

高等学校财经类核心课程教材

应用统计学实验教程

吴先华 李 琰 姚天祥 编著

高等学校财经类核心课程教材

应用统计学实验教程

吴先华 李 琰 姚天祥 编著

YINGYONG TONGJIXUE SHIYAN JIAOCHENG



高等教育出版社·北京
HIGHER EDUCATION PRESS BEIJING

内容简介

随着信息时代的来临,需要对各行业科学分析与精细管理理念逐步深入,统计分析已渗透到生产、生活和科学研究的各个领域。掌握统计学基本理论,借助统计软件分析数据也成为本科学子必须掌握的基本技能之一。本书围绕统计分析的实际操作要求,基于多种统计软件的最新发展,介绍了各种统计方法的主要思想和分析思路,其中 Excel 作为最流行、操作最简便的数据处理和分析软件,SPSS 作为社会科学领域应用最广的专业统计分析软件,是本书介绍统计分析步骤的重点依托工具,另有部分章节引入了基于其他软件(Stata, Eviews)的分析办法作为有益的对比。以此为基础,借用大量实用案例详细讲解了各种方法的操作重点,解释了每个实例的统计输出结果,并且给予读者以统计方法和应用的启示。

本书结构完整、内容全面,有利于读者针对具体问题选择适用的统计软件进行学习,是一本综合性和实用性较强的统计学实验教程。

本书既可作为高等院校经济管理类专业本科生统计实验课程的教材,也可作为非经管专业本科生统计实验课程的教材或参考书,对广大实际工作者也具有参考价值。

图书在版编目(CIP)数据

应用统计学实验教程/吴先华,李琰,姚天祥编著.

--北京:高等教育出版社,2014.2

ISBN 978-7-04-028993-0

I. ①应… II. ①吴… ②李… ③姚… III. ①应用
统计学-实验-高等学校-教材 IV. ①C8-33

中国版本图书馆 CIP 数据核字(2014)第 017467 号

策划编辑 施春花
插图绘制 邓超

责任编辑 范吕
责任校对 杨凤玲

封面设计 王 睢
责任印制 赵义民

版式设计 范晓红

出版发行 高等教育出版社
社 址 北京市西城区德外大街 4 号
邮政编码 100120
印 刷 大厂益利印刷有限公司
开 本 787mm × 1092mm 1/16
印 张 13.75
字 数 300 千字
购书热线 010-58581118

咨询电话 400-810-0598
网 址 <http://www.hep.edu.cn>
<http://www.hep.com.cn>
网上订购 <http://www.landaco.com>
<http://www.landaco.com.cn>
版 次 2014 年 2 月第 1 版
印 次 2014 年 2 月第 1 次印刷
定 价 24.70 元

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换。

版权所有 侵权必究

物 料 号 28993-00

前 言

读者希望从统计类书籍中学到什么？不仅仅是公式、技巧或推理，更多的是一些基本的统计分析思路和统计操作方法，并能应用到他们的专业和日常生活中。根据《中华人民共和国高等教育法》的规定，“高等教育的任务是培养具有创新精神和实践能力的高级专门人才”。由于实验教学在提高学生的综合素质、丰富教学方法等方面发挥巨大作用，故而优秀的实验教材的建设具有重大意义，本书就是一本遵照新时期教学内容及课程体系改革和与时俱进精神的创新性统计学实验教程。

本书主要的目的是让学生更好地了解统计分析在从日常生活到经济管理乃至科学研究等各个领域的重要作用，以及这些作用是如何体现的。在此基础上，掌握获取数据、选择统计软件、整理和分析数据、科学解读数据的原则和方法，最后形成一整套科学分析目标问题的思路体系。具体来看，特点如下：

第一，紧密联系统计学理论课程。与以往统计学实验教材注重软件的应用步骤不同，本书将在各实验章节开篇系统回顾理论知识，并在解决问题的过程中系统讲述各统计参数的联系及其理论意义和现实意义。为增强读者阅读本书和学习统计学的兴趣，每章末还给出了延伸阅读的文献与习题，为读者留下了启发思考的空间。

