

中国人民大学统计咨询研究中心
中国人民大学概率论与数理统计研究所
教育部重点科研基地应用统计科学研究中心

联合推出

数 据 分 析 系 列 教 材

市场研究定量分析

(第二版)

简 明 编著



 中国人民大学出版社

中国人民大学统计咨询研究中心
中国人民大学概率论与数理统计研究所
教育部重点科研基地应用统计科学研究中心

数 据 分 析 系 列 教 材

市场研究定量分析

(第二版)

简 明 编著

中国人民大学出版社
· 北京 ·

图书在版编目 (CIP) 数据

市场研究定量分析/简明编著. — 2 版. — 北京: 中国人民大学出版社, 2015. 1
数据分析系列教材

ISBN 978-7-300-20470-3

I. ①市… II. ①简… III. ①市场学-定量分析-高等学校-教材 IV. ①F713.52

中国版本图书馆 CIP 数据核字 (2014) 第 302304 号



数据分析系列教材

市场研究定量分析 (第二版)

简 明 编著

Shichang Yanjiu Dingliang Fenxi

出版发行 中国人民大学出版社

社 址 北京中关村大街 31 号

电 话 010-62511242 (总编室)

010-82501766 (邮购部)

010-62515195 (发行公司)

网 址 <http://www.crup.com.cn>

<http://www.ttrnet.com> (人大教研网)

经 销 新华书店

印 刷 中煤涿州制图印刷厂北京分厂

规 格 185 mm×260 mm 16 开本

印 张 18.75 插页 1

字 数 400 000

邮政编码 100080

010-62511770 (质管部)

010-62514148 (门市部)

010-62515275 (盗版举报)

版 次 2009 年 5 月第 1 版

2015 年 1 月第 2 版

印 次 2015 年 1 月第 1 次印刷

定 价 35.00 元

版权所有 侵权必究

印装差错 负责调换

数 据 分 析 系 列 教 材 编 委 会

编委会主任 易丹辉

编委会委员（按姓氏笔画排序）

王惠文 吴喜之 张 波

易丹辉 柯惠新 耿 直

黄登源 谢邦昌

总 序

随着社会经济的不断发展、科学技术的不断进步，统计方法越来越成为人们必不可少的工具和手段。在多年教学过程中，老师们也越来越感到运用统计方法解决实际问题的必要，不少人在探索如何运用统计软件介绍和学习统计方法，思考如何运用这些方法解决实际问题。

自 2008 年数据分析系列教材出版以来，得到了不少读者的关注，这次丛书在原来的基础上，根据几位老师教学和科研的实践经验，重新策划完成。丛书特别邀请北京航空航天大学王惠文教授加盟，担任《描述统计》一书的撰写。她多年运用描述统计方法解决实际问题，积累了丰富的经验，将为读者正确运用描述统计方法提供参考。由于几位年轻老师不辞辛苦，这套数据分析系列教材在原来的基础上作了调整，更适合读者的实际需要。

在我们看来，掌握统计方法不仅要理论上弄明白，更重要的在于能够正确有效地运用这些方法，分析说明实际问题。这套丛书正是试图利用实际数据，通过统计软件的实际操作，对能够使用的统计方法加以说明，使读者不仅能够了解相应的统计方法，而且能够通过计算机操作学会运用这些方法处理分析实际数据。希望这套丛书的出版能够为读者提供这样学习的工具。

由于水平有限，难免有不足之处。恳请读者朋友们提出宝贵意见。我们也会循着这样的思路，在教学以及和读者的交流沟通中不断积累、不断提高、不断完善，奉献给读者更多、更好的成果。

感谢为这套丛书的编写付出汗水的研究生，感谢几位认真用心的年轻老师，感谢中国人民大学出版社的大力支持，特别是陈永凤编辑和王伟娟编辑，是她们的努力，让这套丛书在很短的时间里出版。为方便读者，书中的所有例题数据，都将放在中国人民大学出版社的网站上，供读者下载并练习。谢谢读者，希望在数据采集变得越来越容易、大数据时代到来的今天，能够加强沟通和联系，为提高统计方法实际运用的能力和水平共同努力。

易丹辉

前 言

信息技术和互联网的发展为市场分析提供了极其丰富的数据,也对企业的经营管理提出了新的挑战。随着市场竞争形势的改变,越来越多的企业认识到精确营销的重要性,而实现精确营销目标的前提是对市场做出精准的认识和判断,因此市场研究也就越来越为现代企业所依赖。统计分析是认识事物的重要工具,一旦有效地运用于市场营销就可为营销决策者提供精准指导。本书着重介绍市场研究中的定量分析方法和技术,系统地展现了统计分析技术在市场营销中的具体应用。

