代高职高专教育教学的需求,为社会培养出更多优秀的财经专业人才。

中国人民大学商学院

前 言

统计学是经济类专业的必修课,是学习其他专业课的基础,也是未来社会实践的重要应用工具。如何将统计学的教学内容和学生们的实际需要结合起来,这是每个统计学教学工作都面临的一个重要课题。

纵观当前的统计学教材,林林总总不下百十种,各从不同的学术角度、不同的专业层次来构建统计学教学体系。总体来讲,大家都力图突出两个特点:一是受“大统计”学术思想的影响,把“社会统计学派”和“数理统计学派”的方法融合在一个学科体系内;二是强调统计方法应用,把应用性作为教学的重点。但在教材中如何具体反映这两个特点,却存在着诸多问题。例如,许多教材并没有深刻理解“社会统计学派”和“数理统计学派”学术观点的真正含义,只是将两种方法简单拼凑在一起,这样势必造成前后内容上的不协调、不一致。而在统计方法的应用上,大多也仅仅停留在“见点论点,见面论面”的层面上,只是通过例题来讨论各种统计方法的应用过程,忽视了统计方法应用条件的介绍,这样势必导致一个结果:学生学了很多统计方法,却不知道在什么场合下用,面对数据时仍然束手无策。另外,有些教材立足于宏观经济现象的研究,所讨论的统计分析方法也只局限于宏观研究。例如,在“数据收集”时只介绍抽样调查、普查、统计报表等宏观统计中的几种常用调查方法,缺少企业经营管理等社会实践中常用调查方法的介绍,而大多数学生需要的恰恰正是未来实际工作中应用更广泛的研究方法。基于上述问题,本书力图体现如下特点:

(1) 以方法应用为主,以方法原理为辅,立足于学生进一步学习专业课和未来社会实践的实际需要,突出实用性。经济类非统计专业学生学习统计学课程的目的主要有两个,一是专业课学习的需要,二是社会实践中的直接应用。因此,在统计学教学中,教学内容的安排和教学方法的选择必须紧紧围绕这两个目的,强调统计分析方法的实际应用。具体说来,在介绍各种统计分析方法的时候,介绍了应用条件、统计分析的 Excel 操作、数据的解释,并以此作为教学的重点,使学生明白各种统计方法“什么时候用”、“怎么用”。

(2) 立足微观研究,兼顾宏观分析。经济类非统计专业人才培养的目标并不是未来的专业统计工作者,大部分学生在未来的经济生活中运用更多的是微观分析方法,他们需要学会将统计方法应用于对微观现象的研究。鉴于学生的这种实际需要,本书立足于微观分析方法的介绍,突出统计方法在微观研究中的应用。本书除在统计方法的论述上更注重实用性外,在引例、案例和例题的选材上力求接近日常经济生活。在各章最后还编写了“实践技能训练”的内容,目的就是提高学生们的实际动手能力、应用能力,使他们的学习更接近实践需要。

(3) 强调素质教育,以提高学生的统计素质和分析能力为目标,增强学生的归纳概括能力。统计学是一门对数据进行收集、整理、分析的方法论学科,属于归纳性科学。通过学习,不仅可以提高定量分析决策的意识和能力,还可以提高归纳概括能力。本书立足于

这一思想,在内容表述和方法应用上力图使学生明白归纳的过程和特点,引导他们学会合理归纳、正确概括,切实提高归纳分析能力。

本书由肖战峰副教授任主编,各章编写分工如下:第一章、第五章由欧阳菲编写,第二章、第九章由肖战峰编写,第三章、第四章由王蕾编写,第六章由蒋彦青编写,第七章由蒋智霞编写,第八章、第十章由肖培耻编写,肖战峰负责全书的策划、设计、审订、修改、总纂和定稿工作。

由于作者水平有限,加之时间仓促,书中难免存在不当之处,诚恳地希望广大读者提出宝贵意见,给予批评指正。

编 者

目 录

第一章 统计概述	1
本章要点	1
案例导入	2
第一节 统计学的产生和发展	3
第二节 统计学的研究对象和学科性质	7
第三节 统计学的研究方法	10
第四节 统计学的基本概念	13
本章案例	17
实践技能训练	19
思考练习	20
第二章 数据收集	22
本章要点	22
案例导入	23
第一节 统计数据的来源	23
第二节 调查设计	32
本章案例	37
实践技能训练	42
思考练习	42
第三章 数据整理与显示	45
本章要点	45
案例导入	46
第一节 数据的审核	48
第二节 统计分组及频数分布	50
第三节 统计表	59
第四节 统计图	61
实践技能训练	70
思考练习	71
第四章 综合指标	73
本章要点	73
案例导入	74

第一节 总量指标	76
第二节 相对指标	79
第三节 平均指标	86
第四节 标志变异指标	96
本章案例	103
实践技能训练	104
思考练习	104
第五章 时间数列	107
本章要点	107
案例导入	108
第一节 时间数列概述	108
第二节 时间数列的水平指标	111
第三节 时间数列的速度指标	120
第四节 时间数列的分解分析	125
本章案例	140
实践技能训练	143
思考练习	143
第六章 统计指数	148
本章要点	148
案例导入	149
第一节 统计指数的概念、种类和作用	149
第二节 综合指数	152
第三节 平均指数	157
第四节 指数体系与因素分析	161
第五节 几种常用的价格指数	166
本章案例	173
实践技能训练	174
思考练习	174
第七章 抽样与抽样分布	178
本章要点	178
案例导入	179
第一节 抽样方法	179
第二节 三种不同性质的分布	184
第三节 单个总体参数推断的抽样分布	185
本章案例	192
实践技能训练	195
思考练习	195

第八章 参数估计	197
本章要点	197
案例导入	198
第一节 参数估计的一般问题	199
第二节 单总体参数的区间估计	205
第三节 样本容量的确定	211
本章案例	213
实践技能训练	215
思考练习	215
第九章 假设检验	217
本章要点	217
案例导入	218
第一节 假设检验的一般问题	220
第二节 单总体参数的假设检验	227
实践技能训练	237
思考练习	238
第十章 相关分析与一元线性回归分析	240
本章要点	240
案例导入	241
第一节 相关分析和回归分析概述	241
第二节 相关分析	247
第三节 一元线性回归分析	252
本章案例	257
实践技能训练	259
思考练习	259
参考文献	262

第二章

数据收集

本章要点

- » 桌面调查是指收集加工数据的过程。桌面调查中要根据研究的目的,运用科学的方法有针对性地筛选出符合调查要求的数据来。可运用“杜威十进位分类法”对加工数据进行储存和管理。
- » 最常用的收集原始资料的调查方式是抽样调查,它是按照随机的原则抽取调查单位。
- » 调查方法是指从调查单位那里取得所需资料的具体办法,主要有访问法、观察法和实验法。
- » 调查方案不仅是整个数据收集工作的出发点,也是整个数据收集活动的行动指南。在方案设计中,必须要通盘考虑整个调查活动的全过程,切实保证各环节的协调一致。
- » 问卷调查是企业市场调查中常用的方法。调查问卷的设计务必要遵循全面系统、科学合理、缜密严谨、条理清晰的原则。



案例导入

外国公司怎样才能更加稳妥地进入中国市场？

“哈根达斯”是一个国际著名的冰淇淋品牌。在进入中国冷饮市场前曾经做了大量细致的市场调查工作。公司认为首先要确定进入中国市场的“登陆滩”。它们从居民的收入水平、消费习惯、对外来产品的接受能力等方面对中国几个大城市做了调查，结果表明上海是最理想的首选地。同时调查结果还表明，上海对中国其他地方的消费观念的影响作用也十分明显。

接下来应该着手研究的是通过什么渠道将产品推进上海的千家万户。调查结果显示上海市民选购冷饮的基本地方是：食品商场、大卖场、超市和便利店。但是对具体的品牌选择上有明显的“购买场所与品牌”的倾向。一些早期进入上海冷饮市场的国际品牌有自己的专卖店（与著名零售商业合作）。而中外合资的便利店中顾客都是较高收入者和追求新奇的年轻人，这部分人是国际品牌在上海的领先采用者。所以，哈根达斯选择在特定便利店与顾客“见面”的销售方式。最早可以选择那些开设在高中、大学校园附近的和高档住宅区邻近的便利店，稳定一段时间使顾客形成购买习惯。

产品包装设计的调查结果表明，哈根达斯若像可口可乐那样“中国化”可能会影响它的品牌形象，所以他们决定在包装设计上尽量维持原有特征。

最后的调查目标是“上海消费者会接受的价格水平”。“和路雪”也是外来者，它已经成为上海消费者最熟悉的食品商品品牌，价格已经做了几次调整，成为大众化冷饮，而哈根达斯要独树一帜，就必须差异化营销。由于哈根达斯的产品定位是高档冰淇淋，价格不能低，这样才能避免陷入原定目标顾客的“价廉无好货”的思维定势。

近两年多的实践证明哈根达斯制定的营销策略实现了既定目标。在年轻人中的普及率、忠诚度都达到并稳定在一定的水平。

收集统计数据的过程又称为统计调查，它是整个统计研究过程的起点，是统计分析的前提和基础。收集统计数据过程中，要着重考虑的问题是：收集哪些数据？从哪里能得到这些数据？如何获得这些数据？归结起来，就是调查内容、调查方式和调查方法的问题。

第一节 统计数据的来源

统计研究过程中所需要的数据主要有两种：一是从被调查者那里直接获得的资料，称为第一手数据或者原始数据；二是从其他研究者那里取得的研究成果，称为次级数据或者加工数据。收集加工数据的过程又称为桌面调查(Desk Research)。

一、桌面调查

一个统计研究活动，研究者不可能也不必要从被调查者那里获取所有的信息，合理运用别人的研究成果，对加工数据进行系统科学地收集并再加工往往可以达到事半功倍的效果。

桌面调查作为收集次级资料的一种调查方法，可以为企业经营管理和决策提供许多重要

的市场信息,在企业经营过程中起着非常重要的作用,历来被发达国家、著名企业所重视。例如,日本在 20 世纪 60 年代就是通过桌面调查取得了我国大庆油田的重要情报。他们从某画报上刊登的铁人王进喜的照片判断出大庆就在华北地区,并根据某报纸关于工人从火车站将设备人拉肩扛到钻井现场和王进喜在马家窑的言论的报道,推断出了大庆油田的确切位置。从王进喜出席人大会议判断出大庆油田出油了,之后又根据某报纸上刊出的一幅钻塔照片,推算出了油田的产油能力。在此基础上,日本人又估算出我国将在随后的几年中急需进口大量设备,并按照中国的需求特点设计了相关设备,从而在谈判中一举击败了欧美各国的竞争对手,使其设备顺利地打入中国市场。

相对于原始数据的收集,桌面调查有如下特点:

- (1) 桌面调查所收集的是别人加工处理过的数据,而不是原始数据。如各种统计年鉴、行业分析报告等。
- (2) 桌面调查以收集文献性资料为主。
- (3) 桌面调查所收集的数据包括静态数据和动态数据,尤其偏重于从动态角度收集各种反映总体变化的历史数据。

桌面调查中,应根据研究的目的有针对性地收集加工数据。常见的加工方法主要有:

- (1) 各级统计部门和各级各类政府主管部门公布的相关资料。这包括,国家统计局和各级地方统计部门定期发布的统计公报、定期出版的统计年鉴和其他相关统计资料等。此外,财政、工商、税务、银行等各主管部门和职能部门,也都建立有相应的统计机构、研究机构或调查机构,会定期或不定期地公布有关政策、法规、价格、行业动态等信息,这些信息也是桌面调查中重要的收集内容(表 2-1)。

表 2-1 各级统计部门和政府主管部门网址

各级统计部门和政府主管部门	网 址	数据内容
国家统计局	http://www.stats.gov.cn	统计年鉴、统计月报等
国务院发展研究中心	http://www.drcnet.com.cn	宏观经济、货币金融等
中国经济信息网	http://www.cei.gov.cn	宏观经济信息等
华通数据中心	http://data.acmr.com.cn	宏观经济数据等
中国决策信息网	http://www.china-policy.com	决策知识及案例
三农数据网	http://www.sannong.gov.cn	三农信息等
美国人口普查局	http://www.census.gov	人口、家庭等
美国预算编制办公室	http://www.whitehouse.gov/omb	财政、债券等

- (2) 各种经济信息中心、专业信息咨询机构、专业调查机构、各行业协会和联合会提供的市场信息和有关行业情报。

- (3) 国内外有关的书籍、报纸、杂志所提供的文献资料,包括各种统计资料、广告资料、市场行情和各种预测资料等。

- (4) 有关生产和经营机构提供的商品目录、广告说明书、专利资料及商品价目表等。

- (5) 各种国际组织、外国使馆、商会所提供的国际市场信息。

- (6) 电视台、电台提供的相关市场信息资料,如有关市场信息、经济动态、行业发展等方面的栏目、专题节目及广告等。

- (7) 国内外各种博览会、展销会、交易会、订货会等促销会议以及专业性、学术性会议上所发放的文件和材料。

- (8) 互联网与市场信息网络提供的各种信息。

- (9) 企业内部的业务资料、财务资料等方面的信息。

(10) 其他途径获得的信息等。

桌面调查的方法有很多,实践中,应首先弄清楚要收集哪方面的资料,然后针对这些资料的特点运用最合理的手段去收集所需信息。比如要收集政府在畜牧业方面的相关政策,就可以登陆政府网站直接查询出政府历年来所出台的各种相关政策;或者到政府对口主管部门直接查询相关信息。

桌面调查中务必要注意文案资料的储存、积累和管理。储存管理的方式主要有档案式管理和信息化管理两种。档案式管理是指通过建立档案来管理信息资料。这种方式的缺点是工作量大,反应速度慢,效率低下等。信息化管理又称电脑化管理,是指借助于现代化信息技术来管理信息资料。这种方式非常便于文案资料的储存、查找、归类、排序、分析等再加工的需要,不仅可以大大节省存储的时间和空间,而且可以大幅度提高工作效率和精确度。

无论采取档案式管理还是信息化管理都涉及到资料的规划和分类问题。美国图书馆专家麦尔威·杜威所创的“杜威十进位分类法(Dewey Decimal Classification,DDC)”是一种很好的资料分类方法,可为企业文案资料的储存和管理所借鉴。具体方法见表 2-2。

表 2-2 杜威十进位分类法

分 类	项 目
10 账册类	企业内部财务报表、专案计划、年度预算均归属于账册类,尤其是对经营指标、销售统计月报、费用明细表、库存明细表、营业报告、产品计划表、人员分配销售计划表、市场计划书等企业内部行销财务性资料的分析与运用,有助于行销问题的评估与衡量。
20 剪报类	新闻及杂志等剪报资料具有时效性、相关性与连续性。处理方式除了自行分类外,还得注明来源、日期,并编制索引,以便随时查阅。如某公司以五进位数将剪报资料进行分类,其中第一位数代表基本类,第二位数代表大类,第三位数代表中类,第四位数代表小类,第五位数代表细分类。
30 期刊类	各相关的资料皆属该类。期刊类可细分为:31 统计年鉴、32 一般图书、33 政府刊物、34 民间定期刊物、35 统计资料、36 竞争对象刊物、37 证券交易所报告、38 学校团体研究报告、39 其他相关市场研究报告。
40 专刊类	凡国内外专利资料、发明资料、开发资料或开发构想资料皆属此类。此类资料对于产品开发及销售具有很大的利用价值,尤其是对专利资料的分析,其结果往往是新产品研发构想的主要来源。专利资料大多由政府机构定期公布。该类可细分为:41 专利资料、42~49 发明及产品开发资料。
50 目录类	产品目录、产品说明书等资料属于此类。这些资料除了用于了解竞争对手的动向外,更是研究开发新产品所不可或缺的构想来源。该类可细分为:51 产品目录、52 产品说明书、53 技术说明书、54 营业说明书、55 企业说明书、56PR 促进表。
60 企业内外之视听觉类	该类资料包括:61 照片、62 地图、63 缩影影片、64 幻灯片、65 标本模型、66 录音带、67 影片、68 录影带、69 其他。
70 刊物索引类	各种关于期刊或杂志的发行索引均属此类。如美国国际开发处(AID)所出版的 Abstract Index,该索引刊物收录了 100 多个国家的“海外投资环境调查资料”,这些资料可供国际性投资所用。
80 外部机构调查资料类	企业外部各界所发布的调查资料均属此类。如政府机构的统计调查资料,金融机构的金融经济资料,大学、研究机构或民间机构等发布的关于市场或行为方面的资料等。可细分为:81 调查专案机构、82 公会、83 政府机构、84 学术机构、85 民间专业机构、86 金融机构。
90 演讲类	关于演讲会、讲习会或研讨会的资料,均可归入此类。在二手资料中,演讲类资料为最原始的第一手资料,听讲人员或受讲人员应于事后将听讲内容记载于规定格式的报表上。