第二，统计方法与经济管理实际问题的有机结合。本书多选择近期经济管理和社会科学领域受到广泛关注的问题和数据，并将在未来更新的版次中不断更新问题及数据。同时，读者还可以针对特定的统计任务，直接按照目录“跳跃”学习，使得本书的参考价值得到进一步的提升。

第三，统计分析与统计软件相结合。本书不限定单一软件，不但系统地比较了各种软件的较新版本（Excel 2010, SPSS 20.0, Stata 12.1, Eviews 7.0 等）的特点，还在各章节中具体比较了不同软件的应用效果，有助于读者根据不同问题和自身特点选择适当的统计软件学习。

第四，体现了教学改革探索的成果。在教学环节，学生通过与之相匹配的实验教学，可以更深刻地掌握相关的统计基础理论知识，同时也提高了利用理论知识解决实际问题的能力，使得单一的知识传授型教学转变为素质教育为主的实践式教学，在提高学生参与意识的同时极大培养其创新精神和科研能力。

本书由吴先华、李琰、姚天祥编著，孙香玉、张三峰、郭建平、魏娟、来向红、储小俊、盛济川等共同参与编写。以上作者都是南京信息工程大学经济管理学院“统计学”课程组的成员。本课程组于2009年成立，课程组成员在此之前分别承担过不同专业的“统计学”、“应用统计学”、“管理定量方法”、“社会经济统计学”等多门统计类课程的教学，在全面推进教学体系改革思路的指引下，各位任课教师认真总结多年教学过程中的实践经验，在对传统的课程进行改革的基础上，整合了一门内容和结构全新的“统计学”课程。随着电子计算机的应用和普及，统计学实验在

“统计学”课程中所占比重越来越大,探索实验教学的内容和方法、以多种形式开展实验教学成为本课程组必须面对的重要课题。为提高实验教学质量、探讨实验教学规律,总结和交流经济管理统计实验教学经验,迫切需要一本包含更多统计学实验知识点、反映各种软件功能的统计学实验教程,一方面作为理论课程的有益补充,另一方面可为学生统计实践提供指导,在此背景下,本书应运而生。

指导思想明确后,我们多次研讨与修订大纲,在此基础上组织老师撰写初稿,并反复修改。该书的内容已部分应用到课堂实验教学中,并取得了较好的反馈。总体来看,本实验教程体现了理论和实践的结合,较好地反映了经济管理领域的实践活动和统计分析的水平,具有一定的先进性和实用性。

感谢参与编写的教师们,感谢出版社的同仁和为本书编写贡献意见的同行专家,经济管理实验教学尚处于探索、提高阶段,我们衷心地期待经济管理界的专家、学者和广大读者对本书提出宝贵的意见和建议。

编者

2013年10月

郑重声明

高等教育出版社依法对本书享有专有出版权。任何未经许可的复制、销售行为均违反《中华人民共和国著作权法》，其为人将承担相应的民事责任和行政责任；构成犯罪的，将被依法追究刑事责任。为了维护市场秩序，保护读者的合法权益，避免读者误用盗版书造成不良后果，我社将配合行政执法部门和司法机关对违法犯罪的单位和个人进行严厉打击。社会各界人士如发现上述侵权行为，希望及时举报，本社将奖励举报有功人员。