按照理论与实践相结合、市场研究为企业经营管理服务的要求,本书主要体现以下几个特点:

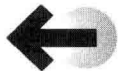
(1) 具有较强的可读性。本书选择市场研究中常用的方法,并予以详细介绍。本书除了第1章从定性角度介绍市场研究中的数据以外,其余各章都介绍市场研究中经常用到的方法,可读性较强。

(2) 具有较强的应用性。本书将市场研究方法与实务相结合,在详细介绍市场研究方法的同时,也兼顾了实践应用。选取一些调研实务数据,将研究方法与实务数据相结合,构建边学边用的方法体系。

(3) 操作性较强。本书将研究方法与计算机相结合,市场研究数据集都很大,借助计算机进行数据分析在市场研究中必不可少。书中介绍了在市场研究中每一种方法的SPSS统计软件操作过程,以便读者自学。

同时各章配有习题,我们在选择习题的过程中注意新颖和实用,习题数据很多是取自编写者近期主持的市场研究项目,以引导读者将定量分析方法应用于市场营销实践,使学生通过练习加深对方法的理解。另外,书中的例题和习题所涉及的数据可以到中国人民大学出版社工商管理分社的网站下载(www.rdjg.com.cn)。

本书可以作为统计、营销、经济管理和财政金融等专业学生学习市场研究定量分析方法的教材,也可以作为市场研究人员、咨询师、企业营销和经营管理人员等的参



考工具书。

本书得以出版要特别感谢教育部重点科研基地中国人民大学应用统计科学研究中心提供的研究条件，还要感谢中国人民大学出版社的编辑为本书付出的辛劳。

本书也包含着编写者多届研究生的贡献，他们在参加讨论、收集资料、分析案例中都体现出了热情和专业素养。2013 届的陈嘉明、王子、李丽、陈丽彬，2014 届的张芸霞参与了本次修订工作。

尽管在编写过程中我们做了很大的努力，希望能让读者满意，但是肯定还会有很多缺陷，热切希望同行专家和广大读者提出宝贵的意见，这些意见将是促使我们进一步研究和写作的动力，也是推动市场研究方法不断进步的动力。

目 录

第 1 章 市场研究中的数据	1
1.1 数据的测度和分类	1
1.2 市场研究数据的来源	3
1.3 缺失数据的处理	7
第 2 章 描述性分析	10
2.1 数据的集中趋势分析	10
2.2 数据的离散趋势分析	16
2.3 数据分布的偏度与峰度	22
2.4 描述性分析的应用及 SPSS 软件操作	23
习 题	25
第 3 章 数据估计与检验	26
3.1 参数估计	26
3.2 参数检验	27
3.3 非参数检验	36
习 题	51
第 4 章 相关性分析	52
4.1 相关分析概述	52
4.2 相关分析	58
4.3 偏相关分析	73



4.4 关联规则	76
4.5 相关分析的 SPSS 软件操作范例	80
习 题	83
第 5 章 线性回归分析	85
5.1 问题的提出	85
5.2 线性回归分析理论原理	86
5.3 线性回归分析的应用	93
附录: 公式推导	107
习 题	109
第 6 章 Logistic 回归分析	110
6.1 Logistic 回归分析概述	110
6.2 Logistic 回归理论介绍	111
6.3 Logistic 回归分析实际应用	113
附录: 公式推导	118
习 题	120
第 7 章 分类分析	121
7.1 聚类分析	121
7.2 判别分析	140
7.3 决策树分析	150
习 题	154
第 8 章 数据简化分析	155
8.1 主成分分析	155
8.2 因子分析	168
8.3 对应分析	181
习 题	195
第 9 章 多维尺度分析	197
9.1 多维尺度分析的基本理论方法	197
9.2 多维尺度分析的实际应用	199
习 题	210



第 10 章 对数线性模型	211
10.1 基本思想和理论	211
10.2 对数线性模型的应用	214
习 题	229
第 11 章 路径分析	230
11.1 路径分析概述	230
11.2 路径分析的概念和理论	232
11.3 路径分析的应用	236
习 题	263
第 12 章 时间序列分析	265
12.1 常见时间序列模型简介	265
12.2 模型预测误差	287
习 题	289