桌面调查可以为企业经营决策提供必要的文案信息,企业决策者一定要具备科学的决策理念,建立完善的管理制度,加强次级资料的收集和管理工作。

二、原始数据的收集

原始数据的收集是指直接从被调查者那里取得的第一手资料。

(一) 调查的方式

收集原始数据的方式有许多,政府统计部门所进行的统计调查和市场研究机构所进行的市场调查在调查方式上稍有不同。政府统计部门所收集的资料主要是社会经济数据,常用的调查方式主要有抽样调查、普查、统计报表等。而市场调查所收集的主要是有关消费者需求、消费行为等市场信息,采用的调查方式主要是非全面调查,其中除了抽样调查外,还有非随机抽样调查如偶遇抽样、判断抽样、配额抽样、个案研究等。

1. 抽样调查

抽样调查(Sampling Survey),就是从总体中按照随机原则抽取一部分单位组成样本进行调查,并根据样本的信息来推断出总体特征的数据收集方法。

抽样调查是运用得最广泛的一种调查方式,是目前我国统计体制中收集资料最重要的方式,也是市场调查常用的调查方式。从效果上看,抽样调查具有省时省力、反应及时、适应面广、准确性高等优点。从方法的角度看,抽样调查有如下特点:

(1) 按照随机原则抽取调查单位。所谓随机原则,就是按照相等的概率来抽取调查单位,或者说,每个总体单位被抽中的概率是相等的。按照随机原则抽样,可以避免抽样中的主观干扰,排除了由于人们判断失误而带来的系统性偏差,使得样本更具客观代表性。随机抽样,是抽样调查的首要前提。

(2) 按照一定的置信度推断总体。抽样调查不仅仅是一种调查的方法,也是一种推断的方法,抽样调查的目的是根据样本的有关信息按照一定的置信度推断出总体的数量特征。

(3) 抽样调查的误差可以事先计算和控制。抽样推断是以样本的资料推断总体的数量特征,虽然存在着一定的抽样误差,但它与其他统计估算不同,抽样误差范围可以事先通过一定资料加以计算,并且能够采取一定的组织措施来控制这个误差范围,保证抽样推断的结果达到一定的可靠程度。

关于抽样调查的具体组织方式,将在第七章详细介绍。

2. 普查

普查(Census),就是为特定目的而专门组织的全面调查。这是一种了解基本国情、国力的重要调查方式。我国进行的普查主要有人口普查、工业普查、第三产业普查、经济单位普查等。表2-3显示了新中国成立后所进行的人口普查情况。

表 2-3 新中国成立后所进行的人口普查

次数 项目	第一次	第二次	第三次	第四次	第五次
时间	1953 年 7 月 1 日	1964 年 7 月 1 日	1982 年 7 月 1 日	1990 年 7 月 1 日	2000 年 11 月 1 日
总人口数(万人)	60 193	69 458	100 817	113 368	126 583

注:总人口数包括现役军人,不含港澳台居民。摘自国家统计局人口普查公报。

普查是对总体中的所有单位都进行调查,可以掌握社会经济现象的基本全貌,为相关政策或措施的制定提供依据。其特点是:

(1) 普查通常是一次性或周期性的。由于普查涉及面广,调查对象多,需要耗费大量的人力、物力、财力和时间,通常需要间隔较长的时间,一般每 10 年进行一次。

(2) 为避免调查数据的重复与遗漏,普查一般要规定出调查标准时间或调查期限,以保证调查结果的准确性。如我国第五次人口普查的时间是 2000 年 11 月 1 日 0 时,调查期限是从 2000 年 11 月 1 日至 2000 年 11 月 10 日。

(3) 普查的数据一般可信用度高,比较规范,因此它可以为抽样调查和其他调查提供数据依据。

(4) 由于普查需要耗费大量的人力、物力、财力和时间,因此普查主要被统计部门用于调查基本的国情、国力等特定现象。市场调查,对于小总体的研究(如某区域市场上某品牌轿车车主调查)也可采用普查的方法。

3. 统计报表

统计报表(Statistical Report Forms),是指按照国家有关法律法规的规定,自上而下地统一布置、自下而上地逐级提供基本统计数据的一种调查方式。

统计报表是收集统计数据的一种重要方式,在我国几十年的政府统计工作中,已形成了一套比较完备的统计报表制度,它已成为国家和各级地方政府部门统计数据的重要来源。

统计报表可作多种分类:按报送范围不同可分为全面统计报表和非全面统计报表;按报送周期不同可分为日报、旬报、月报、季报、年报等;按报送的单位不同可分为基层报表和综合报表;按报表的实施范围不同可分为国家统计报表、部门统计报表和地方统计报表。

4. 偶遇抽样

偶遇抽样又称方便抽样,是指研究者将在一定时间、一定环境里所遇到的或接触到的人选入样本的方法。“街头拦截式访问”就是一种偶遇抽样,即在街头、路口、车站等处拦住过往行人进行调查。偶遇抽样是一种非随机抽样,样本的抽取具有很大的偶然性,代表性差。

偶遇抽样的优点是方便省事,常被用于正式调查开始前的探索性研究,或对总体情况不甚了解时采用,正式调查中很少采用。

5. 判断抽样

判断抽样又称典型调查,是根据调查的目的和要求,在对研究对象进行全面分析的基础上,依靠研究人员的主观能力和经验,有意识地选择部分有代表性的单位进行调查。

判断抽样是一种非随机抽样,调查单位是根据研究人员的主观判断选择出来的,选择调查单位时注重的不是样本的数量而是调查单位的“典型性”,即选中的调查单位应该是能够充分地、集中地体现出调查对象总体某些方面共性特征的最有代表性的单位。判断抽样最常见的方式是“划类选典”,它是先对调查对象总体进行分类,然后从各类中选择少数具有代表性的典型单位进行深入细致地调查,以便找出事物的发展规律并以此对调查对象总体进行推断估计。比如,为了了解不同年龄的手机消费者的消费特点,可将消费者按年龄分为老、中、青、少,然后在每一年龄类型中判断选择出若干有代表性的典型单位,对其进行深入调查以寻找出该类消费者的本质特点,为营销策略的制定提供依据。又如,为了总结某行业中企业管理中成功的经验和失败的教训,可在该行业中选择经营成功的典型企业和经营失败的典型企业作为调查单位进行深入调查。

判断抽样的特点是:

(1) 它是一种深入细致的调查方式。它强调的是调查单位的典型代表性,因而调查范围小调查单位少,有利于深入研究事物内在的、本质的规律和特征。

(2) 它是一种比较灵活的调查方式,既可以注重于现象的量的分析,也可以从质的方面分析数量和数量关系形成的原因;可以是对某一问题纵向方面进行研究,深入少数单位了解问题的历史和现状,也可以是就某一问题从横向方面进行探讨,了解问题在不同条件下的不同表现。

(3) 调查单位是根据调查的目的和任务,在对调查总体进行全面分析的基础上,有意识地选择出来的。

(二) 调查的方法

调查方法是指从被调查者那里取得统计数据的具体方法。常见的调查方法有访问法、观察法和实验法。

1. 访问法

访问法就是调查者与被调查者直接或者间接地接触,以询问回答的方式取得数据的一种调查方法。具体包括面谈访问、座谈会、邮寄调查、电话调查、留置调查等。

(1) 面谈访问。面谈访问是调查员与被调查者通过面对面交谈的方式取得数据的调查方法。面谈访问有结构式访问和无结构式访问两种。结构式访问是指被调查者按照事先设计好的、有固定格式的标准化调查问卷或表格回答问题。结构式访问中的问卷和表格有自填式(由被调查者根据卷面自行填答)和访问式(由调查人员提问、被调查者作答的方式收集资料)两种。关于调查问卷和表格的设计将在第二节中介绍。结构式访问的优点是能够对调查过程加以控制,从而获得比较可靠的调查结果;调查结果便于归类、汇总、分析。非结构式访问是指事先不制作统一的问卷或表格,没有统一的提问顺序,调查人员只是给出一个题目或提纲,由调查人员和受访者自由交谈,以获取所需资料的调查方法。非结构式访问的优点是调查内容的自由度高,如果引导得当,可以获得大量有用信息。缺点是调查结果难以归类汇总,不便于统计分析。

(2) 座谈会。座谈会又称焦点访谈法,是指挑选一组有代表性的被调查者集中于调查现场,在主持人的组织下,就某个主体进行讨论、发表意见,从而获得所需资料的调查方法。座谈会需要注意的问题:

① 挑选主持人时要注意其驾驭会议的能力,既要鼓励和引导每个与会者充分发言,又要避免谈话中偏离主题。态度和善、大方、热情、灵活,对与会者的发言要敏感。

② 参加座谈会的人数和会议时间视座谈会内容和主持人能力而定,一般控制在 8~12 人,时间控制在 1.5~3 小时。

③ 要注意与会人员的代表性。

④ 做好座谈会的会前准备、会中记录和会后整理分析工作。

座谈会的优点是资料收集快、效率高,取得的资料广泛深入、结构灵活,如果主持人引导得当可收集到大量有用信息。缺陷是对主持人的素质能力要求高,涉及到隐私和敏感话题时难以取得有效资料,调查结果难以归类汇总。

(3) 电话调查。电话调查是以打电话的方式从被调查者那里取得资料的调查方法。电话调查是近些年随着电话在普通家庭中的普及而迅速发展起来的调查方法。其优点是操作方便,资料回收率高,便于随机抽样,调查的成本低。缺陷是对调查员的要求较高,调查内容难以深入全面。

电话调查分为传统电话调查和计算机辅助电话调查(Computer-Assisted Telephone Interviewing, CATI)。传统电话调查是借助于普通电话、普通的印刷问卷和普通的书写用笔,由经过培训的调查员按照规定的随机拨号方法和规定的提问方式(严格按照问卷中问题顺

序,以规定的语气和语速逐字逐题访问)来收集资料。CATI是在一个装备了相关设备的场所中进行。CATI的相关设备包括:一台起总控作用的计算机主机,若干台与主机相连的终端(又称CATI站),耳机式电话,鼠标,若干台起监视作用的计算机和配套的音像设备等等。CATI调查还需要安装相应的软件系统:自动随机拨号系统、问卷设计系统、自动访问管理系统、自动数据录入和简单统计系统等。每个访问员各自坐在一台终端前,当自动拨号系统将被访者的电话拨通后,访问员通过设备开始提问,并将被访者的答案键入计算机,计算机会自动显示恰当的下一道问题。CATI是目前西方发达国家十分流行的调查方法。

访问法还有其他的调查形式,如邮寄调查(将问卷通过邮政系统发放出去,被访者填答后邮寄回来)、留置调查(调查员将问卷留给被访者并告知填答要求,约定被访者填好后,调查员再登门当面回收)等。实际应用中,应充分考虑每种调查方法的优缺点,结合研究的目的和被访者的特点,选择出最合适的调查方法来。

2. 观察法

观察法是调查者在现场对被调查者的情况直接观察、记录,以取得信息资料的一种调查方法。

观察法可从不同的角度进行分类:

(1) 按照观察者的身份,可把观察法分为参与观察和非参与观察。所谓参与观察是指观察者直接加入到某个群体中,以内部成员的角色参与他们的各项活动,在共同生活中进行观察、收集资料。非参与观察是指观察者以旁观者的身份,置身于调查群体之外进行的观察。

(2) 按照标准化程度,可将观察法分为结构式观察和无结构式观察。结构式观察是事先制定好观察计划并严格按照规定的内容和程序实施的观察。其优点是标准化程度高,便于操作实施,也便于归类分析。无结构式观察是指对观察的内容、程序事先不作严格规定,依现场的具体情况随机决定观察。无结构式观察的优点是灵活机动,信息量大,如果观察者的能力强素质高,可探寻出现象内在的属性。缺陷是对观察员的素质和能力要求高,所得资料不系统、不规范,难以归类、分析。

(3) 按照观察的方式,可将观察法分为直接观察和间接观察。直接观察是指观察者直接对被观察者的活动进行观察。间接观察是指通过对自然物品、社会环境、行为痕迹等进行观察,以便间接反映被调查者的状况和特征。例如,通过对某个城市市容卫生的观察,就能从侧面了解该市居民的精神面貌。

“神秘购物法”(Mystery Shopping Studies)是观察法在实际经营管理中的一种具体应用。具体做法是:由对被调查行业有较深入了解的调查员,以普通顾客的身份亲历被调查企业的服务及产品,在真实的消费环境中以专业的视角感知企业与顾客接触的每一个真实时刻,并将其消费经历、感受、评价等以《顾客经历报告》的形式反馈给委托人。由于被检查或需要被评定的对象无法确认“神秘”顾客,较之领导定期或不定期的检查,能够更真实、客观并系统地反映出目标对象的真实状况。

某大学曾经受某服装研究部门的委托,运用“神秘购物法”对“××”专卖店的营业员进行了结构式观察。观察的内容主要包括:营业员的礼貌和笑容、营业员的推销技巧、店铺和货品的整洁程度、管理人员的态度等。研究者事先设计了20个项目(见表2-4)及相应的评分标准,制作出了暗访操作流程(见图2-1),由经过培训的大学生担任观察员进行调查。^①

^① 范伟达. 市场调查教程[M]. 上海:复旦大学出版社,2002:210~214.

表 2-4 暗访项目及评分标准

项 目	评 分	标 准
1	优	有营业员立即面对顾客打招呼
	良	有营业员销售后才面对顾客打招呼
	中	有营业员打招呼,但不面对顾客
	差	不打招呼
2	优	衣着统一,佩戴胸卡,发饰整洁,化妆自然
	良	四缺一
	中	四缺二
	差	四缺三或以上
3	优	全体店员积极服务或随时准备服务顾客
	良	大部分店员积极服务或随时准备服务顾客
	中	店内有依靠现象
	差	店内有聊天或干私事现象
4	优	礼貌用语,面带微笑,当顾客讲普通话时,营业员也讲普通话
	良	三缺一
	中	三缺二
	差	全部没有
.....		
16	优	店内收银台附近,有标牌,且很整洁
	良	店内收银台附近,有标牌,且很不整洁
	中	店内收银台附近,有标牌,且很脏
	差	无标牌
17	优	店内货架、橱窗、门面招牌、地面整洁
	良	一项欠缺
	中	二项欠缺
	差	三项或四项欠缺,或有一项严重损害商店形象
18	优	货品全部摆放有条不紊,分门别类,货架不空置,货品及模特无污渍、无损坏
	良	有一个货架(货品、模特)未达到要求
	中	有两个货架(货品、模特)未达到要求
	差	货品乱放一气,或三个以上货品及模特有污渍、有损坏
19	优	试衣室整洁、门锁安全、设施齐全
	良	三缺一
	中	三缺二
	差	全部没有
20	优	灯光充足、音响适中、温度适宜、走道畅通(无杂物堆放)
	良	一项有欠缺
	中	二项有欠缺
	差	三项或四项有欠缺,或有一项严重不足