反盗版举报电话 (010)58581897 58582371 58581879

反盗版举报传真 (010)82086060

反盗版举报邮箱 dd@hep.com.cn

通信地址 北京市西城区德外大街4号 高等教育出版社法务部

邮政编码 100120

目 录

第一章 数据库介绍、数据收集..... 1	第三章 描述统计 44
第一节 原理介绍 1	第一节 原理介绍 44
一、数据的来源 1	一、数据整理 44
二、数据收集的原则 1	二、描述统计分析 44
三、数据预处理的原则 2	三、结果输出 45
第二节 方法与步骤 2	第二节 基本操作指南 46
一、数据库与数据的搜集 2	一、Excel 基本操作指南..... 46
二、数据的整理 9	二、SPSS 基本操作指南 49
第三节 案例介绍 15	第三节 数据的基本描述统计 50
案例一 中国健康与养老追踪调查 (CHARLS) 15	一、数据整理 51
案例二 农业保险市场数据的搜集与处理..... 16	二、频数分析 55
文献阅读 21	三、描述分析 61
习题 21	四、探索分析 64
第二章 统计分析软件介绍 22	第四节 统计图形的选择和实现 69
第一节 SPSS 软件 22	一、整体分析图 70
一、数据文件的操作与建立 22	二、各部分分析图 73
二、数据的录入与保存 26	文献阅读 74
第二节 Stata 29	习题 75
一、Stata 软件 29	第四章 参数估计 78
二、Stata 的界面 30	第一节 原理介绍 78
三、Stata 语法的基本格式 32	一、点估计 78
四、Stata 的数据输入与保存 32	二、区间估计 79
五、Stata 的帮助功能 33	第二节 方法与步骤 79
第三节 Eviews 34	一、点估计的 SPSS 实现 79
一、Eviews 简介 34	二、区间估计的实现 82
二、运行 Eviews 34	第三节 案例介绍 89
三、Eviews 的主要功能 36	案例一 快餐平均满意度的置信区间 89
四、退出 Eviews 36	案例二 以单样本 T 检验过程和探索过程 求均值的置信区间 91
五、案例 37	文献阅读 94
文献阅读 43	习题 94
习题 43	第五章 假设检验 96

第一节 原理介绍	96	案例二 税收增长的原因	147
一、原假设(零假设)和备择假设	96	案例三 居民储蓄	150
二、拒绝域和临界点	96	文献阅读	152
三、第一类错误和第二类错误	96	习题	152
四、双边检验和单边检验	96	第八章 因子分析	154
第二节 方法与步骤	97	第一节 原理介绍	154
一、假设检验的 Excel 实现	97	第二节 方法与步骤	156
二、假设检验的 SPSS 实现	99	一、基本步骤	156
第三节 案例介绍	100	二、因子分析的基本操作	156
案例一 独立样本 T 检验	100	第三节 案例介绍	159
案例二 正态分布假设检验	103	一、操作步骤	159
文献阅读	105	二、结果解读	160
习题	105	文献阅读	163
第六章 方差分析	106	习题	164
第一节 原理介绍	106	第九章 主成分分析	165
一、基本思想	106	第一节 原理介绍	165
二、方差分析的应用条件	107	第二节 方法与步骤	166
第二节 方法与步骤	107	第三节 案例介绍	166
一、单因素方差分析	107	一、利用 SPSS 进行因子分析	166
二、双因素方差分析	112	二、利用因子分析的结果进行主成分分析	167
第三节 案例介绍	122	文献阅读	169
案例一 我国城乡居民消费支出结构差异	122	习题	170
分析	122	第十章 时间序列分析	172
案例二 公众气象服务满意度差异及影响	126	第一节 原理介绍	172
因素分析	126	一、概念介绍	172
文献阅读	129	二、时间序列的几种预测方法	172
习题	130	三、预测方法的评估	173
附录 调查问卷	131	第二节 方法与步骤	173
第七章 相关与回归分析	132	一、Excel 计算方法与步骤	173
第一节 原理介绍	132	二、SPSS 计算方法与步骤	189
一、相关分析	132	第三节 案例介绍	199
二、回归分析	133	案例一 社会消费品零售总额预测	199
第二节 方法与步骤	135	案例二 基于指数平滑法的能源需求预测	203
一、相关分析	135	分析	203
二、回归分析	136	案例三 基于季节调整分析的我国蔬菜及	207
第三节 案例介绍	143	水果出口量预测	207
案例一 文化素质与经济发展水平的相关	143	文献阅读	211
分析	143	习题	211