1.1 数据的测度和分类

一、数据的测量尺度

市场研究过程中有各种各样的数据，我们发现这些数据有不同的测量尺度。数据的测量尺度是指测量某一物体的准则或者依据。比如测量门的高度时，我们可以用米尺，也可以用掌距来测量。测量尺度是为了更好地反映事物的特征，更好地表述、记录以使我们的研究变得方便。一般来说，市场研究中一般有下列四种数据的测量尺度。

1. 名义尺度

名义尺度是按照事物的特征或者属性的不同，赋予不同的名称，作为该事物的一种标记，从而可以区别于其他事物，而且具有相同属性或者特征的数据可以归为同一类，因此名义尺度也可以称为“类别尺度”。例如，人的性别可以区分为男性和女性；婚姻状况可以区分为已婚和未婚，这些都是应用名义尺度来命名的数据。所以在用名义尺度测量或者描述事物的特征时，先将该事物按照其特征进行归类，并表示类别的名称，然后用代码进行表示。

2. 顺序尺度

顺序尺度是将事物按照其特征和属性排成顺序或者等级。比如将中国家电企业销售量最高的 5 家公司排名为 1~5。顺序尺度的主要功能是用来进行等级排列，比较先后的顺序。在等级或者顺序的排列中，可以比较个体之间的地位，说明“大于”或“小于”的关系和差异，但是个体之间的差异并无相同的单位。例如一次考试中，全



班第一名和第二名的成绩差异，未必等于第二名与第三名之间的差异。

3. 等距尺度

等距尺度是一组具有连续性、单位相等的数值。这些数值不仅显示大小顺序，而且数值之间具有相同的距离。例如，以等距尺度测量一个班的英语考试成绩，范围在 0~100 分之间，根据各个学生回答问题的情况给予一定的分数，从学生的分数当中可以看出他们成绩的高低，而且可以了解到学生之间成绩的差距。因此等距尺度的主要特征在于采用连续且等距的数据说明变量特征或属性的差异性。

4. 比率尺度

比率尺度具有等距尺度的全部特征，而且具有“真正零点”。比率尺度的数值之间有相等的比率，不仅可以加减，而且可以作乘除运算。例如，人的身高，可以采用比率尺度来测量，以 0 代表没有高度，0 以上的不同数值代表实际高度，身高 200 厘米是身高 100 厘米的两倍。年龄也可以采用比率尺度来测量，零岁是真正的零点。因此，比率尺度提供事物的信息量最多，作用也最大，但是在实际研究中并不多见。一般来说，物理特征（重量、长度等）的测量比较可能采用比率尺度，但心理特征的测量多以等距尺度为主，因为人类的心理特征很难找到真正的零点。

二、数据的类型

测量某种特性的尺度通常都包含一组项目（items），每一项目都陈述了某种可以引发个人反应的特殊情境或社会对象，同时附有可供个人选择的一组反应类别。按照某种指派规则，对于个人对某一特定项目的反应所评定的分数，称为项目分数（items score）。项目必须经过一道选择的程序，去掉过分冗长、语义含糊不清或可能引起混淆的叙述句，项目经过适当量化后称为变数。

针对问题中各种特性在不同对象间差异的描述，我们必须发展一种客观的测量方法以提供个别性质的描述及进一步的信息。性质可能是形状、大小或态度等的特性，变数用来描述某种性质的个别差异。性质在某些情况下相当具体，但在其他情形，特别是心理变数就不具体了。对于不同对象其性质只有一种，此性质称为常数。如对全人类而言，性别是变数，但只对男性而言，性别则是常数；又如在数学测试上不同学生得到不同的分数，反映在分数上的数学学习能力是变数，但是所有得 42 分的学生的数学学习能力是常数。

市场研究中数据可以分为这样几种类型：按照所采用的测量尺度不同，可以将数据分为定类数据、定序数据和数值型数据；按照数据的收集方法可以分为观测数据和实验数据；按照描述对象与时间的关系，可以将数据分为时间序列数据和截面数据；按照数据的连续性不同，可以将数据分为连续数据和间断数据。

1. 定类数据、定序数据、数值型数据

依据前面的数据的测量尺度，将市场研究中的数据分为定类数据、定序数据和数值型数据三种类型。定类数据是按照事物的特征进行分类的结果。数据表现为类别，



可以用文字来描述,也可以用数字代码来代替。在市场研究中进行数据分析时,会对定类数据进行数据编码。定序数据是根据顺序尺度来进行分类的,它不仅能够表示事物的类别,还能表示类别之间的顺序。例如考试中的优、良、中等;产品的一等品、二等品、三等品等。数值型数据即依据等距尺度和比率尺度等测量尺度对事物进行计量的结果,其结果可以表现为具体的数值,表明事物的数量特征。