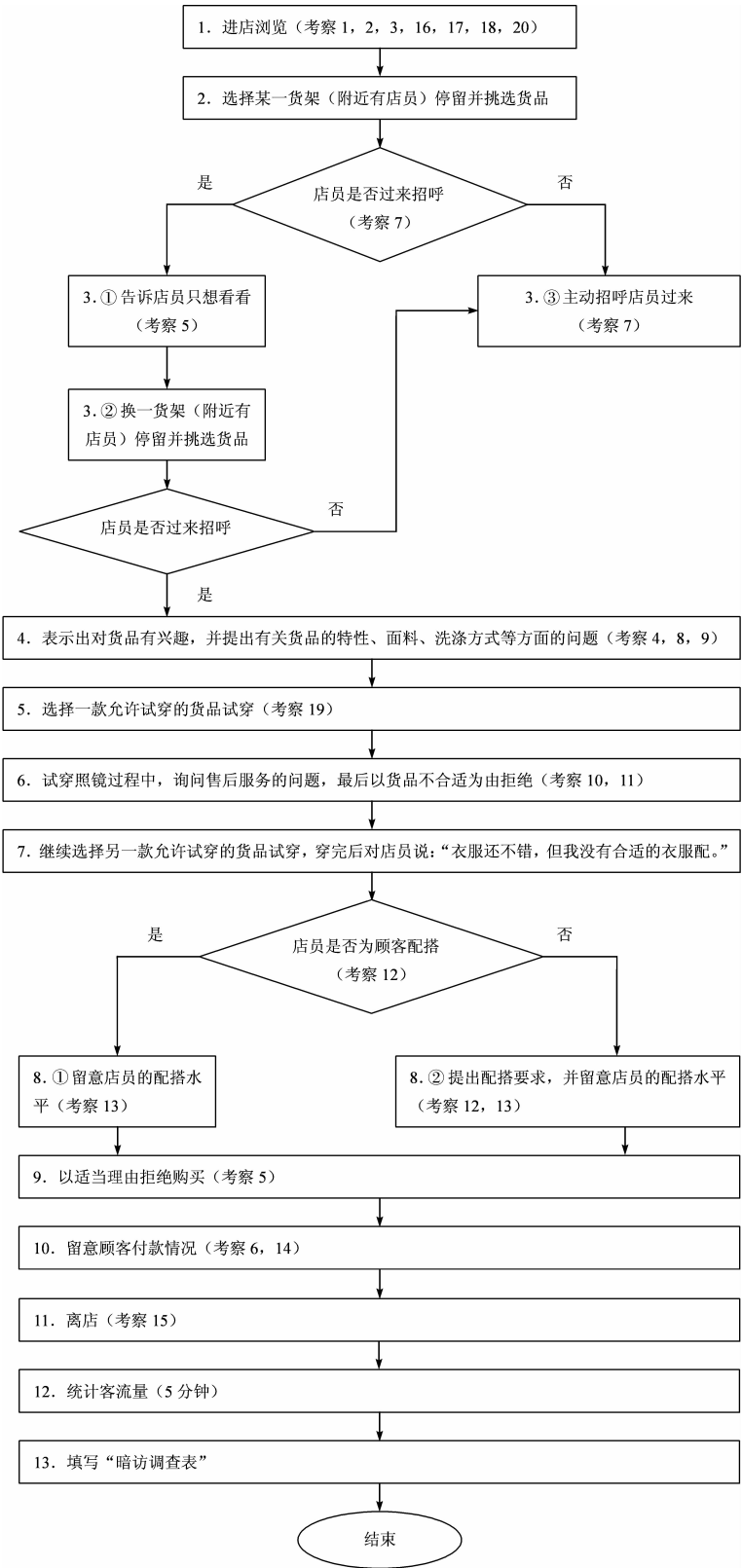


图 2-1 “暗访”操作流程

观察法的优点是灵活方便,资料客观真实,直接记录了调查的事实和被调查者在现场的行为,调查结果更接近实际。缺陷是对观察员的素质能力要求较高。另外,在应用观察法时,要对观察对象进行长时间观察,切忌“走马观花”。

3. 实验法

实验法是根据自然科学中科学实验的原理,从影响调查对象的若干因素中选出一个或几个因素作为实验因素,在其余因素不变化的条件下,了解实验因素的变化对调查对象的影响程度。

例如,一位日本咖啡店的老板发觉不同的颜色能使人产生不同的感觉,于是他就做了一个实验。他请来 30 位试验者,请他们每人喝四杯相同的咖啡,但这四个装咖啡的杯子的颜色是不同的,有红色、黄色、青色和咖啡色。然后询问每个试验者:“哪种杯子的咖啡浓度正好?”大家异口同声地回答:“青色杯子的咖啡太淡”,“黄色杯子的咖啡正好”,“红色杯子的咖啡太浓”,还有一部分人认为:“咖啡色杯子的咖啡太浓”。于是,咖啡店就改用了黄色的杯子,减少了原料,降低了成本。这就是实验法的一种应用。

实验法有多种,如事前事后对比实验,控制组同实验组对比实验,有控制组的事前事后对比实验,随机对比实验等。

实验法的优点是,通过实验获得的数据比较真实客观,调查结果具有较强的说服力,可以探索现象之间是否存在相关关系。缺陷是调查时间长、费用高,由于影响调查对象的因素较为复杂,因此对实验因素的控制比较困难。

第二节 调查设计

一、调查方案设计

统计调查是一项复杂而又细致的工作,会涉及到许多环节,调查参与者众多。为了在调查过程中统一认识,统一内容,统一方法,统一步调,顺利完成调查任务,就必须在调查前设计一个统一的调查方案,以保证统计调查工作有计划、有组织地进行。

一个完整的调查方案应包括以下内容:

1. 明确调查目的和调查任务

明确统计调查的目的和任务是制订调查方案的首要问题。对任何社会经济现象的研究,都可以根据不同的目的、不同的任务从不同的角度去收集资料。如消费者调查中,可以从消费需求的角度进行调查,也可以从消费行为或消费心理等角度进行调查,这就需要明确本次调查的目的是什么。调查的目的和任务不同,调查的内容和范围也就不同。目的不明任务不清,就无法确定向谁调查,调查什么,以及用什么方式方法进行调查。这就会使调查工作带有很大的盲目性,调查到的资料可能是不需要的,需要了解的情况得不到充分的反映,这样不仅会造成人力、物力、财力的大量浪费,而且还会延误工作。

例如,我国进行了五次人口普查,每次的调查目的都不一样。因此,调查项目数也不相同。第一次,只设四个项目,第三次,设有十三个项目,第四次设有十五个项目,到了第五次,就增加到了四十九项。总之,列入计划的调查项目是依据调查目的而确定的。

2. 确定调查对象和调查单位

调查的目的和任务明确以后,就要确定调查对象和调查单位。调查对象就是需要调查的社会现象的总体,它是由性质上相同的许多调查单位所组成。调查单位就是构成社会现象总体的个体,是调查项目的具体承担者,也就是在调查对象中所要调查的具体单位。例如:调查目的是收集某市小学状况的资料,那么,该市所有小学就是调查对象,而全市的每一所小学就是调查单位。又如,调查的目的是收集国有工业企业的设备资料,则调查对象是一切国有工业企业的所有设备,而国有工业企业的每一台设备则是调查单位。

确定调查对象和调查单位时要注意以下问题:

第一,由于经济现象具有复杂多变的特点,因此,在许多情况下调查对象是比较复杂的,必须根据研究目的严格规定调查对象的含义,指出它与其他有关现象的区别,以免造成调查中由于界限不清而发生的差错。如“×市城市居民家用电脑消费观念调查”,以×市城市居民为调查对象,就要明确“城市居民”的含义,划清“城市居民”和“非城市居民”概念的界限。

第二,调查单位的确定取决于调查目的和调查对象,调查目的和对象变化了,调查单位也要随之改变。例如,要进行“×市城市居民家庭家用电脑消费现状调查”,调查单位是×市每一户城市居民家庭,而不再是每一个城市居民了。

第三,不同的调查方式会产生不同的调查单位。如果采用普查方式,则总体内所有单位都是调查单位;如果采取抽样调查方式,则抽取出的样本单位是调查单位。

3. 确定调查项目

调查项目是指对调查单位所要调查的主要内容。在进行调查之前,必须根据调查的目的,明确规定出调查项目。一般说来,确定调查项目应注意以下几点:

第一,确定调查项目,要有取得资料的可能性。凡是被确定为调查内容的项目,必须能够取得确切的资料。否则,即使需要,但没有条件取得确切资料的项目,也不该列入。同时,对每一个调查项目都应该有确切的含义和统一的解释,以免因为调查人员理解不同而致使调查结果不一致。

第二,被确定为调查内容的每个项目间,应该彼此衔接,以便对现象的相互联系从整体上了解,也便于有关项目相互核对,提高调查资料的质量。

第三,调查项目之间时间上要有可比性,即本次调查项目和过去同类调查项目之间要互相衔接,以便进行动态对比。

第四,能确定的项目必须与调查目的有关。只登记与问题有关的标志,不应包括可有可无备而不用的标志。

调查项目确定后,就可以按照调查目的和要求,将调查项目制作成调查问卷或调查表。关于调查问卷的制作,将在本节第二部分“调查问卷的设计”中具体介绍。

4. 确定出样本容量、调查方式和调查方法

在调查方案中,要从调查目的出发,结合所研究现象的特点,确定出样本容量(调查单位的数量)、调查的方式和方法来。

设计调查方案务必要注意的是,调查过程中各环节之间必须要协调一致、相互支持。比如,确定样本容量时,就需要对调查的方式、调查的方法、调查的时间和费用、调查结果的准确性和精确性要求等诸多问题进行通盘考虑。如果采取随机抽样调查,就要根据调查的置信度要求和允许误差要求计算出必要样本容量(具体见第八章),并结合调查目的和调查成本以及

调查对象的特点,制定出具体的抽样方案来。如果是非随机抽样,则要根据研究者的经验判断和主观能力,结合调查对象的特点和调查方法,设计出详细的调查方式来。

5. 确定调查时间和调查工作期限

调查时间是指调查资料所属的时间。如果所要调查的是时期现象,就要明确规定该现象是从何年何月何日起到何年何月何日止。如果所调查的是时点现象,就要明确规定统一的标准调查时点。

为了保证调查工作的及时性和有序性,还要规定出调查工作期限。调查工作期限是指进行调查工作的时限,包括从方案设计、收集资料到提交调查报告的整个工作进度。规定调查期限要考虑调查项目的复杂性和调查资料的时效性。在可能的情况下,调查期限应尽可能缩短。

6. 确定调查经费预算

企业所进行的市场调查,要涉及到调查经费的筹措和使用,对费用进行估算也是调查方案的内容之一。调查费用的多少要视样本容量、调查范围以及调查内容的复杂程度而定,没有一个绝对的标准。进行费用估算时,一般要考虑以下几个方面:调查方案设计费,抽样费,问卷设计费,问卷印刷费,调查费(培训费、交通费、调查人员工资、礼品费等),数据录入费,数据统计分析费,调查报告费,办公费,其他费用等。

7. 制定调查的组织计划和管理措施

这是调查工作顺利实施的保证。调查的组织工作主要应包括以下内容:调查项目组的构建,调查员的组织和培训,调查前的准备工作(包括调查动员、资料查询、文件印刷等),调查经费的筹措,管理制度的制定等。

调查方案的设计是调查工作的出发点,也是整个调查活动的行动指南,务必要站在全局的角度对调查工作的各环节、各阶段通盘考虑,制定出合理的工作程序来。

二、调查问卷设计

问卷调查是收集原始数据的常见方法。它是将调查的内容转化成一系列具体的问题,并将这些问题写在一份卷宗内,称为调查问卷(Questionnaire),以书面的方式来收集所需数据的调查方法。问卷调查的效果取决于问卷的设计水平,设计出一份合格的问卷是提高调查结果有效性的重要前提。

(一) 问卷的结构

调查问卷有多种不同的形式,按照填答的方式不同,可分为自填式问卷(由被调查者自行填写)和访问式问卷(调查时由调查员提问,被调查者作答,再由调查员填写)。按照问卷的标准化程度,可分为结构式问卷(有固定格式的、有一定提问方式的要求和顺序要求的标准化调查问卷)和无结构式问卷(所问的问题没有严格的设计和格式要求)。这里主要介绍自填式结构化问卷的设计。

一份完整的调查问卷通常包括如下部分:

(1) 开头部分。开头部分一般包括问卷名称、问卷编号、调查组织名称、城市编号、访问员、问卷复核人、问卷编码员等内容。

(2) 开场白/介绍。开场白/介绍是对调查基本情况的说明,它包括问候语、调查的主题和组织、访问员的身份、访问的请求、对调查结果的保密承诺以及答谢的方式等。开场白/介

(3) 筛选/过滤部分。该部分是为了选择符合调研要求的受访者而设的。如,上述调查中,在询问正式问题之前先提出过滤性问题:

A. 是 B. 否(终止访问)

(4) 主体部分。这是调查的基本内容,是问卷中最重要的部分。

(5) 背景资料部分。该部分包括被访者的基本资料,如性别、年龄、民族、文化程度、婚姻状况、职业、家庭人数、收入、单位等。

(6) 结束语。结束语的任务是要告诉被访者访问完毕,表示对被访者的谢意。语言应自然、有礼貌。如“访问到此结束,非常感谢您的参与,这里有一份小礼物送给您,请签收。谢谢您,再见”。

（二）问卷的设计程序

问卷设计应遵循科学的程序,确保其可信度和有效度。一般来讲,问卷设计应按照以下程序:

(1) 准备阶段。该阶段是根据研究的需要,确定调查主题的范围和项目,将所需资料一一列出,分析哪些是主要资料,哪些是次要资料,哪些资料需要通过问卷来收集、向谁收集等。同时还要分析调查对象的各种特征,并以此作为拟定问卷的基础。在此阶段,应充分征求各相关人员的意见,力求使问卷切实可行,能够满足研究的需要。

(2) 初步设计。在准备工作的基础上,设计出初稿。主要是确定问卷结构,拟定并编排问题顺序等。

(3) 试答和修改。初步设计出来的问卷往往存在许多问题,需要在小范围内进行试验性调查,以便弄清楚问题所在。发现问题后,应对问卷做必要修改,使问卷更加完善、合理。

(4) 复印。复印就是将定稿后的问卷印刷出来,制作成正式问卷。

问卷设计的具体步骤见图 2-2。

（三）问题的提问方式

问卷中的问题有多种类型,总体上可分为两大类:封闭式问题和开放式问题。封闭式问题是在问题后设计了若干可能的答案,被访者只能从备选答案中选择出一个或几个来。如:

在购买罐头时,您认为哪些因素很重要?(多选)

A. 价格适中 B. 信赖的品牌 C. 到处都买得到

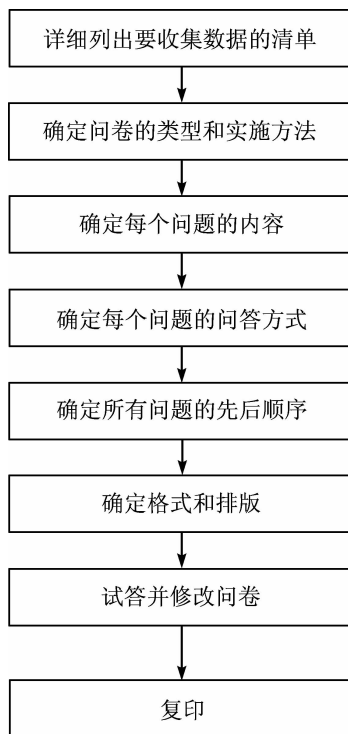


图 2-2 问卷设计的程序

- D. 口味好
- E. 包装精美
- F. 挑选余地大
- G. 含有特殊营养成分
- H. 著名企业的产品
- I. 开启方便
- J. 经常做广告
- K. 进口产品
- L. 不含防腐剂
- M. 不含人工色素
- N. 其他

封闭式问题的优点是,标准化程度高,回答方便,易于归类和分析,有利于提高问卷的回收率和有效率。缺点是被访者只能在限定的答案范围内选择,自由度小,可能无法反映其他各种有目的的、真实的想法。

开放式问题是指所提出的问题不列出答案,由被访者自由作答。如“您认为这种包装材料的饮料有哪些优缺点?”开放式问题的优点是被访者可以自由按自己的想法和方式回答问题或发表意见,不受限制,所得数据主动、具体、信息量大,有助于询问那些潜在答案很多、答案比较复杂、或者尚未弄清楚各种可能答案的问题。缺点是难以归类汇总,也不便于定量分析。

在提问问题时,务必注意提问的方式。具体来说,应注意以下几点:

- ① 避免提笼统抽象的问题。如“您对××超市的印象如何”这样的问题过于笼统,被访者难以回答。
- ② 避免使用含糊不清的概念。如问题“您是否经常去超市购物”中“经常”是个模糊不清的概念,可以改问“您上月去超市购物了几次”。
- ③ 问题要通俗易懂,尽量不要使用专业性术语。如问题“您对当前彩电销售中的低价倾销策略有何看法”中,“低价倾销”是个专业术语,一般的被调查者不会理解。
- ④ 避免引导性提问。如“消费者反映××牌电脑物美价廉,您是否有购买计划”暗示了被调查者选择肯定性的答案。
- ⑤ 切忌语言生硬、令人难堪。如:
您家至今不买电脑的原因是什么?
A. 买不起 B. 没有用 C. 不太懂 D. 软件少
该问题的提问方式显然令被调查者反感,自然难以获得配合了。
- ⑥ 避免问断定性问题。如对于问题“您一天抽多少支烟”,如果被调查者不抽烟,自然就难以回答了。
- ⑦ 避免多重提问。如“您的父母退休了吗”实际上问了两个问题:“您的父亲退休了吗”、“您的母亲退休了吗”,会使被调查者难以回答。
- ⑧ 避免问敏感性、禁忌性问题。如“您逃过几次税”等。

(四) 问题的排列顺序

在设计问卷时,要讲究问题的排列顺序,使问卷条理清晰。问题排序的原则是:

- ① 最初的提问应当是被访者容易回答且较为关心的内容。
- ② 提问的内容应从简单逐步转向复杂深入。
- ③ 对相关的内容应进行系统的整理,使被访者不断增加兴趣。
- ④ 作为调查核心的重要问题应排在前面。
- ⑤ 专业性强的具体细致问题应尽量放后面。
- ⑥ 敏感性问题也应该放在后面。
- ⑦ 封闭式问题放在前面,开放式问题放在后面。



本章案例

【案例 1】

西安市度假村市场调查方案^①

一、研究目的

本次市场调查的目的是为陕西某集团鲸鱼沟度假村开发项目提供市场分析基本资料,以帮助其在该项目的市场定位和市场监管策略方面作出正确抉择。需要调查研究的具体内容有以下几个方面:

(1) 西安市度假村的基本情况,包括:度假村的数量,规模,分布,高中低档次结构,客房出租率,经济效益率。

(2) 各层次度假村的典型调查,调查内容有:各度假村的设计接待能力,实际接待能力,服务设施,服务功能,不同时间段(周一至周四,周五至周日)的客源结构,客房出租率,出租价格及其变化等。

(3) 西安市消费者的收入状况和消费结构,度假消费状况,消费者对目前西安市度假村的评价及要求。

(4) 西安市集团度假消费状况。

二、研究方法

本次研究包括三方面的调查,第一是度假村调查;第二是西安市消费者调查;第三是西安市集团消费调查。度假村的调查范围为西安市区和长安县;消费者调查和集团消费调查的范围是西安市区内的企事业单位和城市居民户。

1. 度假村调查

这部分调查为了取得度假村的现状资料。调查对象主要是度假村,其次还要通过有关管理机构获取一些现成资料。调查方法采用面谈访问和案头调查法,并按照典型调查方法在不同规模、不同层次的度假村中分别选择两个调查对象(按照大、中、小规模分,共选 6 个),对其经营状况进行深入调查。

2. 消费者调查

由于去度假村休假的消费者主要是城市居民,故调查总体包括西安市六大区非农居民,具体范围是新城区、碑林区、莲湖区、雁塔区的小寨路街道办、灞桥区的纺织城街道办等地区居民。合计调查总体数为 46.3 万户。针对总体庞大的特点,我们决定采用随机抽样方法选择调查对象。根据有限总体所需样本经验数据表查得,在可靠性水平为 95%,允许误差范围为 $\pm 5\%$ 的情况下,当总体是 50 万个时,样本数为 384 个。由于本调查总体小于 50 万个,故样本数确定为 380 个。

根据本次调查内容要求,采用分群分层相结合的抽样设计方法。首先在以上调查总体中按随机方法抽取居委会(家委会),然后以家庭收入高低分层,并根据各层居民户占总体比例分配样本,再用等距抽样法在选中的居委会(家委会)中抽取入户调查对象。

^①张庚森. 市场营销调研[M]. 大连:东北财经大学出版社,2002. 37~38.

3. 集团消费调查

集团消费是度假村客流的主要来源。由于不同行业、不同规模、不同所有制的团体组织其集团消费内容和方式不同,因此,对集团消费调查采用相互交叉配额抽样方法,按照各行业、各种规模、各种团体组织的比例分配样本数。考虑到集团消费尽管有差异,但与消费者市场比较,模仿消费、攀比消费现象较为普遍,同一类型团体组织的消费相似性较强,故按照每一类型团体组织分配1~2个样本计,样本总数确定为40个,并采用面谈访问法进行调查。

三、调查时间

本次调查从4月30日开始,到5月30日结束,共计1个月,具体时间安排略。

四、费用预算

序 号	项 目	金 额(元)	备 注
1	方案设计费	4 000	
2	选样费	4 000	
3	问卷印刷费	1 170	
4	度假村调查费	3 000	
5	集团调查费	4 800	
6	消费者调查费	2 000	
7	小礼品	1 200	
8	交际费	1 000	
9	上机处理费	3 000	
10	调查报告费	5 000	
11	总计	29 170	

【案例 2】

××大学学生电脑消费倾向调查问卷

您好！我是××大学商学院 04 级统计学专业学生，我们需要就××大学学生电脑消费倾向这个话题征求您宝贵的意见，以便电脑厂商更好地为大学生提供服务。请您客观陈述您的观点，我们保证除了作为市场研究的资料外不作他用，并对您的回答严格保密，请您不必有任何顾虑！

感谢您在百忙之中抽出时间配合我们的工作,我们将在访问结束时向您提供一件小礼物以示感谢。谢谢!

××大学商学院 04 级统计学专业调查小组

一、被调查者背景资料

您的性别:男☐ 女☐ 院系:

年级： 大一 ☐ 大二 ☐ 大三 ☐ 大四 ☐ 研一 ☐ 研二 ☐

1. 您平均每月的生活费是()。

- a. 400 元以下 b. 401~500 元 c. 501~600 元
d. 601~700 元 e. 701~800 元 f. 800 元以上

二、市场占有率以及潜在市场分析

2. 你现在是否有电脑? 有 ☐ 无 ☐ (请跳至第 5 题)

3. 您的电脑是()。

- a. 组装机 b. 品牌台式 c. 品牌笔记本

如是品牌,品牌是()。

4. 您如果将要购买电脑(或想更换),会选择()。

- a. 兼容机 b. 品牌台式机 c. 笔记本

三、被调查者购买过程决策分析

下面是关于购买前的问题,无电脑者请直接跳至第 30 题。

5. 您购买电脑主要用途是()。(可多选,最多 3 个)

- a. 上网聊天 b. 看新闻 c. 游戏
d. 影视节目 e. 学习 f. E-mail
g. 电子商务 h. 其他

6. 您主要是通过()获得要购买的电脑的信息的?

- a. 专业电脑报刊 b. 普通杂志 c. 专业电脑网站
d. 大型门户网站 e. 电视 f. 咨询他人
g. 其他

7. 您会选择()购买电脑?

- a. 节假日活动期间 b. 宣传有特价时 c. 有特别需要时
d. 无所谓 e. 其他

如果您购买的是品牌机(包括手提),请继续,否则直接跳至第 12 题。

8. 您在购买台式品牌机时最看重()。

- a. 品牌 b. 外观 c. 性能(配置)
d. 价格 e. 质量 f. 售后服务

9. 您会选择在()购买品牌机。

- a. 电脑城 b. 家电百货超市 c. 品牌机专卖店
d. 网上电脑直销 e. 其他

10. 您愿意花()购买品牌机。

- a. 3 000 元以下 b. 3 000~4 000 元 c. 4 000~5 000 元
d. 5 000~6 000 元 e. 6 000~7 000 元 f. 7 000~8 000 元
g. 8 000 元以上

11. 从购买电脑至今出现问题的频率大概是()。

- a. 经常 b. 很少 c. 基本没有

12. 您电脑出现问题的种类()(可多选)。

- a. 黑屏 b. 蓝屏 c. 速度过慢
d. 死机 e. 外部设备(键盘鼠标等)

13. 出现了问题您是如何处理的?()

- a. 打售后电话 b. 找同学 c. 找学校里的修理公司
d. 自己处理 e. 其他

14. 您对您现在的电脑使用的满意度()。

- a. 非常满意 b. 满意 c. 一般 d. 差

15. 请您列出几条感觉特别不满意的地方_____。

四、售中及售后

16. 您对您购买电脑的环境的第一印象是()。

- a. 整洁,商品分类清楚 b. 一般 c. 混乱,无鲜明特色

17. 您对您购买过程中促销员的服务态度是()。

- a. 非常满意 b. 基本满意 c. 不满意 d. 非常不满意

18. 您对您购买过程中工作人员的办事效率是()。

- a. 非常满意 b. 基本满意 c. 不满意 d. 非常不满意

19. 您希望如何将电脑运回寝室(),实际是()。

- a. 公交车 b. 的士 c. 商场班车
d. 搬运公司 e. 其他

20. 您的电脑出问题寻求售后服务的一般响应时间()(即从告知客服到维护人员到达)。

- a. 半天 b. 一天 c. 两天
d. 更长 e. 不向客服求助

21. 您对售后服务水平的评价是()。

- a. 非常满意 b. 基本满意 c. 不满意 d. 非常不满意

22. 您对售后的建议是_____。

五、联想专项调查

非联想电脑用户请跳至第 30 题,联想品牌用户(包括手提)请继续。

23. 您购买联想电脑的时间是()。

- a. 一年前 b. 两年前 c. 三年前 d. 四年前

24. 您第一次知道联想这个品牌是通过()渠道。

- a. 听朋友说 b. 电视广告 c. 网络媒体
d. 专业报纸杂志 e. 其他

25. 您购买的渠道是()。

- a. 代理商 b. 专卖店 c. 家店商场
d. 大型超市 e. 网上直销 f. 其他

26. 如果下次您购买品牌电脑,您是否会购买联想品牌的电脑?()

- a. 会 b. 不会 c. 视情况而定

如果您是网上直接订购的,请继续回答第 27 题至第 29 题,否则直接跳至第 30 题。

27. 您选择的网上直接购买的原因是()。

- a. 价格低廉 b. 购买方便 c. 安全可靠 d. 其他

28. 您认为联想的网上订购系统操作是否简便快捷?()

- a. 很方便 b. 简便 c. 有点麻烦
d. 很难操作

29. 您对联想网上订购(网上商城)系统有什么建议(功能上的或界面设计上的等)?

30. 您是否经常接触联想电脑的宣传广告？()
- a. 经常 b. 比较经常 c. 不太经常 d. 不经常
31. 您认为联想品牌的优势是(),不足之处是()。
- a. 技术实力 b. 品牌优势 c. 售后服务
- d. 价格优势 e. 其他(请注明)
32. 同等价位下,您认为联想最缺的是()。
- a. 电脑整机的外观 b. 性能 c. 操作的舒适感
- d. 馈赠礼物 e. 不清楚 f. 其他
33. 如果让您设计联想电脑,您首先会加入()。(最多选三项,如选了“不清楚”则无需选其他)
- a. 更快的 CPU b. 更大容量的内存 c. 更好的独立显卡
- d. 更酷(漂亮)的外观 e. 更大的硬盘 f. 宽屏 LCD
- g. 全功能的刻录机 h. BOSS 级的音箱 i. 独立声效卡
- j. 不清楚 k. 其他
34. 同等配置下您会选择()。
- a. 一体机 b. 小型机 c. 传统的机型 d. 不清楚
- 下面问题请联想用户继续回答,其他人请直接跳至最后一题。
35. 在您购买和使用联想产品的过程中遇到过哪些问题？()
- a. 产品本身质量不合格 b. 广告与实物不符 c. 有价无货
- d. 没有遇到问题 e. 性能问题 f. 其他
36. 如遇到产品的质量问題,您是否通过相关售后服务或机构进行处理？
- 是□ 否□
37. 对处理结果您满意吗？()
- a. 满意 b. 基本满意 c. 不满意
38. 您对联想服务的建议是_____。

六、品牌形象

39. 针对下列每一种说法或评价,请选择您认为最合适的一个品牌。(单选)

(请根据您的 个人的见闻 或经历填写)	联想	华硕	惠普	神舟	方正	三喜	戴尔	清华 同方	宏基	七喜	TCL	其他 (可注 明)
	A	B	C	D	E	F	G	H	I	J	K	
1. 您最喜欢的品牌电脑												
2. 最值得信赖的品牌电脑												
3. 最能保证质量的品牌电脑												
4. 服务质量最好的品牌电脑												
5. 价格最合理的品牌电脑												
6. 技术最先进的品牌电脑												
7. 最适合大众消费的品牌电脑												
8. 最具发展潜力的品牌电脑												

问卷到此结束,感谢您花费了宝贵的时间来帮我们完成这份问卷。



实践技能训练

项目名称:

本校学生手机消费调查

实训内容:

1. 培养学生调查设计的能力。
2. 加强学生对调查方案设计、调查问卷设计的全面认识。
3. 培养学生系统思考问题、统揽调查全局的能力。

调查目的:针对本校学生手机消费的现状和需求进行调查,弄清楚当前大学生在手机性能、款式、价格、服务等方面的消费特点,为手机生产厂商的新产品研发和营销策略的制定提供决策依据。(注意调查的内容是手机而不是手机卡,注意手机生产商和运营商的区别。)

实训内容:针对上述目的,设计出调查方案和调查问卷。

实训目的:

1. 每个同学根据自己对调查目的的理解,自行设计调查方案和调查内容。
2. 为深化对调查项目的理解,可在老师的引导下先进行课堂讨论。
3. 调查方案设计中,要注意诸环节之间的协调一致、相互支持。
4. 调查内容要力求深入、全面,切忌空泛。

实训成绩:

每位同学的成绩由两部分组成:课堂讨论成绩(40%)和卷面成绩(60%)。



思考练习

一、填空题

1. 收集加工数据的过程称为_____调查。
2. 抽样调查是从总体中按照_____原则抽取一部分单位组成样本进行调查。
3. _____是指按照国家有关法律法规的规定,自上而下地统一布置、自下而上地逐级提供基本统计数据的一种调查方式。
4. 常见的调查方法有_____法、_____法和_____法。
5. 面谈访问有_____和_____两种。
6. “神秘购物法”(Mystery Shopping Studies)是_____在实际经营管理中的一种具体应用。
7. _____就是需要调查的社会现象的总体,它是由性质上相同的许多调查单位所组成。
8. 有固定格式的、有一定提问方式的要求和顺序要求的标准化调查问卷称为_____问卷。

二、单项选择题

- 按照随机原则抽取调查单位的调查方式是()。
A. 普查 B. 抽样调查 C. 判断抽样 D. 偶遇抽样
- 检查产品寿命应使用哪种调查方式()?
A. 普查 B. 抽样调查 C. 判断抽样 D. 统计报表
- 下面哪种调查方法属于观察法()?
A. 面谈访问 B. 座谈会 C. 神秘购物法 D. 随机对比实验
- 人口普查规定统一的标准时间是为了()。
A. 避免登记的重复与遗漏 B. 确定调查对象的范围
C. 确定调查的单位 D. 登记的方便
- 为了解不同职业群体的消费特征,将调查单位按职业分为公务员、教师、医生、工人等类型,根据研究者的经验和主观判断能力在每类型内选择出最有代表性的单位进行调查,这种调查方式属于()。
A. 随机抽样 B. 偶遇抽样 C. 重点调查 D. 判断抽样
- ××市大学生学习状况调查中,调查对象是()。
A. ××市所有高等学校 B. ××市所有大学生
C. ××市所有大学生的学习成绩 D. ××市所有高等学校内的班级
- “神秘购物法”是一种()。
A. 面谈访问法 B. 观察法
C. 留置调查法 D. 实验法
- “您通常买什么品牌的旅游鞋?”这种提问方式()。
A. 是正确的提问方式 B. 所问问题过于抽象、笼统
C. 使用了含糊不清的概念 D. 使用了断定性提问

三、多项选择题

- 桌面调查的特点是()。
A. 桌面调查所收集的是别人加工处理过的数据
B. 桌面调查所收集的是原始数据
C. 桌面调查以收集文献性资料为主
D. 桌面调查以收集物质性资料为主
E. 桌面调查偏重于从动态角度收集各种反映总体变化的历史数据。
- 下列属于非随机抽样的是()。
A. 抽样调查 B. 普查
C. 判断抽样 D. 偶遇抽样
E. 个案研究
- 下列属于访问法的是()。
A. 面谈访问 B. 座谈会
C. 电话调查 D. 观察法
E. 实验法
- 在工业企业设备调查中()。
A. 全体工业企业是调查对象 B. 工业企业的所有设备是调查对象