第一章 数据库介绍、数据收集

第一节 原理介绍

研究统计学是通过搜索、整理、分析数据等手段,以达到推断所测对象的本质,甚至预测对象未来的一门综合性科学。数据是统计学的研究处理对象,因此数据的收集工作尤为重要。

一、数据的来源

数据的来源包括直接来源与间接来源,无论数据来源于何处,一定要注意统计口径的一致,以及数据的真实性和可靠性。有些人以为一手数据更加可靠,实际上,如果有权威的二手数据而且能够满足研究要求,应该使用二手数据,经济上可行而且更加使别人能够信服。往往因为二手数据无法满足我们对于变量的要求,这时候才需要进行实验或者调查获得一手数据。这时获得数据的每一个步骤都要小心翼翼,才能保证最终的结果真实可靠。

如果使用二手数据,研究者应该熟知自己研究领域拥有的数据库资源。例如国家层面的数据库资源包括哪些,行业层面的数据库包括哪些,如何知道宏观经济数据。除了政府提供的数据库之外,还有很多权威的研究机构进行过比较大规模的数据调查,比如中国科学院农业政策研究中心、北京大学国家发展研究院等,这些学者或者机构采集的数据也有很多公开供学术界使用,为做微观数据研究的学者提供了一笔宝贵的财富。而且现在已经通过网络把这些资源进行了整合。

二、数据收集的原则

问题和目标明确后,接下来要考虑的是进行研究所需要的数据。因此,数据的搜集应该根据研究目标和研究问题进行。仔细梳理分析框架,做到数据搜集有的放矢,以最低的成本获得所需要的可行的有质量保障的数据。

这里面特别强调的是根据你的研究目标来搜集数据,而不是本末倒置,先拿到数据然后提炼问题。对于既有研究问题的验证,如果有可用的二手数据,应该根据研究思路和框架仔细思考是否自己希望的关键变量都能包含,对于非关键变量是否可以找到替代的变量。这个数据是否能验证支持自己的假说,如果不能是什么原因。而不要根据数据的结果来编造一个故事。如果没有二手数据可以使用,那么就需要进行调查或者实验。

在问卷调查中,应该注意:第一,根据研究目标设计变量,然后再设计问卷。问卷设计好后,需要进行预调查来检验问卷是否有歧义;第二,问卷设计时,避免大家答案都一样的问题;第三,问卷设计时,尽量不要有开放式的问题,如果有问及原因或者主观感受的东西,也要事先了解大概的问题范围,变成选项请受访者选择,这样做的好处是便于统计分析。否则开放式的问题就变成了无用的问题。第四,问卷设计时注意数据录入的问题,因此要做好问卷的编码,并且尽量少设计多选题,所有的选项用 1、2、3、4,而不是 A\B\C\D;第五,问卷设计时,尽量客观,少用主观选项,因为问卷的问题往往会诱导受访者回答;第六,调查方案中一定注意抽样的问题。因为抽样会对你的数据有非常大的影响,受到资源的约束(包括财力、精力和政治资源)我们往往很难做到完全的随机。这时候应该是尽量地重视自己抽取样本的质量,尽可能地设计好经济合理科学的抽样方法。否则,利用有偏误的样本得到的总体,会给决策带来很大的误导,最终变成一种自娱自乐。最后,实施调查时,一定注意调查员的培训,不仅要讲解问卷,更重要的是讲解自己设计的思路。调查结束后,要注意数据的检查,逻辑的检验。

三、数据预处理的原则

在数据的整理方面,值得一提的是,永远要保留纸质问卷和原始资料。另外数据的初查和整理,会让研究者少走弯路。如果不进行数据的筛查、整理和初步的计算,而直接进行分析,往往会得出错误的结论。

因此数据拿到后,首先要大致浏览,进行逻辑上的判断,看看可能会出现什么问题,核对问卷和数据,然后再通过数据排序、筛选等方式进行数据的检查,并利用 Excel 等工具进行效度、信度的检查。