2. 观测数据和实验数据

观测数据是由访问员的现场观察而获取的数据,观测数据的内容丰富多样,可以用文字来表示,也可以用声音、图片等来表示。实验数据是由研究人员根据实验的结果而得到的数据,市场研究中通常是由研究人员控制一种变量,而观察另一种变量的反应变化。例如,在北京家乐福商场中,研究人员提高飘柔洗发水的价格,看该洗发水销量的变化。

3. 时间序列数据和截面数据

时间序列数据指的是在不同的时间上收集的数据,它所描述的是某种现象随时间而变化的特征。例如,三元牌牛奶在北京市场 2000—2006 年的市场占有率变化,就是一个典型的时间序列数据。截面数据指的是在相同或者相似的时间点上收集的数据,例如 2006 年各品牌牛奶在北京市场的销售量。

4. 连续数据和间断数据

在对一些变量进行测量时,变量的单位可以再划分成无限多个细小单位。例如身高,单位可以用米,也可以用厘米、毫米,变量的值是连续的。这类变量称为连续变量,其相应的数据称为连续数据。但是对另外一些变量进行测量时,变量的单位不能再划分成无限多个细小的单位,变量值是间断的。如测量学生的心跳次数,只能是 1 次、2 次,而不可能是 1.5 次、2.5 次。这类变量称为间断变量,其相应的数据称为间断数据。间断数据的数字形式只能为整数。

1.2 市场研究数据的来源

统计数据分为原始数据和二手数据。原始数据系指研究者直接由调查或实验而收集得来的数据,亦称一手数据。例如,民意调查机构要预测选举结果,得直接向选民询问调查,所得的数据就是原始数据。原始数据不易收集,有赖于完善的收集技术与筹划,其所花费的人力、时间、金钱很多,收集来的数据比较全面且合用。而二手数据系指其他私人、机关、团体所收集整理后的数据。二手数据是由专家收集的,因此有相当高的正确性,而且这种数据的查询既省时省力又省钱,但是却经常不能完全适合所要研究的问题。

此外,数据的性质包含静态数据与动态数据。静态数据系指在某一特定时间所收集的数据。静态数据通常都由调查而来,其调查方法依照调查的对象是否为整个全体,可分为普查与抽查两类。动态数据系指在某一期间内连续不断地收集得来的数



据。例如,户籍机关每天必须登记人口的出生、死亡、婚姻、迁移等动态统计数据。动态数据多由登记或观测而得。研究某现象的瞬息状态,多采用调查方式来收集原始数据,如普查某年12月31日某城市的总人口;分析某现象在某期间内连续发生的状态,则宜采用登记方式或观测来收集统计数据,如统计某地某年人口的出生与死亡数目。统计某商店某月每天营业额,或观测某地商店消费顾客的特征等,必须经常记录其变动情形。

一、原始数据收集的方法

1. 观察法

观察是一种对已经发生的事件或现象作记录的活动,从对日常生活的观察中,可以得到很多信息。例如,学生对老师上课的反应,老师说话的声调,同学之间对话的声音,闻到的香水味、饭菜的香味等,都可形成信息的基础。观察对象不只是看到的,也可以是听到的、读到的、碰到的、闻到的。观察内容包括历史数据或目前记录的数据,如财务记录、历年的指标、手抄的或录影带式的记录等。另外,面部的表情、讲话的习惯、演讲的内容、声音的频率、音量大小、音质、讲话的速度、说话的间隔及节奏、表达的风格等都可以成为观察的内容。

观察法可以在现场观察到事情的发生,而不必通过受测者的报告或其他人的转述,可以避免受测者对信息的筛选和报告不完全等情况,也可以观察到一些受测者没有记录到的细节。观察法是数据收集方法中限制最少的。但是观察法要求在现场目睹事件的发生,观察者所看到的通常是较为表面的或象征性的现象。利用观察法得到的数据是一种主观评价或记录的数据。应视场合决定观察法是否适用。

例如,我们可以观察到商店内顾客将何种商品放入购物篮,但是只用眼睛看却无法得知顾客是基于何种理由选择了这件商品;我们从外表可知来店中消费的顾客性别或大概的年纪,但是他们住在何处、为什么会来这家店消费等信息,却无法由观察法得知。