- C. 每个工业企业是调查单位
- D. 每台设备是调查单位
- E. 每个工业企业员工是调查单位
5. 封闭式问题的优点是()。
- A. 标准化程度高
- B. 回答方便
- C. 易于归类和分析
- D. 有利于提高问卷的回收率和有效率
- E. 被访者回答问题的自由度大

四、简答题

1. 调查方案包括的主要内容有哪些？
2. 抽样调查、判断抽样各有什么特点？
3. 问卷设计中,问题的提问方式应注意什么？
4. 问卷设计中,问题的排列顺序应遵循什么原则？

五、判断题

1. 所谓随机原则就是按照主观判断的方式抽取调查单位。()
2. 普查是为某特定项目专门组织的调查。()
3. 偶遇抽样是一种随机抽样。()
4. 开放式问题具有标准化程度高、调查结果便于归类分析等优点。()
5. 问卷中,问题的排列顺序要遵从“先易后难”的原则。()

第三章

数据整理与显示

本章要点

- » 数据的整理,就是根据统计研究的目的,对收集到的资料进行科学加工,使之系统化、条理化,的工作过程。
- » 数据的审核是统计整理的初始阶段,也是整个统计整理工作的基础。审核的主要内容包括数据的检查 and 数据的校订。
- » 统计分组,是指根据统计总体内在的特征和统计研究的任务需要,将统计总体按照一定的标志划分为若干组成部分的一种统计方法。频数分布,是指在统计分组的基础上,把总体全部单位按组归类整理,汇总出各组的总体单位数,并将其按分组顺序加以排列而形成的统计数列。不同类型的统计数据,其统计分组的方法也不相同。
- » 统计表和统计图是表达频数分布最常用的形式,EXCEL 所提供的图表功能使得数据的整理和显示工作得到了很大的简化。



案例导入

重庆市城市居民服务性消费支出现状

十六大以来伴随着重庆社会经济发展速度的加快,人们收入水平不断提高,重庆城市居民的服务性消费支出保持平稳增长。据国家统计局重庆调查总队的抽样调查资料显示,2002—2006年的五年间重庆市城市居民年人均服务性消费支出由1 883.88元增加到2 969.28元,净增1 085.4元,增幅达57.6%,年均增长9.5%,比商品消费支出增幅快2个百分点,成为推动居民消费支出持续增长的首要原因。服务性消费占消费支出的比重逐年增加,由2002年的29.6%上升到2006年的31.6%,增长2个百分点,保持平稳上升态势(见表3-1)。

表 3-1 2002—2006 年重庆城市居民服务性消费支出比重

年份 名称	2002	2003	2004	2005	2006
可支配收入(元)	7 238.07	8 093.67	9910.09	10 243.99	11 569.74
消费支出(元)	6 360.20	7 118.06	7 973.05	8 623.29	9 398.67
服务性消费支出(元)	1 883.88	2 019.74	2 443.77	2 689.82	2 969.28
服务性消费支出比重(%)	29.6	28.4	30.7	31.2	31.6

从服务性消费内部构成看,教育文化娱乐服务占的比重最大(见表3-2),2002年占到40%,2003年以后始终保持在35%以上,占整个服务性消费支出的三分之一强。居民用于教育文化娱乐的支出由2002年的792.24元/人增加到2006年的1 049元/人,增长32.4%,年均增长3.4%。在教育文化娱乐服务中,教育费用是主要部分,占教育文化娱乐服务的70%以上,占整个服务性消费的26%以上;这与近年来人们重视对子女的教育和培养,教育投入加大;成人更加注重自身的学习提高,购买各种书籍、学习软件,听讲座,参加培训班,丰富自身文化知识,不断提升个人素养和综合素质是分不开的。

表 3-2 2002—2006 年重庆城市居民各类服务性消费支出比重(%)

年份 名称	2002	2003	2004	2005	2006
全部服务性消费	100.0	100.0	100.0	100.0	100.0
1. 餐饮服务	12.1	12.5	12.7	13.2	14.9
其中:在外用餐占总计	12	12.4	12.6	12.9	14.7
2. 衣着加工服务	0.3	0.2	0.2	0.2	0.2
3. 家庭服务	1.5	1.7	1.1	1.3	1.9
其中:家政服务费占	61.4	56.6	50.2	67.0	67.8
4. 医疗服务	6	5.4	5.7	7.4	7.6
5. 交通通信服务	29.2	32.5	32.2	30.1	30.1

续表

名称 \ 年份	2002	2003	2004	2005	2006
6. 教育文化娱乐服务	42	37	35.6	36.2	35.3
其中:教育费用占	75.4	71.8	73.7	73.5	73.5
7. 居住服务	6.6	8.8	10.4	9.4	7.8
其中:居住服务费占	14.1	23	21.7	21.9	29.6
物业管理费占	53.4	54.4	68.0	65.2	74.3
8. 其他服务	2.3	1.9	2.1	2.2	2.2

交通通讯服务在服务性消费支出中居第二位。近年来随着重庆市交通条件的不断改善,人们出行的便利程度大幅提高,有效地激发了人们出行的热情,重庆市城市居民的交通费支出五年间增加了 63.46 元,增长 25.9%。移动电话、固定电话、小灵通等通讯设备以及互联网的广泛普及,在通讯资费整体下降的情况下,人均通讯服务消费支出保持年均 3.7% 的增速,2006 年人均通讯服务费达 572.13 元,比 2002 年增加 88.3%,平均每天消费 1.59 元/人,占交通通讯服务的 64%,占整个服务性消费支出的 19% 左右。交通通讯服务支出在服务性消费中的比重达 30% 以上。

餐饮消费仍是城市居民常规消费的主流。重庆人热情好客又好吃,重庆美食名扬天下。重庆人均在外用餐的消费占服务性消费的比重常年保持在 12% 以上,2006 年达到 14.7%,年均保持 3.7% 的增幅。

家庭服务需求快速增长。随着居民收入的不断增加、职业女性摆脱家务劳动的意愿不断增强,居民对家庭服务的需求不断上升,各类家政服务纷纷走进居民家庭生活。2006 年重庆市城市居民的人均家庭服务支出为 55.6 元,比 2002 年增加 28.24 元,增长 103.2%,以家政服务为主要消费方式的支出快速增长,2006 年比 2002 年增加 20.92 元/人,占家庭服务支出的 67.8%。家庭服务消费的增幅居各类服务性消费首位。

居住服务需求旺盛。近年来,重庆的城市建设不断加快,安居工程、廉租房工程的推广、各种中高档楼盘相继落成,在改变城市面貌、美化城市环境的同时,也改善了人们的住房条件,有效地带动了居住服务消费快速增长,2006 年人均居住服务支出比 2002 年增加 108.65 元,累计增长 87.6%,其中人均居住服务费支出增长 292.7%,年均增速达 4.5%。在居住服务费中物业管理费的增长最快,2006 年比 2002 年增长了 4.46 倍,人均物业管理费支出达 51.09 元,年均增长 4.8%。

医疗服务消费逐年上升。由于医疗费上涨和人口老龄趋势加快,医疗服务消费需求逐年上升,消费支出不断增加,2002—2006 年重庆城市人均医疗服务支出增加 112.69 元,增幅达 100.4%,仅次于家庭服务消费的增幅。2006 年医疗服务支出占服务性消费的比重达 7.6%。

统计数据的整理,就是根据统计研究的目的,对所收集到的资料进行科学加工,使之系统化、条理化的工作过程。数据整理通常是对统计调查所取得的原始数据进行整理,对某些已加工的综合数据进行再整理。

数据整理是统计工作过程的重要阶段,它是实现从个体单位标志值过渡到总体数量特征的必经阶段。由于统计调查所获得的数据多是分散的,它只能反映一些个别单位的特征,无法揭示被研究现象的本质和规律性,达不到统计研究的目的。这就需要对数据进行去粗取精、去伪存真、由此及彼、由表及里的加工改造。同时,数据整理又是统计分析的前提,不对数据进行整理,就无法进一步计算有关的统计分析指标,无法进行有关的统计分析。并且,即使收集的数据准确无误,而整理的方法不当,也难得出满意的分析结果。因此,数据的整理,是统计研究必不可少的中间环节,数据整理的质量直接影响着统计分析的效果。

随着电子计算机技术的应用,数据整理的概念和作用都大大地扩展了。数据整理已不是局限于满足单项统计研究目的而进行的综合汇总加工,而是将大量的原始数据和再加工数据进行有序系统化,建立统计数据库,以满足多方面、多层次的反复需要。这样不但大大提高了数据整理的计算汇总速度,而且大大提高了统计信息资源的开发利用程度。

第一节 数据的审核

数据的审核是统计整理的初始阶段,也是整个统计整理工作的基础。所谓数据的审核,就是根据统计研究的目的和事先的研究设计,检查所收集的统计数据是否符合研究的要求,并在此基础上进行必要的修正和补充。这里主要讨论调查问卷的数据审核,审核的主要内容

包括数据的检查和数据的校订。

一、数据的检查

数据的检查是指对所收集的数据的完整性和准确性进行检查。

数据检查的目的是甄别出符合研究要求的有效数据,剔除无效数据的干扰和影响,为进一步的统计整理打好基础,从而提高统计分析结果的准确性。针对问卷调查而言,数据检查的主要内容就是问卷的完整性和准确性。检查的主要步骤是:

(1) 事先制定出检查的规则,明确规定出哪些问卷可以接受、哪些问卷必须拒绝等,并对每一份问卷进行检查。

(2) 根据检查规则,将所有的问卷分为三个部分:可以接受的、明显要作废的、对于是否可以接受有疑问的。一般而言,下面情况的问卷是要作废的:

所回收的问卷是明显不完整的,如缺失了一页甚至更多内容;

问卷从整体上看是回答不完全的;

问卷的几个关键部分的内容回答不完全的;

问卷显示调查员或被调查者没有理解或遵循访问要求的,如没有按要求跳答等;

问卷中的答案机会没有什么变化的;

问卷是在事先规定的截止日期以后回收的;

问卷是由不符合要求的被访者回答的;

.....

如果问卷作废后,有效问卷数量达不到规定的样本容量,则需要补充调查,使有效样本量达到规定的数目,否则将无法达到准确性和精确度的满意要求。

(3) 对可以接受的问卷和有疑问的问卷做进一步的校订。

二、数据的校订

所谓数据的校订,是指根据研究目的和研究设计,对数据做进一步的补充和修正,以满足统计研究的要求。

针对调查问卷,数据校订的步骤是:

(1) 对每一份问卷,检查出不满意的答案(如果有的话)。如下答案属于不满意答案:

字迹模糊的;

回答不完全的;

不一致的;

模棱两可的;

跳答错误的;

.....

(2) 处理不满意答案。对于不满意的答案,通常有两种处理方法:按缺失值处理、整个问卷作废。

在下列情况中,不满意答案按缺失值处理:

有不满意答案的问卷数量很小;

整份问卷中,不满意的答案所占比例很小;

有不满意答案的变量(问题)不是关键变量。

对于缺失值,通常采取如下方法进行处理:用一个样本统计量的值去代替缺失值、从一个统计模型计算出来的值去代替缺失值、将这份问卷保留,但缺失的答案仅在相应的分析中做必要的排除。

在下列情况中,则要将包含不满意答案的整个问卷作废:

不满意的问卷比例很小(不超过 10%);

样本容量很大;

不满意的问卷和满意的问卷之间没有明显的差别;

每份不满意的问卷中,不满意答案的比例很大;

关键变量的答案是缺失的。

经过检查和校订后的问卷,被视作有效问卷,就可以做进一步的整理和分析了。

对于通过其他渠道取得的二手数据,除了对其完整性和准确性进行审核外,还应着重审核数据的适用性和时效性。二手数据可以来自多种渠道,有些数据可能是为特定目的通过专门调查所取得的,或者是已经按特定目的的需要做了加工整理。对于使用者来说,首先应弄清楚数据的来源、数据的口径以及有关的背景材料,以便确定这些数据是否符合分析研究的需要、是否需要重新加工整理等,不能盲目生搬硬套。此外,还要对数据的时效性进行审核,有些时效性较强的问题,如果所取得的数据过于滞后,就失去了研究的意义。一般来说,应尽可能使用最新的统计数据。数据经过审核后,确认适合实际需要,才有必要做进一步的加工整理。

对审核过程中发现的错误应尽可能予以纠正。调查结束后,当数据中发现的错误不能予以纠正,或者有些数据不符合调查的要求而又无法弥补时,就需要对数据进行筛选。数据筛选包括两方面内容:一是将某些不符合要求的数据或有明显错误的数据予以剔除;二是将符合某种特定条件的数据筛选出来,对不符合特定条件的数据予以剔除。

第二节 统计分组及频数分布

一、统计分组和频数分布

（一）统计分组和频数分布的概念

统计分组,是指根据统计总体内在的特征和统计研究的任务需要,将统计总体按照一定的标志划分为若干组成部分的一种统计方法。统计资料分组的目的,就在于把同质总体中的具有不同性质的单位分开,把性质相同的单位合并在一起,保持各组内统计资料的一致性和各组之间统计资料的差异性,以便进一步研究调查对象的数量表现与数量关系,进而正确认识调查对象的本质及其规律性。例如,在我国人口普查中,作为个体的每个人,在年龄、性别、民族、文化程度以及居住地等诸多调查标志上不完全相同。为反映我国人口总体内部的差异,就需要按照不同的标志对全国人口进行分组。如,按性别可分为男、女两组,按年龄、民族可划分为若干组,这就有助于对我国人口的性别、年龄、民族等各方面的结构及其比例关系的认识。

频数分布,是指在统计分组的基础上,把总体全部单位按组归类整理,汇总出各组的总体单位数,并将其按分组顺序加以排列而形成的统计数列。分布在各组中的总体单位数,叫做频数,亦称次数;频数与总频数的比值,叫做比率,亦称频率。表 3-3 反映的是某单位职工文化程度的频数分布状况。

表 3-3 某单位职工文化程度分布表

文化程度	频数(人)	比例(%)
高中	10	7.69
专科	56	43.08
本科	48	36.92
硕士研究生	16	12.31
合 计	130	100.00

频数分布是对统计分组结果的表达,是进一步统计分析的重要基础。

（二）统计分组的作用

统计分组的作用主要表现为:

一是区分总体类型,现象的类型是多种多样的,不同类型的现象存在本质差别,通过统计资料的分组就可以把不同类型的现象区别开来。如表 3-4 是我国 2006 年末人口的年龄分布状况。

表 3-4 我国 2006 年末人口的年龄分布

年 龄	人口数(万人)	比例(%)
0~14	25 961	19.8
15~59	90 586	68.9
60 以上	14 901	11.3

二是反映总体内部结构,通过分组,统计总体被划分为若干组成部分,计算各组成部分的总量在总体总量中所占的比重,即可反映总体结构特征与总体结构类型。见表 3-5 我国 2006 年国内生产总值的产业构成。

表 3-5 我国 2006 年国内生产总值的产业构成

产 业	增加值(亿元)	比重(%)
第一产业	24 700	11.8
第二产业	102 004	48.7
第三产业	82 703	39.5
合 计	209 407	100

三是可以分析总体在数量现象之间的依存关系,现象之间总是相互联系、相互依存、相互制约的,分组就是要在现象各种错综复杂的联系中,找出它们内在的联系和数量关系。具体做法,可将一个可变标志(自变量)作为分组标志,来观察另一个标志(因变量)相应的变动状况。如居民家庭收入与就业人数有密切的联系。通过分组就可以反映这两个标志之间相互联系的程度和方向。例如表 3-6 反映了某地销售额与流通费用之间的关系。

表 3-6 某地销售额与流通费用

按销售额分组(万元)	商店数(个)	每百元商品销售额中支付的流通费(元)
50 以下	10	21.2
50~100	20	20.1
100~200	30	19.2
200~300	25	18.5
300 以上	15	16.0

(三) 统计分组的原则

要保证分组的科学性,要遵循“穷尽原则”和“互斥原则”。“穷尽原则”是指全部分组必须容纳所有总体单位,即总体中的每个总体单位都必须有组的归属。如劳动者按文化程度分组,若只分为小学、中学毕业和大学毕业三组;那么,未上过小学的以及大学以上文化程度的劳动者就无组可归。这种分组未做到“穷尽”。互斥原则是指在特定的分组标志下,总体中的任何一个单位只能归属于某一组,不能同时归属于几个组。把鞋子分为男鞋、女鞋、童鞋三类,就不符合互斥原则,因为童鞋也有男鞋与女鞋之分。