有时候我们会需要进行初步的计算整理,例如根据样本的各项收入算出总收入、纯收入。原始的表格一定要保留,以备以后查看。对于根据变量进行整理、精简后的用于模型的数据,可能更加方便分析、整理、绘图等。这时记住自己所作的变换。

第二节 方法与步骤

数据的搜集在抽样方面会使用到 Excel 工具或者其他的统计、计量软件,主要是基于随机抽样中随机数的产生,而很多方面都不涉及软件的使用问题。因此,这里不介绍软件层面的操作问题。

一、数据库与数据的搜集

在生产生活当中,我们经常要搜集数据,比如市场营销团队要评估营销方案;药物制造商要确定一种新药是否比现用药物疗效更好;老师要了解学生学习某门课程的情况;国家要了解当前的人口数量以及结构状况等。一般而言,搜集整体中每个数据是比较困难的,耗时耗力。所以,大多数情况下我们是从样本中搜集数据的。

依据研究目的搜集数据是统计工作的第一步骤,因为如果搜集的数据中有偏差、模糊不清或有其他的错误时,即使使用最复杂的统计方法,也无法得出有用的信息。统计数据主要来源于两种渠道:一种是来源于别人的调查或者实验数据,称为间接数据或者是二手数据;另一种来源于科学实验、观察研究或者直接调查的数据,称为直接数据或者一手数据。

1. 间接数据的搜集

间接统计数据主要是公开出版或者报道的数据,可以通过年鉴、期刊、报纸、广播、电视等途径搜集获取。间接资料比较容易得到,相对来说比较便宜,并能很快地获取。有些间接数据,例如由国家统计局普查结果所提供的数据,是不可能由任何一个调查公司按原始数据去收集的。

尽管间接数据不可能提供特定调研问题所需的全部答案,但间接数据在许多方面都是很有用的。二手数据有一定局限性和缺点,由于二手数据是为其他目的而不是为手边的问题而收集的,因此,二手数据对当前问题的帮助在一些重要方面是有缺陷的。另外,二手数据的相关性可能不够,收集二手数据的目的、性质和方法不一定适合当前的情况。而且,二手数据也可能缺乏准确性,或者有些过时了。因此,在使用二手数据之前,有必要先对二手数据进行评价。

在网络广泛普及的今天,通过网络搜集间接数据已经成为了一种主要手段。一般,在网络上搜集数据可以采取两种方式。第一种方式是直接进入专业数据库网站查询数据。比如,要搜集江苏省南京市 2009 年全市地区生产总值,可以进入网址 <http://www.njtj.gov.cn/>,进入该局统计局的主页面,如图 1-1 所示。