如上所述,利用观察法所能得到的资料是从外表上的判断而来,因此会受到限制。

观察法可以分为单独进行的观察法和与其他调查法结合的方式。与其他调查法结合的方式是先以观察法进行初步调查,之后再使用其他调查方式。

使用观察法的时候,必须事先确认是否只用观察的方式就能够达到调查的目的,或得到想要的结果,再进一步确定是否使用其他方式调查。若是没有将调查目的理清,最后很容易落入得到的成果稀少、不知道下一步该如何做的窘境。

以下将说明由观察法能够得到的信息及其做法。

(1) 单独进行的观察法,如进行自动贩卖机的销售情况调查。调查时,可以在设置很多台自动贩卖机的地方派一些调查员,观察使用自动贩卖机的人买了什么种类、多少数量的商品,还有顾客的性别及大概的年纪、从服装上可以判断出来的职业、购买时间,等等,连续观察12~14小时。

(2) 以观察法为基础进行其他的调查。观察法的第二种类型是观察法与其他调查法的组合方式,先以观察法得到线索,再进行下一步调查。

举例而言,考虑新商品的处理方法时,可以在市中心竞争对手的商店观察其卖场中的商品,然后买入主要的用品,并调查商品的制造商或销售处,询问其批发处或进货方法等,这是现在比较盛行的做法。

调查开始之前,需先决定是按想要知道的内容调查还是依循调查目的来调查,这是非常重要的。

2. 实验法

实验法是在控制环境下,操作一个或一个以上的变数,并观察这种干扰会造成何种影响的方法。例如,考虑提高在化学实验中某种产生物的强度,假设在种种条件中,对强度影响较大的因素为反应温度与催化剂含量,我们要找到这两种因素的最佳条件。为了寻求各因素的最佳条件,必须变更各因素的条件来做实验。



例 1—1

某通信公司推出一种新型增值业务,收费标准为 19 元/月。一段时间后欲对收费标准作调整,设定了两种新收费标准,分别为 16 元/月和 21 元/月,公司采取随机抽样选定三个分公司市场作为实验单位,分别执行一种收费标准。19 元/月的分公司市场即为“控制单位”,其他两家则为“实验单位”。至于哪家分公司采取何种收费标准,可用简单随机抽样法决定。实验期间定为每月 1 次,共进行 4 次实验,即 4 个月。实验目的在于推断何种收费标准创造的收益最大,确定一种合理的收费标准,降低日后大量推广的风险。因此市场实验是有意义的。

3. 调查法

通常调查有三种方式可供选择:

(1) 派员调查。即选用专人施以有关训练,然后派出去调查。此种调查方式适用于调查单位多且分布广的场合,可节省时间,但其调查结果的准确度与访问员的素质有密切关系,通常不如亲自调查的结果,且费用较高。

(2) 邮寄调查。即将事先设计好的调查表寄给被调查者,请其自行填报。此种调查方式简便且可节省经费,但被调查者常会有误答、乱填或根本不答的情况,故很难获得圆满的结果。为避免此缺点,调查者除慎拟调查表外,还常利用抽奖或赠纪念品的方式以提高问卷的回收率。如果经费与时间充足,可采取先派人分发问卷而后要求被调查者邮寄回来,或者先邮寄问卷而后派人收回的方式。

(3) 电话询问调查。能在最短时间内得到信息,过程直接而简便,但涉及敏感性或必须深入了解的题材,可能不易获得。

适合电话调查的场合

最近常常听到“电话行销”这个词。电话行销是指使用电话作为行销手段的方法。所谓电话调查,可以实现电话行销之中的一个机能。

我们常在日常生活中见到或是听到电话调查的结果。例如,大家都曾经在报纸等



媒体上看过的内阁或政党支持率、选举结果的预测等。电话调查的好处是可以在短期内在范围广大的地域调查众多的对象。只要通话的人数增加，如果不介意电话费，就有可能在一天内从一地向全国范围内的几千人进行调查。

另外也常常实施购物调查、商圈调查或商品所有状况的调查等。短期内以较低成本便能够收集到很多的回答，这是最大的好处。

了解顾客满意度的调查，也可以借由电话进行，只不过必须减少问题的数量。换句话说，问题数量无法太多是电话调查的不足之处。

常常忘记规则的电话

大家都曾经接过进行问卷调查兼具产品促销的电话吧？对方以“想麻烦做下问卷调查”的名义开始，不知什么时候就变成商品简介及约定见面时间了。遭拒后对方能立即作罢，还算是不错的，但是仍有强行约定见面时间的情况发生，令人非常为难。毫不惭愧地进行这些活动的公司，让一般消费者对电话调查产生不信任感，这实在是件憾事。