二、定性数据的分组

(一) 定类数据的分组

由于定类数据只有类别的属性,其类别本身就是分组,因此,对于定类数据只需按照其各个

类别的含义进行汇总,计算出相应的频数和频率即可。表 3-7 是我国 2005 年客运量资料。

表 3-7 我国 2005 年客运量

分 类	客运量(万人)	比例(%)
铁路	115 583	6.26
公路	1 697 381	91.90
水运	20 227	1.10
民航	13 827	0.74
合 计	1 847 018	100

原始的定类数据的整理和分组,可借助于 Excel 来进行。下面举例说明操作方法。

【例 3-1】 某班 40 个同学分别来自湖北、湖南、安徽、山东。表 3-8 是这 40 个同学的原始数据,试借助于 Excel 将其整理成频数分布。

表 3-8 某班同学的籍贯

山东	湖南	山东	安徽
安徽	湖北	湖北	山东
安徽	湖北	湖北	湖北
湖南	山东	安徽	湖北
湖北	山东	湖北	湖北
山东	湖南	湖北	山东
山东	湖南	安徽	湖南
湖北	湖南	安徽	山东
山东	安徽	湖南	安徽
安徽	山东	湖北	山东

解:为方便整理,将学生籍贯合并为一列,表达在 Excel 工作表中的 A2:A41。并用数字代码来表示学生籍贯,设 1——湖北,2——湖南,3——安徽,4——山东,将各学生的籍贯代码填入 Excel 工作表中 B2:B41。Excel 现在把代码视为数值型数据。为建立频数分布和条形图,Excel 要求将 4 个代码单独作为一列,以作为“接收区域”,设将 4 个代码输入到工作表的 C2:C5。见表 3-9。

表 3-9 籍贯及代码的 Excel 工作表

	A	B	C	D	E	F
1	籍贯	代码	代码上限			
2	山东	4	1			
3	安徽	3	2			
4	安徽	3	3			
5	湖南	2	4			
6	湖北	1				
7	山东	4				
8	山东	4				
9	湖北	1				
10	山东	4				
11	安徽	3				
12	湖南	2				
13	湖北	1				
14	湖北	1				
15	山东	4				
16	山东	4				

Excel 自动对代码数据值小于或等于每个代码的数据进行计数,这样得到的合计数即频

数。具体步骤如下：

① 打开 Excel 工作表中“工具”下拉菜单中的“数据分析”选项，出现如图 3-1 所示对话框。

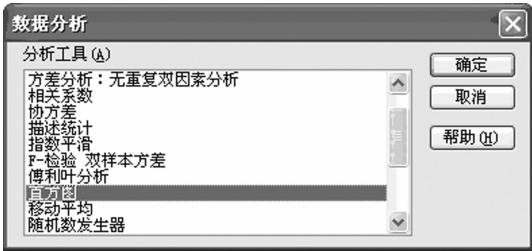


图 3-1 “数据分析”对话框

② 在“数据分析”对话框中选择“直方图”命令，并点击“确定”按钮，出现如图 3-2 对话框。



图 3-2 “直方图”对话框

③ 在该对话框中“输入区域”一栏填入数据区域 B2:B41；在“接收区域”一栏填入代码区域 C2:C5；在“输出区域”一栏填入结果输出的区域（本例为 D6）；其他选项根据需要选择。点击“确定”按钮，得结果，见图 3-3。

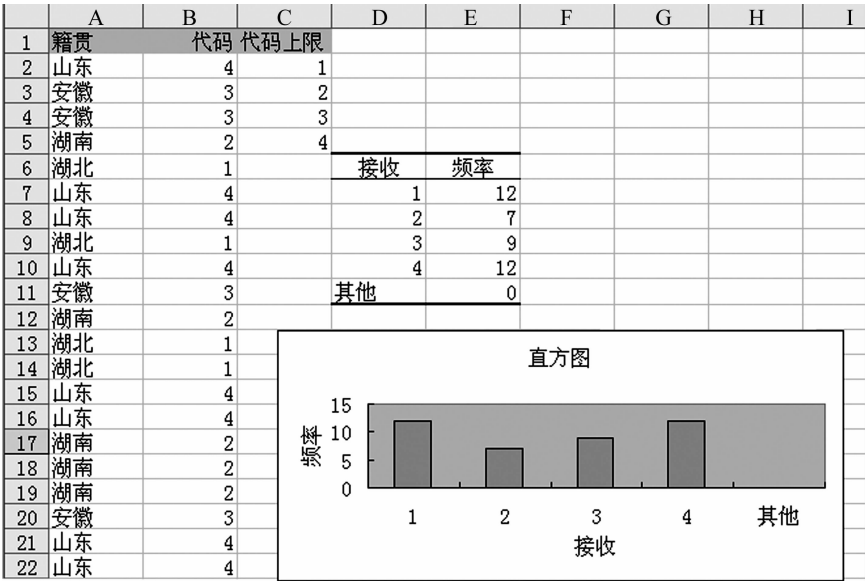


图 3-3 Excel 整理的结果

对输出结果进行还原并适当改造可得频数分布,见表 3-10。

表 3-10 某班同学籍贯分布表

籍 贯	频数(人)	比例(%)
湖北	12	30.0
湖南	7	17.5
安徽	9	22.5
山东	12	30.0
合 计	40	100

(二) 定序数据的分组

定序数据的分组方法和定类数据相类似,只是进一步要求各分组按顺序排列。如表 3-3 所反映的某单位职工文化程度的频数分布。

对于定序数据,除可汇总计算出频数分布和频率分布外,还可计算累积频数和累积频率。

1. 累积频数

累积频数(Cumulative Frequencies)就是依分组顺序将各组频数逐级累加起来。

其方法有两种:一是从变量值小的分组向变量值大的分组累加频数,称为向上累积;二是从变量值大的分组向变量值小的分组累加频数,称为向下累积。通过累积频数,可以很容易看出某一类别(或数值)以下及某一类别(或数值)以上的频数之和。

2. 累积频率或百分比

累积频率或累积百分比(Cumulative Percentages)就是将各分组的频率或百分比逐级累加起来,也有向上累积和向下累积两种方法。

【例 3-2】 在一项有关住房问题的研究中,研究人员在甲、乙两个城市各抽样调查 300 户家庭,其中的一个问题是:“您对您家庭目前的住房状况是否满意?”

(1) 非常不满意;(2)不满意;(3)一般;(4)满意;(5)非常满意

调查结果经整理如表 3-11 和表 3-12 所示。

表 3-11 甲城市家庭对住房状况的评价

回答类别	甲城市					
	户数 (户)	百分比 (%)	向上累积		向下累积	
			户数(户)	百分比(%)	户数(户)	百分比(%)
非常不满意	24	8	24	8.0	300	100
不满意	108	36	132	44.0	276	92
一般	93	31	225	75.0	168	56
满意	45	15	270	90.0	75	25
非常满意	30	10	300	100.0	30	10
合计	300	100	—	—	—	—

表 3-12 乙城市家庭对住房状况的评价

回答类别	乙城市					
	户数 (户)	百分比 (%)	向上累积		向下累积	
			户数(户)	百分比(%)	户数(户)	百分比(%)
非常不满意	21	7.0	21	7.0	300	100.0
不满意	99	33.0	120	40.0	279	93.0
一般	78	26.0	198	66.0	180	60.0
满意	64	21.3	262	87.3	102	34.0
非常满意	38	12.7	300	100.0	38	12.7
合计	300	100.0	—	—	—	—

(三) 数值数据(定距数据)的分组

定距数据的分组方法有两种:单项分组和组距分组。

1. 单项分组

单项分组就是每一个组中只有一个变量值。它适用于离散型变量的数据并且数据的范围不太大的情况下的分组。它是按变量值大小顺序排列的,一般有多少个不同的变量值就分为多少个组。如表 3-13。

表 3-13 某班同学年龄分布

年 龄	人数(人)	比例(%)
15	3	3.75
16	23	28.75
17	45	56.25
18	9	11.25
合 计	80	100.00

2. 组距分组

组距分组就是每个分组是一个数值区间。它适用于连续型变量或变动范围较大的离散型变量的数据分组。如表 3-14 就是一个组距分组。

表 3-14 某班同学数学成绩分布

成 绩	人数(人)	比例(%)
60 以下	2	2.50
60~69	15	18.75
70~79	36	45.00
80~89	20	25.00
90 以上	7	8.75
合 计	80	100.00

组距分组中,关键要处理好如下几个问题:

(1) 组数(Class Data)与组距(Class Width)。组距分组是用变量值变动的一个区间代表一个组,每个组的最大值为组的上限,最小值为组的下限。每个组的上限与下限间的距离称

为组距。具体分组时,先要找全距(R),即全部数据的最大值与最小值的距离;然后确定组数(m),实际工作中,主要凭经验确定,也可按不同的组数进行试验,比较其次数分布表,看哪一个能够更好地显示出分组数据的特征。另外有一个经验公式——“斯特杰斯规则”(Sturges' rule),可以作为确定组数的参考:

$$m=1+3.322\lg N$$
 (N 为总次数)

组数与组距(i)的关系是: $i=R/m$,两者成反比变化。

(2) 等距分组与异距分组。等距分组是指每个分组的组距都相等,否则,就是异距分组。

分组时究竟采取等距分组还是异距分组,应根据统计研究的目的和研究现象的特点来确定。一般当客观现象的变动比较均衡时,宜采用等距分组;而当客观现象的变动不均衡时,则宜采用异距分组。

在等距分组的频数分布中,由于各组的组距都相等,所以通过频数可以直接进行对比、分析。而在异距分组的分布中,各组频数受组距大小的影响,就无法直接对比、分析。如表 3-15 是某社区初婚年龄的分布,这是一个异距分组,如果用频数(人数)直接分析,就会得出“四五十岁初婚的人数比二十七八岁初婚的人数还多”的奇怪结论。而实际上,27~28 岁的组距仅为 1 岁,也就是说该组包括的年龄范围小;而 40~50 岁的组距为 10 岁,包括的年龄范围大,自然其频数也应该大些。所以,在异距分组中,不应该通过频数来分析、比较,而应该计算出各组的频数密度,通过频数密度来比较分析。

$$\text{频数密度}=\frac{\text{该组频数}}{\text{该组组距}}$$

表 3-15 某社区初婚年龄的频数分布

初婚年龄	人 数	频数密度
...	...	...
27~28	30	30
...	...	...
40~50	35	3.5
...	...	...

上例中,显然 27~28 岁的频数密度远远大于 40~50 岁的频数密度,这个结果显示,该社区的初婚年龄的分布是非常正常的。

(3) 组限(Class Limit)和组中值(Class Midpoint)。组限即每个分组的区间端点,其中,小的那个端点称为下限,大的那个端点称为上限。按照分组的组限情况,组距分组可分为两种:开口组和闭口组。开口组是指有上限没下限或有下限没上限的分组,闭口组是指既有上限又有下限的分组。如表 3-14 中,“60 分以下”和“90 分以上”都是开口组,而其他各组都是闭口组。在闭口组中,上限与下限的差称为组距,如“70~79 分”的组距为 9 分。而开口组的组距,依“以相邻的闭口组组距来界定”这一原则来确定。因此,在表 3-14 中,“60 分以下”的组距为 9 分,因为它相邻的闭口组“60~69”的组距为 9 分。“90 分以上”的组距也为 9 分。要注意的是,这一原则只是为了确定开口组的组距而设,在实际汇总的时候,当然所有“60 分以下”的成绩都要归入这一组,而不能仅仅汇总 51~60 分之间的成绩。

此外,按照各分组中组限是否重合,可将组距分组方法分为间断组距式分组和连续组距式分组。间断组距式分组是指各组的组限不重合,如表 3-14。间断组距式分组适合于离散型

变量的数据分组(如,某班成绩只取整数,可采用表 3-14 的方式分组)。连续组距式分组是指各组组限是重合的,如表 3-16。它适用于连续型变量的数据分组(如,如果成绩数据有小数形式,可采取表 3-16 的连续组距式分组)。

表 3-16 某班同学数学成绩分布

成 绩	人数(人)	比例(%)
60 以下	2	2.50
60~70	15	18.75
70~80	36	45.00
80~90	20	25.00
90 以上	7	8.75
合 计	80	100.00

在连续组距式分组中,由于相邻组的组限是重合的,为满足“互斥原则”,就应遵循“上限不在内”原则,即各组上限不包括在本组之内。如在表 3-16 中,70 分应归入第三组“70~80”内。

在统计分析时,通常会以组中值来代表各组标志值的平均水平。所谓组中值,就是各组的中心值,计算公式:

闭口组的组中值=(上限+下限)/2

缺下限的开口组组中值=上限-邻组组距/2

缺上限的开口组组中值=下限+邻组组距/2

解决了上述问题,就可以编制出组距分布了。

【例 3-3】 某厂对 110 件产品的高度进行测量(单位:厘米),数据如下,编制次数分布表。

154 133 116 128 85 100 105 150 118 97 110 131 119 103 93

108 100 111 130 104 135 113 122 115 103 90 108 114 127 87

127 108 112 100 117 121 105 136 123 108 89 94 139 82 113

110 109 118 115 126 106 108 115 133 114 119 104 147 134 117

119 91 137 101 107 112 121 125 103 89 110 122 123 124 125

115 113 128 85 113 143 80 102 132 96 129 83 142 112 120

107 108 111 100 97 111 131 109 145 93 135 98 142 127 106

110 101 116 110 123

解:第一步,先将 110 个数据排序,找出最大值 154 和最小值 80,这个数列的全距 $R=154-80=74$ 厘米。

第二步,根据斯特杰斯规则确定组数: $m=1+3.322\times(\lg 110)=7.78$,再根据组数与组距的关系确定组距: $i=R/m=74/7.78=9.51$ (厘米)。根据以上的计算结果,组数定为 8 组;组距定为 10 厘米。

要注意的是,在用经验公式计算 m 和 i 时,计算结果的取舍,不采用四舍五入法,而采用舍去进一法,即:只要有小数,就把小数舍去,并在整数位上加 1。这种做法保证次数分布表有足够宽的覆盖区间。另外,一般说来组距宜于取整百整十,起始组的下限也宜于取整百整十,这样看起来比较舒服。还有,有些数据本身是有特殊或固定的分组要求的,如学生成绩如果

出现“54~62”这样一组,则将不同性质的学生混在了一起,即在这组里有成绩不合格的学生,又有成绩合格的学生,这样的分组肯定是错误的。

第三步,根据所定组数和组距确定组限。第一组下组限定为 80,第一组上组限则为 90 (即 80+10);第二组下组限就是第一组上组限,第二组上组限为 100……;依此类推,第八组下组限是 150,其上组限则为 160。这样共有 8 个下组限和 8 个上组限。由于有重合值,故只有 9 个组限值。

第四步,进行归组,即将各个变量值归入相应的组中,比如 154 归入第八组(150~160); 133 归入第六组(130~140);……;依此类推。最后的结果用次数分布表显示,见表 3-17。

表 3-17 产品高度的频数分布表

产品高度 x (厘米)	件数 f (件)
80~90	8
90~100	9
100~110	26
110~120	30
120~130	18
130~140	12
140~150	5
150~160	2
合 计	110

如果数据量很大,归组的工作量会很大,Excel 中有“FREQUENCY”函数,可以完成这一任务。

组距分布中,也可以计算累积频数和累积频率。各组的向上累积频数,是指小于该组上限的总频数;而向下累积频数,是指大于该组下限的总频数。

表 3-18 某种产品高度的频数分布表

产品高度 x (厘米)	频数 f (件)	频率(%)	向上累计		向下累计	
			频数(件)	频率(%)	频数(件)	频率(%)
80~90	8	7.3	8	7.3	110	100.0
90~100	9	8.2	17	15.5	102	92.7
100~110	26	23.6	43	39.1	93	84.5
110~120	30	27.3	73	66.4	67	60.9
120~130	18	16.4	91	82.8	37	33.6
130~140	12	10.9	103	93.7	19	17.2
140~150	5	4.5	108	98.2	7	6.3
150~160	2	1.8	110	100.0	2	1.8
合 计	110	100	—	—	—	—