图 1-1 南京统计局网站的部分主页面

寻找其中的统计年鉴并点击,如图 1-2 所示。

进入年鉴页面并选择 2010 年统计年鉴,如图 1-3 所示。



图 1-2 操作步骤 1



图 1-3 操作步骤 2

进入 2010 年南京统计年鉴,并点击国民经济核算,如图 1-4 所示。

点击进入全市地区生产总值(2009 年),如图 1-5 所示。

得到 2009 年的全市地区生产总值以及各产业数据,数据表如图 1-6 所示。



图 1-4 操作步骤 3

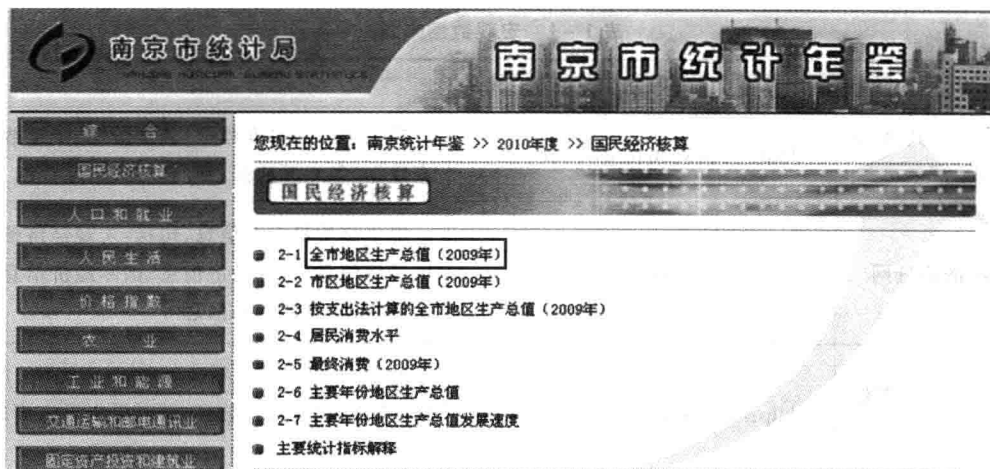


图 1-5 操作步骤 4

全市地区生产总值(2009 年)			计量单位:亿元
指 标	2009 年	2009 年为上年% (按可比价计算)	占地区生产总值 比重%
地区生产总值	4230.26	111.5	100.0
第一产业	129.18	104.1	3.1
第二产业	1930.66	110.1	45.6
工业	1640.53	109.3	38.8
建筑业	290.13	116.0	6.8

图 1-6 搜集到的数据表结果

第二种通过网络获取二手数据的方式是使用搜索引擎。比如,我们可以使用百度网站,搜索江苏省南京市 2009 年经济数据。在浏览器地址栏输入 www.baidu.com,回车即进入百度搜索网站,在搜索框中输入“南京市 2009 年全市地区生产总值”,单击“百度一下”按钮,可以在搜索显示出的结果中,选择链接进入查看符合自己的搜索目标。

可以看到搜索引擎的结果会指向第一种方法的专业数据库网站。但是除此以外,它还会搜索到很多相关的结果。在间接数据的搜集上,如果目标明确而且能找到对应专业数据库网站的,就没有必要使用搜索引擎,因为它可能提供一些无法判别其真实可靠性的结果。当然,如果目标并不明确,或者专业的数据库网站没有直接提供的数据,可以利用搜索引擎去搜索别人类似的二手数据,再通过其他手段去确认它的真实可靠性。

表 1-1 和表 1-2 整理了经济数据中微观数据和宏观数据常用的网络资源。有很多数据库都是比较专业的,大部分都是公开免费使用的。

表 1-1 宏观数据来源

网站名称	网址	简介
中华人民共和国国家统计局	http://www.stats.gov.cn/	提供全国的月度数据、季度数据、年度数据、普查数据、各专题数据、部门数据、国际数据
中国经济信息网	http://www.cei.gov.cn/	提供高质量专业信息,包括:宏观经济信息、地区经济信息、行业经济信息、国际经济信息、专业统计数据、企业产品信息等
中国宏观经济信息网	http://www.macrochina.com.cn/	涵盖 20 世纪 90 年代以来宏观经济、产业经济、投资消费、统计数字、研究报告等方面的详尽内容
华通数据中心	http://data.acmr.com.cn/	集宏观数据、行业数据、企业数据、普查数据于一体,建立从宏观到微观的一系列商用信息库

表 1-2 微观数据来源

网站名称	网址	简介
中国调查数据网	http://www.chinasurveycenter.org/csdn_cn/	致力于改变中国微观调查数据零散和不易获得的现状,建立一个基于互联网的数据共享平台,以便更多学者能够使用微观调查数据进行实证研究
中国健康与养老追踪调查	http://charls.ccer.edu.cn/zh-CN	旨在收集一套代表中国 45 岁及以上中老年人家庭和个人的高质量微观数据,用以分析我国人口老龄化问题,推动老龄化问题的跨学科研究
中国健康与营养调查	http://www.cpc.unc.edu/projects/china	搜集 1989 年之后中国农村和城镇的营养与健康数据