电话调查是公司或商店希望能达到“对顾客有帮助，让顾客更加高兴”的目的所实施的，因此对社会有贡献的部分也不少。但是另一方面，对于违背这个意义进行调查的企业，则令人感到相当愤怒。

再者，因为使用电话不用看到对方的脸，就可以轻易侵入他人的生活，所以一不注意，便可能会有意想不到的状况发生。通电话时，为了表达诚意、不失礼，一开始说出自己的姓名及目的，取得对方的信任后再进行访问，这是最基本的礼节。进行问卷调查的时候，当然也是要取得对方的许可，对方没有同意受访，单方面强硬地进行调查，是违反规则的。

不适合电话调查的情况

那么，什么样的情况不适合电话调查呢？譬如说，一定要观察，否则无法了解的状况，只用电话当然是无法明白的。店面卖场陈列了什么样的商品，不实际看一下卖场，就没有办法掌握。

同样，调查内容是要根据图、照片或商品来评断的情况，也不可能利用电话进行访问。再者，访问的内容复杂，或是问题很多的时候，中途被拒绝的机会就会增加。问题数以15~20题为限。但是这个数量适用于习惯受访的人，初次访谈的对象，最好是这个题数的一半左右。

设计问卷的时候，调查者常常会这个也想知道，那个也想问问，这往往是失败的根源。受访者答应接受调查，照理会予以配合，但是如果刚好在最忙碌的时候突然接到电话，被问了一堆细琐的事项，任谁也不会有耐心，更遑论用心回答。另外，不集中项目问完的结果是易失去效率，内容的正确性也会有问题。

总之，请记住，电话调查的访问项目越是简单、少量，回收率就越高。

二、二手数据

二手数据是在原始数据的基础上经过加工整理而得到的，它具有数据量大、收集

容易、内容丰富等特点,尤其是目前互联网比较流行的情况下,很多数据都是以电子数据库的形式存在的,收集二手数据就更加方便了。

1. 二手数据的收集方法

(1) 利用公开的二手资料。二手数据量非常大,而且比较复杂,每年政府都会发布大量行业方面的年报,而且企业也会公布大量的数据,各大市场研究公司发布的研究报告,行业协会发布的年报等都可以成为二手数据收集的途径。

(2) 利用公司内部的数据库。许多公司建有内部的数据库,这些数据直接记录了该企业的日常销售、运行等方面的数据,可以直接提取出来运用。

(3) 利用互联网。利用互联网收集二手数据是一种非常流行的方式,利用互联网收集数据成本低、见效快,而且互联网能够共享大量的信息资源,所以应该充分利用好这个工具。

2. 二手数据的特点

(1) 收集时间短、费用低。二手数据可以通过政府发布、企业年报及新闻媒介等途径获得,尤其是许多官方公开的数据不仅具有权威性,而且是免费的,能够通过发达的网络快速收集到,所以相对于一手的调查数据,二手数据的收集时间比较短,费用也相对低廉。

(2) 二手数据缺乏相关性和准确性。二手数据是根据过去某专一研究目的而收集的,很可能不能直接应用于现在的研究,而且有许多二手数据是没有精确性的,存在较大的错误,如果没有一定的鉴别经验直接拿过来用,是存在较大风险的,这也是二手数据的一个缺陷。

1.3 缺失数据的处理

一、数据缺失的危害和机制

1. 数据缺失的危害

市场调查过程中,各种原因会导致数据的缺失。调查的初始阶段,访问员无法接触到受访者,或者被调查者地址不详,或者已经搬迁,导致无法接触到目标受访者而产生数据缺失;调查阶段受访者拒绝回答某些问题,如关于受访者的收入及其他较为隐私性的问题,这些问题的空白也是数据缺失的表现;在数据整理阶段也会出现数据缺失的现象,研究人员将明显不符合实际的数据剔除。所以即使做到万无一失的控制,仍然会有数据的缺失。倘若是较小规模的缺失,可能对调查结果的影响不太大,如果是大规模的数据缺失,后果就比较严重了,会导致调研结果和真实结果完全不相符。具体来说,数据的缺失主要表现在:第一,增大了估计误差;第二,导致有偏差的区间估计。