(四) 简单分组与复合分组

对统计数据进行的分组是按照不同的标志进行的,分组标志就是进行分组的标准和依据。按分组标志的多少,统计分组可分为简单分组与复合分组。简单分组是按照一个分组标志对所研究的对象进行分组。如人口按性别分为男、女两组。复合分组是按照两个或两个以上的分组标志对所研究的对象进行分组。这种分组,先按一个分组标志对所研究对象进行分组,再按第二个分组标志进一步分组,然后再次层叠地按第三个分组标志分成更小的组。如表 3-19 是对高校教师进行的复合分组。

表 3-19 高校教师的复合分组

第一标志(职称)	第二标志(年龄)	第三标志(性别)
高级职称	45 岁以下	男
		女
	45 岁以上	男
		女
非高级职称	45 岁以下	男
		女
	45 岁以上	男
		女

复合分组也可表达成表 3-20 的形式。

表 3-20 某学校教师频数分布表

职称	年龄	45 岁以下		45 岁以上	
	性别				
		男	女	男	女
高级职称					
非高级职称					

第三节 统 计 表

经过整理的统计数据资料,可以通过统计表和统计图来表达。用统计表来表达频数分布不仅可以节省大量的文字叙述,而且更为集中醒目、条理分明,也便于资料的对比分析与积累。

一、统计表的 概念和结构

1. 统计表的 概念

统计表是指把经过整理的数 据资料,通过表格的形式表达出来。统计表可分为广义统计表和狭义统计表两种。广义的统计表,包括统计工作各阶段中所用的一切表格;狭义的统计表专指分析表和容纳各种统计资料的表格,通常所说的统计表即狭义的统计表。统计表是统计分析

的重要工具,它不仅能清楚、有条理地显示统计资料,并能直观地反映统计分析的特征。

2. 统计表的结构

统计表的构造,从形式来看,其构成要素包括:总标题、横行标题、纵行标题、数字资料四个部分。总标题置于表的正上方,是统计表的名称,它简明扼要地说明全表的基本内容。横行标题置于表的左端,是横行的名称,即总体各组名称,表示统计研究的对象,亦称主词。主词所在栏称为主词栏。纵行标题置于表的右上端,是纵行的名称,连同表中数字,称为宾词。宾词所在栏称为宾词栏。见表 3-21。

表 3-21 1998 年底某地社会劳动者分布状况 ←总标题

横行标题	按三次产业分组	社会劳动者		纵行标题
		绝对数(万人)	比重(%)	
	第一产业	500	41.7	数字资料
	第二产业	400	33.3	
	第三产业	300	25.0	
	总 计	1 200	100.0	
主 词		宾 词		

二、统计表的分类与设计规则

1. 统计表的分类

(1) 简单表,是指主词未经任何分组的统计表,也叫做一览表。简单表的主词一般按时间顺序排列,或者按个体的名称排列。它是对原始资料进行初步整理所采用的形式。

(2) 简单分组表,是指主词只用一个标志分组形成的统计表,也叫做分组表。运用简单分组表可以说明不同类型现象的特征,以揭示现象内部的结构,以便分析现象之间的相互关系,如表 3-21 所示。

(3) 复合分组表,是指主词按两个或两个以上标志进行分组的统计表,也叫做复合表。复合分组表可以通过更多的标志,对总体进行更深入地分析与研究。如表 3-20。

注意简单表和简单分组表的区别。简单表是指未分组的统计表,如第五章中所言的时间数列表;而简单分组表是指按照一个标志进行分组的统计表。

2. 统计表设计规则

统计表的设计应尽可能做到简洁、清晰、准确、醒目,便于使用者进行比较、分析以及阅读。设计时应遵循如下规则:

(1) 统计表的各类标题应十分简明,并确切地反映与概括资料的主要内容以及所属的地区和时间。纵横各栏的排列特别要注意表述资料的逻辑性。

(2) 主词各行和宾词各栏,一般先列各个项目,后列总体。若无必要列出全部项目时,就应先列总体,后列其中一部分重要项目。内容不宜罗列太多和过于庞杂。

(3) 表中应有计量单位栏。当表中只有一种计量单位时,可在表的右上端注明。若有几个计量单位时,横行的计量单位可专设“计量单位”一栏,纵栏的计量单位可与纵栏标题写在一起,用小括号标明。

(4) 表中数字填写要整齐,上下位数要对齐,同栏数字的单位,小数位要一致。如遇相同

数字必须照填,不能用“同上”或“同左”代替。无数字的空格要用“—”表示。如遇缺乏资料的空格时,要用“……”表示,以免被误认漏报。

(5) 当统计表栏数较多时,通常要加编号,并说明其相互关系,主词栏与计量单位栏可用甲、乙、丙等文字标明;宾词各栏可用(1)、(2)、(3)等数码标明。

(6) 统计表的表式为开口式,即表的左右两端不封闭(不画纵线),表的上下端线通常用粗线或双线,表内如有两个或两个以上不同的内容,也要用粗线或双线隔开。

(7) 借用他人数据资料时,统计表应加注解,说明资料出处,一般在统计表的下端注明“资料来源”。

第四节 统计图

统计图是统计资料的另外一种表达方式,它可以简洁直观地表示统计表中枯燥的数据,帮助我们从众多的数据中发现规律,更迅速、更有效地传递信息,给人以明确而深刻的印象。

一、统计图的结构

统计图基本包括以下几部分:

(1) 标题。统计图一般包括图表标题、数值轴(X、Y)标题。

(2) 坐标轴和网格线。坐标轴和网格线构造了绘图区的骨架,借助坐标轴和网格线,我们可以更容易读懂统计图。

(3) 图表区和绘图区。统计表的所有内容都在图表区内,包括绘图区。统计图绘制在绘图区内。

(4) 图例。图例用来标明图表中的数据系列。当只有一个序列时,可以省略图例。有时统计图中的数据系列可能有多个,我们可以用不同颜色、形状的图例来区别不同的数据系列。

统计图的种类很多,常用的有用于辅助统计分析的直方图、趋势图、散点图;有擅长直观表现数据的柱形图、饼图、圆环图等。Excel 提供了 14 种标准图表类型,见图 3-4,每种标准图表类型还可以包含几种不同的子类型,我们可以根据数据的类型以及自己的需要决定采用哪种图形来表达数据。



图 3-4 Excel 中标准图表类型

二、定类数据的统计图

适用于定类数据的统计图主要有饼图、条形图等。

(一) 饼图

饼图(Pie Chart)也称圆形图,是用圆形及圆内扇形的面积来表示数值大小的图形。圆形图主要用于表示总体中各组成部分所占的比例,对于研究结构性问题十分有用。

在绘制饼图时,总体中各部分所占的百分比用圆内的各个扇形面积表示,这些扇形的圆心角度数是按各分组(即各类)的百分比乘以 360° 来确定的。例如,关注服务广告的人数占总人数的百分比为 25.5% ,那么其对应的扇形圆心角度数就应为 $360^\circ \times 25.5\% = 91.8^\circ$ 。

【例 3-4】 为研究广告市场的状况,一家广告公司在某城市随机抽取 200 人就广告问题做了邮寄问卷调查,其中的一个问题是:“您比较关心下列哪一类广告?”

a. 商品广告; b. 服务广告; c. 金融广告; d. 房地产广告; e. 招生招聘广告; f. 其他广告。

这里的变量就是“广告类别”，不同类型的广告就是变量值。调查数据经分类整理后形成频数分布表，见表 3-22。

表 3-22 某无奢望居民关注广告类型的频数分布

广告类型	人数(人)	比 例	频率(%)
商品广告	112	0.560	56.0
服务广告	51	0.255	25.5
金融广告	9	0.045	4.5
房地产广告	16	0.080	8.0
招生招聘广告	10	0.050	5.0
其他广告	2	0.010	1.0
合 计	200	1	100

表 3-22 的频数分布可用饼图来表达,这可以借助于 Excel 中的“图表”功能来操作(其实在 Microsoft Office 中的 Word、Excel 和 Powerpoint 等系统中均具有“图表”功能,这里仅就 Excel 做一介绍)。具体步骤是:

(1) 将表 3-22 数据录入 Excel 工作表中, 点击“插入”, 然后点击“图表”, 出现如图 3-5 对话框。



图 3-5 图表向导——图表类型

(2)在对话框中选择“饼图”及“子图表类型(本例选择三维饼图)”,然后单击对话框中的“下一步”,出现如图 3-6 对话框。



图 3-6 图表向导——图表源数据

(3)在对话框中,“数据区域”一栏输入工作表中广告类型及频数所在的区域 \$A\$1:\$B\$6,然后点“下一步”,出现图 3-7 对话框。



图 3-7 图表向导——图表选项

(4)点击“下一步”,出现如图 3-8 对话框。然后选择图形输出的方式(本例选择“作为其中的对象插入”),然后点击“完成”,即得所需的饼图(见图 3-9)。



图 3-8 图表向导——图表位置

绘制的圆形图如图 3-9 所示。

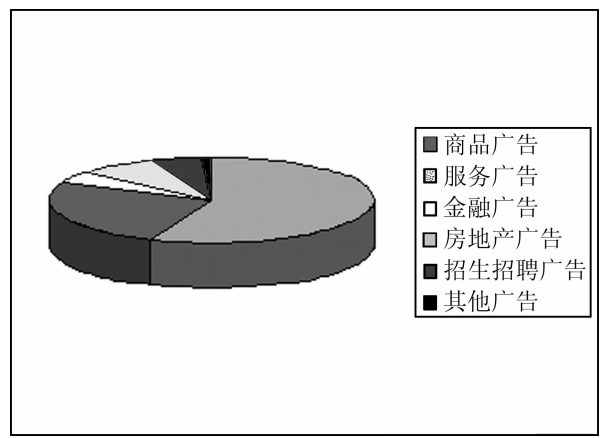


图 3-9 某城市居民关注不同类型广告的人数分布

(二) 条形图

条形图 (Bar Chart) 是用宽度相同的条形的高度或长短来表示频数分布的统计图形。条形图可以横置也可以纵置,纵置时也称为柱形图。

在表示定类数据的分布时,用条形图的高度来表示各类别数据的频数或频率。绘制时,各类别可以放在纵轴,称为条形图;也可以放在横轴,称为柱形图。

根据表 3-22 绘制的条形图如图 3-10 所示。

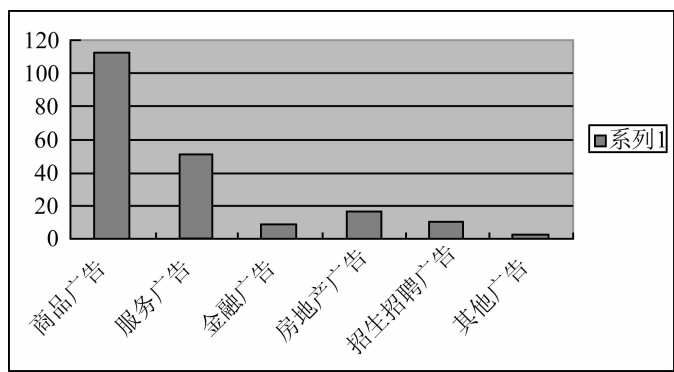


图 3-10 某城市居民关注广告类型频数分布的条形图

Excel 中,条形图的制作方法和饼图类似,这里不再赘述。

三、定序数据的整理与显示

由于定序数据中不仅有类别属性还有顺序属性,因此,上述所介绍的饼图不能适用于定序数据。条形图虽然可以用来表达定序数据的频数分布,但要注意,其分组(即类别)必须要按照顺序排列。

此外,条形图不仅可以表达定序数据的频数分布,而且还可以表达累积频数分布。如本章第二节例 2 中表 3-11 和表 3-12 的累积频数就可以表示为图 3-11 和图 3-12。

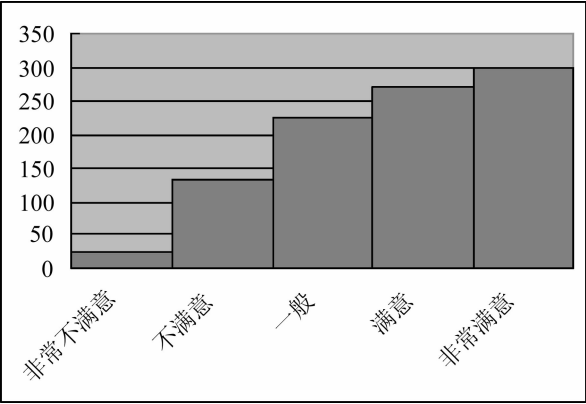


图 3-11 甲城市对住房状况评价的累积分布图

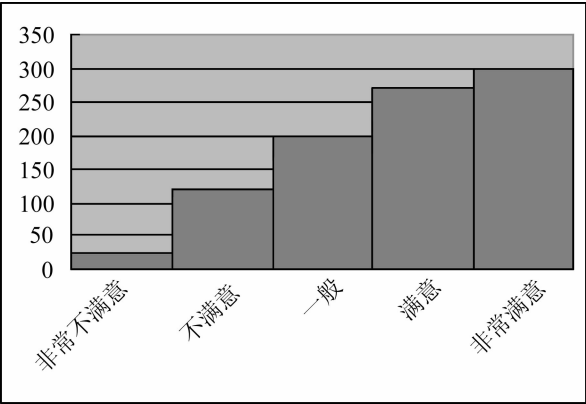


图 3-12 乙城市对住房状况评价的累积分布图

对于多总体定序数据,还可以采取环形图(Doughnut Chart)来表达。

环形图(Doughnut Chart)与圆形图类似,但又有区别。环形图中间有一个“空洞”,总体中的每一部分数据用环中的一段表示。圆形图只能显示一个总体各部分所占的比例,而环形图则可以同时绘制多个总体的数据系列,每一个总体的数据系列为一个环,因此环形图可显示多个总体各部分所占的相应比例,从而有利于进行比较研究。例如根据表 3-11 和表 3-12 数据绘制两个城市家庭对住房状况评价的环形图如图 3-13 所示:

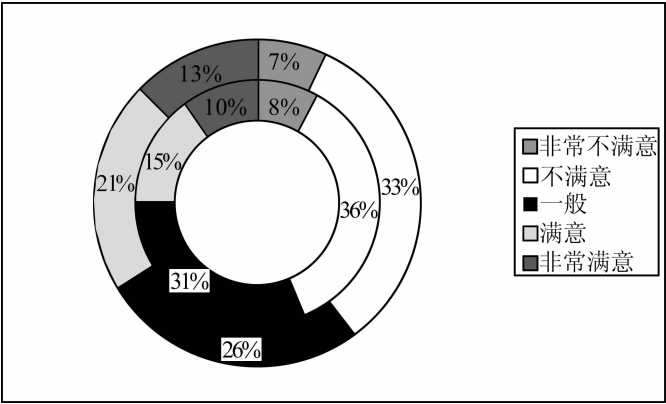


图 3-13 甲、乙两城市家庭对住房状况的评价

在图 3-13 中,外边的一个环表示的是乙城市家庭对住房状况评价各类所占的百分比,里边的一个环则为甲城市家庭对住房状况评价各类所占的百分比。

四、定距数据的统计图

定距数据可分为离散型和连续型两类。离散型定距数据可视为定序数据,可采取条形图等形式来表达。连续型定距数据频数分布的统计图则有直方图、折线图等多种类型。

(一) 分组数据

1. 直方图

直方图(Histogram)是用矩形的宽度和高度来表示频数分布的图形。在平面直角坐标中,以横轴表示数据分组,纵轴表示频数或频率,各分组以组距为宽,以频数或频率为高(异距分布中,以频数密度或频率密度为高)画矩形,即得直方图。比如,根据例 3-3 中表 3-17 的频数分布绘制成的直方图如图 3-14 所示。