2. 直接数据的搜集

直接统计数据可以通过三种途径获得:科学实验、观察研究或者直接调查。科学实验往往要使用复杂的统计程序去设计实验来获得和核实实验结果,本书不过多地涉及这方面的知识。观察研究,往往是心理学领域或其他相关领域常用的方法。这里主要是指社会经济类数据的调查方法。直接调查是通过设计调查问题的方式,来获取与问题相对应的数据。直接调查包括普查、重点调查、抽样调查、统计报表等方式。

下面主要介绍抽样调查取得直接数据的方法。调查步骤如图 1-7 所示。

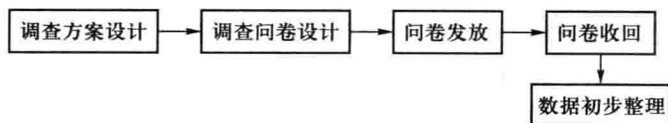


图 1-7 调查步骤

设计一份问卷包括一系列逻辑步骤,如图 1-8 所示。

调查问卷设计的方法与过程:

(1) 问卷设计概述。调查问卷,又称调查表,是调查者根据一定的调查目的精心设计的一份调查表格,是现代社会用于收集资料的一种最为普遍的工具。按照不同的分类标准,可将调查问卷分成不同的类型。

根据市场调查中填写问卷的方式的不同,可将调查问卷分成自填式问卷和访问式问卷两大类。所谓自填式问卷,是指由调查者将问卷发给(或邮寄给)被调查者,由被调查者自己填写的问卷;而自填式问卷由于可以借助于视觉功能,在问题的制作上可更加详尽、全面。而访问式问卷则是由调查者按照事先设计好的问卷或问卷提纲向被调查者提问,然后根据被调查者的回答

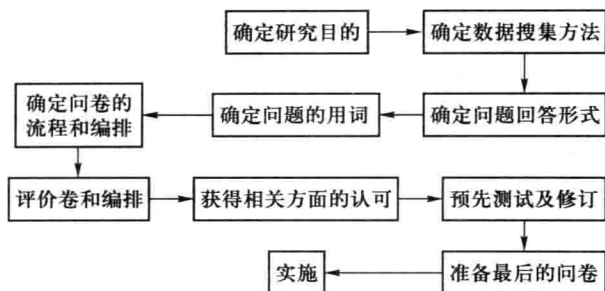


图 1-8 问卷设计步骤

进行填写的问卷。一般而言,访问式问卷要求简便,最好采用两项选择题形式。

根据问卷发放方式的不同,可将调查问卷分为送发式问卷、邮寄式问卷、报刊式问卷、人员访问式问卷、电话访问式问卷和网上访问式问卷六种。其中前三类大致可以划归自填式问卷范畴,后三类则属于访问式问卷。①送发式问卷就是由调查者将调查问卷发送给选定的被调查者,待被调查者填答完毕后再统一收回。②邮寄式问卷是通过邮局将事先设计好的问卷邮寄给选定的被调查者,并要求被调查者按规定填写后回寄给调查者。邮寄式问卷的匿名性较好,但回收率较低。③报刊式问卷是随报刊发送的问卷,并要求报刊读者对问题如实作答后回寄给报刊编辑部。报刊式问卷有传递渠道稳定、匿名性好,费用省等优点,因此有很大的适用性,缺点也是回收率不高。④人员访问式问卷是由调查者按照事先设计好的调查提纲或调查问卷对被调查者提问,然后根据被调查者的口头回答填写问卷。人员访问式问卷的回收率高,也便于设计一些可深入讨论的问题,但不便于涉及敏感性问题。⑤电话访问式问卷就是通过电话对被调查者进行访问调查的问卷。此种问卷要求简单明了,在问卷设计上要充分考虑四个因素:通话时间限制、听觉功能的局限性、记忆的规律、记录的需要。电话访问式问卷一般应用于问题相对简单明确,但需及时得到调查结果的调查项目。⑥网上访问式问卷是在互联网上制作,并通过互联网来进行调查的问卷。此种问卷不受时间、空间限制,便于获得大量信息,特别是对敏感性问题,相对而言更容易获得满意的答案。

(2) 问卷的基本结构。

① 问卷的基本要求。一份好的问卷调查表应能从形式和内容两个方面同时取胜。从形式上看,要求版面整齐、美观、便于阅读和作答,这是总体上的要求,具体的版式设计、版面风格与版面要求,这里暂不赘述。从内容上看,一份好的问卷调查表至少应该满足以下几方面的要求:问题具体、表述清楚、重点突出、整体结构好;确保问卷能完成调查任务与目的;调查问卷明确正确的政治方向,把握正确的舆论导向,注意对群众可能造成的影响;便于统计整理。

② 问卷的基本结构。问卷的基本结构一般包括四个部分,即说明信、调查内容、编码和结束语。其中调查内容是问卷的核心部分,是每一份问卷都必不可少的内容,而其他部分则根据设计者需要可取可舍。

a. 说明信。说明信是调查者给被调查者写的简短的说明文字,主要说明调查的目的、意义、

选择方法以及填答说明等,一般放在问卷的开头。

b. 调查内容。问卷的调查内容主要包括各类问题,问题的回答方式及其指导语,这是调查问卷的主体,也是问卷设计的主要内容。

问卷中的问答题,从形式上看,可分为开放式、封闭式和混合型三大类。开放式问答题只提问题,不给具体答案,要求被调查者根据自己的实际情况自由作答。封闭式问答题则既提问题,又给出若干答案,被调查者只需在选中的答案中打“√”即可。混合型问答题,又称半封闭式问答题,是在采用封闭式问答题的同时,最后再附上一项开放式问题。

至于指导语,也就是填答说明,是指用来指导被调查者填答问题的各种解释和说明。

c. 编码。编码一般应用于大规模的问卷调查。因为在大规模问卷调查中,调查资料的统计汇总工作十分繁重,借助编码技术和计算机,可大大简化这一工作。

编码是将调查问卷中的调查项目以及备选答案给予统一设计的代码。编码既可以在问卷设计的同时就设计好,也可以等调查工作完成以后再进行。前者称为预编码,后者称为后编码。在实际调查中,常采用预编码。

d. 结束语。结束语一般放在问卷的最后面,用来简短地对被调查者的合作表示感谢,也可征询一下被调查者对问卷设计和问卷调查本身的看法和感受。

③ 问卷调查的实施。

a. 预调查。在进行正规的大规模的调查之前,最好进行预调查。预调查可以发现问卷的问题及时进行修正,也可以对一些问题进行演练,增加培训时大家对于问题的统一认识。也有助于问卷设计者增加对问卷数据的主观认识。

b. 调查员的培训。大规模的调查往往不能由一个人完成,一般要聘请一定数量的调查员。由于调查员不是问卷的设计者,不一定能从问卷中全部领会调查者需要重点获得的信息。因此需要对调查员进行培训。通过研究框架讲解、问卷题目讲解和角色演练等方式,让调查员充分理解问卷问题,尽量减少调查员的偏误。

c. 问卷调查质量的控制。在调查实施的过程中,应该在调查前、调查中和调查后进行问卷质量的控制。调查前主要是通过问卷设计、预调查和调查员培训进行质量控制,并事先对调查的单位进行前期的联络工作。调查中,负责人应该正确地组织好调查,最好安排有经验的人作为组长。同时,及时发现问题,每天调查任务结束后最好进行问卷的自查和互查,发现问题及时修正。整个问卷调查结束后,在数据录入前,也要进行问卷的总体检查,发现遗漏、空白选项进行核实,并尽可能地补充信息。

二、数据的整理

1. 数据编码

无论是 Excel, 还是 Stata、SPSS、Eviews 等计量软件,对于数据的处理能力强,对于文字处理起来不是很方便,因此我们进行数据录入时,经常需要对文字进行编码。最常见的编码是调查地址的编码。为了数据录入的快捷,也为了计算机识别的便利,我们往往需要用一串数字代表调查的