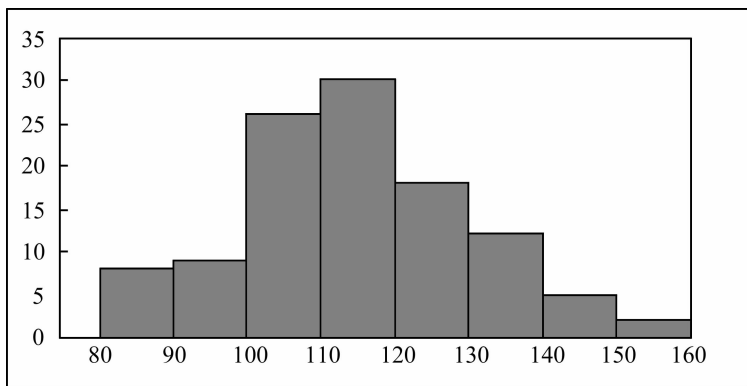


图 3-14 某产品频数分布的直方图

直方图与条形图不同,条形图是用条形的长度(横置时)表示各类别频数的多少,其宽度(表示类别)是固定的;直方图是用面积表示各组频数的多少,矩形的高度表示每一组的频数密度,宽度则表示各组的组距,因此其高度与宽度均有意义。此外,由于分组数据具有连续性,因此直方图中的矩形必须是连续排列,而条形图则可以是分开排列。

2. 折线图

折线图也称频数多边形图(Frequency Polygon),是在直方图的基础上,把直方图中各矩形顶部的中点(即组中值)用直线连接起来,即得折线图。需要注意,折线图的两个终点要与横轴相交,具体的做法是将第一个矩形的顶部中点通过竖边中点(即该组频数一半的位置)连接到横轴,最后一个矩形顶部中点与其竖边中点连接到横轴。这样才会使折线图下所围成的面积与直方图的面积相等,从而使两者所表示的频数分布一致。例如,在图 3-14 的基础上绘制的折线图如图 3-15 所示。折线图可以反映频数的分布趋势。

当对数据所分的组数很多时,组距会越来越小,这时所绘制的折线图就会越来越光滑,逐渐形成一条平滑的曲线,这就是频数分布曲线。分布曲线在统计学中有着十分广泛的应用,是描述各种统计量和分布规律的有效方法。

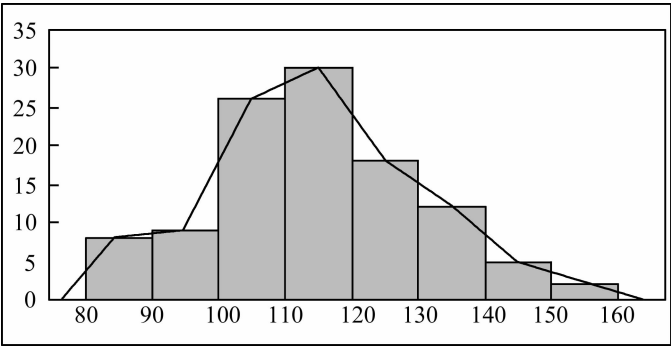


图 3-15 某产品频数分布的折线图

(二) 未分组数据

1. 茎叶图

对于未分组的原始数据,可以用茎叶图(Stem-and-Leaf Display)来显示其分布特征。虽然通过直方图可以大体上看出一组数据的分布状况,但直方图没有给出具体的数值。而这里介绍的茎叶图,既能给出数据的分布状况,又能给出每一个原始数值。

茎叶图由“茎”和“叶”两部分构成,其图形是由数字组成的。通过茎叶图,可以看出数据的分布形状及数据的离散状况,比如,分布是否对称,数据是否集中,是否有极端值等等。绘制茎叶图的关键是设计好树茎,通常是以该组数据的高位数值作为树茎。树茎一经确定,树叶就自然地长在相应的树茎上了。

【例 3-5】 某车间工人某天日产量数据如下:

107 108 108 110 112 112 113 114 115 117 117 117 118
118 118 119 120 120 121 122 122 122 122 123 123 123
124 124 124 125 125 126 126 127 127 127 128 128 129
129 130 131 133 133 134 134 135 137 139 139

试制作茎叶图。

解:将每个数据分解为两部分:前两位作为一部分,个位数作为一部分,如 107 分解为 10 | 7,118 分解为 11 | 8 等。然后以前两位为树茎,以个位为树叶,制作出茎叶图如图 3-16 所示。

树茎	树叶	数据个数
10	7 8 8	(3)
11	0 2 2 3 4 5 7 7 7 8 8 8 9	(13)
12	0 0 1 2 2 2 2 3 3 3 4 4 4 5 5 6 6 7 7 7 8 8 9 9	(24)
13	0 1 3 3 4 4 5 7 9 9	(10)

图 3-16 某车间工人日加工零件数的茎叶图

上面的茎叶图显得过于拥挤,可以把它扩展,形成扩展的茎叶图。比如,可以将图 3-16 扩展 1 倍,即每一个树茎重复两次,一次有记号“*”,表示该行叶子上的数为 0~4,另一次有

记号“·”，表示该行叶子上的数为 5~9，于是可得到图 3-17(a)。

对于一组数据，茎叶图到底有多少行比较合适呢？经验表明，如果数据的个数为 $n(20 \leq n \leq 300)$ ，则茎叶图的最大行数不超过 $L = [10 \times \lg n]$ 。式中，方括号表示括号中数据的整数部分。如本例中 $n = 50$ ，则 $L = [10 \times \lg 50] = [16.99] = 16$ 。按照这个标准，茎叶图的行数应为 16 行。因此，我们还可以再“拉长”或“扩展”茎叶团。比如把图 3-17(a)中的茎拉长 5 位，即每个茎上的数重复 5 次。其中有记号“*”的茎，叶子上的数为 0 和 1；有记号“t”的茎，叶子上的数为 2 和 3；有记号“f”的茎，叶子上的数为 4 和 5；有记号“s”的茎，叶子上的数为 6 和 7；有记号“·”的茎，叶子上的数为 8 和 9。可绘制得扩展的茎叶图，见图 3-17(b)。

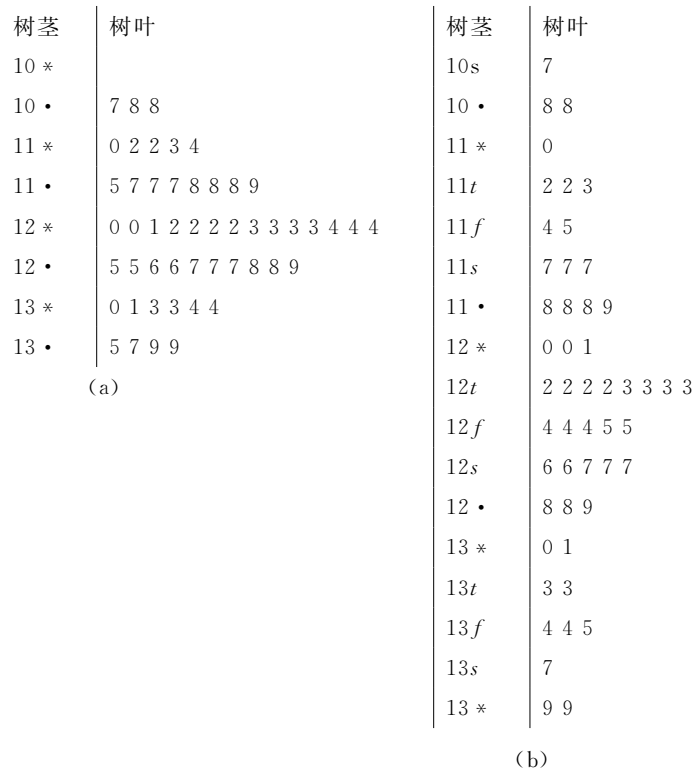


图 3-17 扩展的茎叶图

图 3-17(a)相当于对原始数据按照组距为 5 进行的分组，也就是每一行的数据宽度为 5；而图 3-17(b)则相当于按照组距为 2 进行的分组，即每一行的数据宽度为 2。当然，根据经验公式确定的行数只是一个大致的标准。在实际应用中，茎叶图行数的确定还要根据数据的分散状况及数据分布的特征来确定，总之要以能充分显示出数据的分布特征为目的。

2. 箱线图

箱线图(Box Plot)是通过一个箱线把一组数据的 5 个特征值(最大值、最小值、中位数、两个四分位数)表达出来,从而反映原始数据分布状况的统计图形。

箱线图分为简单箱线图和多重箱线图。简单箱线图适用于单变量定距数据,而多重箱线图适用于多变量定距数据。

(1) 简单箱线图。简单箱线图是将一组数据的 5 个特征值通过一个箱子和两条线段绘制出来。其绘制方法是:首先找出一组数据的 5 个特征值,即数据的最大值、最小值、中位数和两个四

分位数(中位数 M_e 是一组数据排序后处于中间位置上的变量值,四分位数是处在数据 25% 位置和 75% 位置上的两个值,分别称为下四分位数和上四分位数,下一章详细介绍其计算方法);然后,连接两个四分位数画出箱子;再将两个极值点与箱子相连接。见图 3-18 所示。



图 3-18 简单箱线图

例如,根据例 3-5 的数据绘制的箱线图如图 3-19。

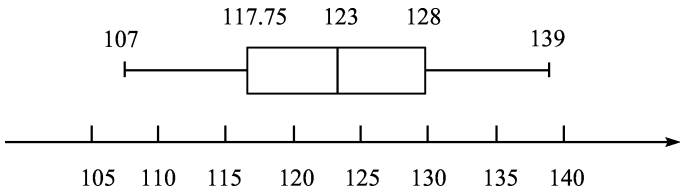


图 3-19 50 名工人日产量箱线图

通过箱线图的形状,可以看出数据分布的特征。图 3-20 就是几种不同分布的箱线图。

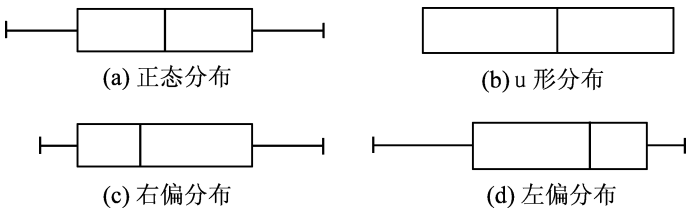


图 3-20 不同类型分布的箱线图

(2)多重箱线图。对于多变量定距数据,可以将各变量的数据的箱线图并列起来,用多重箱线图来表示。

【例 3-6】 从某班学生中随机抽取 11 人,对 8 门主要课程的考试成绩进行调查,所得结果见表 3-23。试绘制各科考试成绩的箱线图,并分析各科考试成绩的分布特征。

表 3-23 某班同学考试成绩

课程名称	学生编号										
	1	2	3	4	5	6	7	8	9	10	11
英语	76	90	97	71	70	93	86	83	78	85	81
经济数学	65	95	51	74	78	63	91	82	75	71	55
西方经济学	93	81	76	88	66	79	83	92	78	86	78
市场营销学	74	87	85	69	90	80	77	84	91	74	70
财务管理	68	75	70	84	73	60	76	81	88	68	75
基础会计学	70	73	92	65	78	87	90	70	66	79	68
统计学	55	91	68	73	84	81	70	69	84	62	71
计算机应用基础	85	78	81	95	70	67	82	72	80	81	77

解:为绘制箱线图,首先计算出 11 名同学各科成绩的五个特征值,见表 3-24。

表 3-24 各科成绩特征值

课程名称	最小值		Q _L			M _e			Q _U		最大值
英语	70	71	76	78	81	83	85	86	90	93	97
经济数学	51	55	63	65	71	74	75	78	82	91	95
西方经济学	66	76	78	78	79	81	83	86	88	92	93
市场营销学	69	70	74	74	77	80	84	85	87	90	91
财务管理	60	68	68	70	73	75	75	76	81	84	88
基础会计学	65	66	68	70	70	73	78	79	87	90	92
统计学	55	62	68	69	70	71	73	81	84	91	94
计算机应用基础	67	70	72	77	78	80	81	81	82	84	95

根据表 3-24 的计算结果绘制的箱线图如图 3-21 所示。

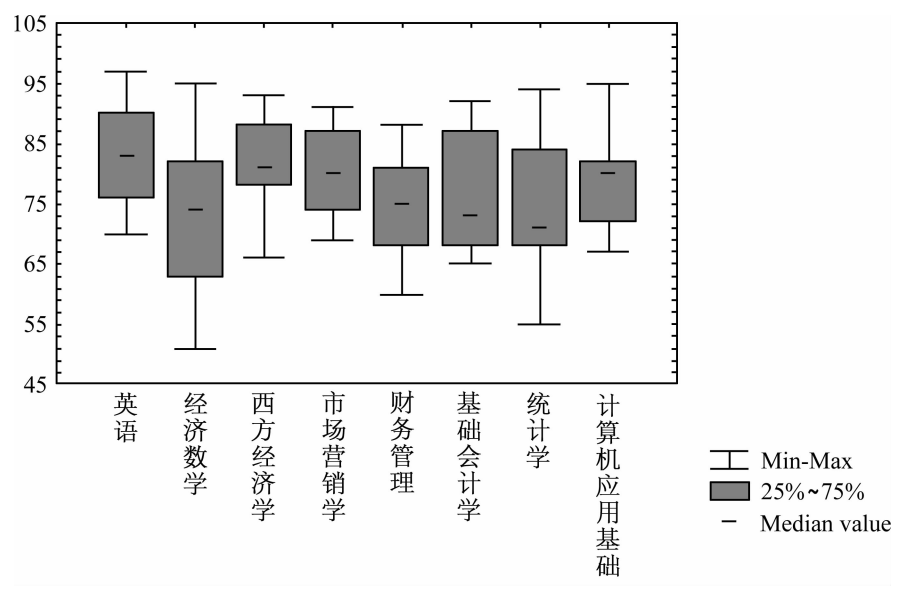


图 3-21 各科成绩箱线图

由图 3-21 可以看出,在 8 门主要课程中,平均考试成绩较高的是英语和西方经济学,较低的是经济数学和统计学。从考试成绩的离散程度来看,英语和市场营销学的考试成绩比较集中,大体上为对称分布;而经济数学、基础会计学、统计学和计算机应用基础等课程的考试成绩则比较分散。

实践技能训练

项目名称:

本班学生统计学考试成绩的统计分析

实训目的:

通过本次训练达到:

1. 深入理解和熟练掌握频数分布表的制作及分组标志及类型。
2. 加强理论和实践的联系,提高学生将统计方法应用于日常学习与生活。

实训内容与要求:

请你对本班全体同学的统计学考试成绩进行统计整理,分析本班该课程考试情况。

1. 根据本班实际人数情况确定样本容量。
2. 将该班学生分为不及格、及格、中、良、优五组,编制一张频数分配表。

其中,学校规定:60 分以下为不及格,60~70 为及格,70~80 分为中,80~90 分为良,90~100 分为优。

**思考练习****一、填空题**

1. _____ 是统计整理的初始阶段,也是整个统计整理工作的基础。
2. 所谓 _____,是指根据研究目的和研究设计,对数据做进一步的补充和修正,以满足统计研究的要求。
3. 为保证统计分组的科学性,统计分组要遵循 _____ 原则和 _____ 原则。
4. 定距数据的统计分组可分为 _____ 分组和 _____ 分组两种。
5. 从形式上看,统计表由 _____、_____、_____ 和 _____ 四部分组成。
6. 折线图适用于表达 _____ 数据的频数分布。

二、计算题

1. 某车间 40 名工人定额完成率(%)统计如下:

97	88	123	115	119	158	112	146
117	108	105	110	107	137	120	136
125	127	142	118	103	87	115	114
117	124	129	138	100	103	92	95
113	126	107	108	105	119	127	104

将上述资料整理成频数分布,组距可划分为:100 以下、100~110、110~120、120~130、130 以上,并绘制直方图。

2. 从某大学一年级学生中随机抽取 36 人,对公共理论课的考试成绩进行的调查,结果如下:

67	90	66	80	67	65	74	70	87
85	83	75	58	67	54	65	79	86
89	95	78	97	76	78	82	94	56
60	93	88	76	84	79	76	77	76

要求:根据以上数据将考试成绩等距分为 5 组,组距为 10,并编制成绩次数分布表,绘制次数分布直方图。

3. 某企业生产的产品需经 6 道工序,为提高该产品的质量,寻找出一季度全部废品产生的原因,特对各道工序进行了质量检查,结果如下表所示。

废品数	
工序名称	废 品 数
A	2 606
B	1 024
C	355
D	59
E	28
F	25
合 计	4 097

要求作出累计频率分布表,并绘制直方图。

三、简答题

- 1. 统计表由哪几个主要部分组成?
- 2. 制作统计表应注意哪几个问题?
- 3. 直方图与条形图有什么区别?
- 4. 饼图和环形图有什么不同?