

浅谈影响样本容量大小的因素和对调查误差的认识

薛长青

〔摘要〕 抽样调查是企业进行市场调查的主要方式,样本容量的确定是采用抽样调查方式必须首先考虑的问题,影响样本容量确定,既有定性方面又有定量方面的因素,了解这些因素,有助于制定科学的抽样设计;市场调查由于受各方面因素影响,调查结果总会存在误差,既有抽样误差又有非抽样误差,正确认识和评价这些误差,实践中尽可能避免,从而有利于提高市场调查的准确性。

〔关键词〕 样本容量 调查误差

一、影响样本容量的因素

样本容量的大小涉及到调查中所要包括的人数或单元数。确定样本量的数目,既有定性方面的考虑,又有定量方面的考虑。

(一) 定性方面因素

1. 决策的重要性。通常越是重要的决策,就越需要更多的信息,因此就需要较大的样本量。但随着样本量的增加,费用相应上升,并且样本量越大,增加一个调查单位所得到的精度上的增益也就越小,这里,精度与费用的关系也是必须考虑的。

2. 调查本身的性质。一般来说,探索性调查,多采用定性方法,所需样本量较小;结论性调查,如描述性的调查,所需样本量较大。

3. 变量的个数。调查所涉及到的变量较多,样本量就应该大一些,以减少抽样误差的累积效应。

4. 数据分析的性质。当采用多元统计方法对数据进行复杂的高级分析时,就需要较大的样本量。

5. 发生率和回收率。发生率指的是调查所需“合格”人员的比例,发生率决定了对给定的样本量所需的接触次数;完成率是指合格调查对象完成调查的比率。发生率和完成率同时考虑,就包含着实际应接触的调查对象人数应当是所要求样本量的一个新的倍数。

6. 资源限制。市场调查要受到人力、物力和财力的限制。

(二) 定量方面因素

即按统计方法来确定样本量的多少,通常事先规定好精度水平,然后依据公式来计算样本量。在确定具体样本容量时先要明确以下几个因素:

(1) 允许误差。是指根据抽样资料估计总体参数所允许的误差范围,用 ΔX 表示。

(2) 置信度。是指保证抽样误差超过允许范围的一定概率,又称概率保证度,它表明抽样误差落在允许误差范围内的可靠性程度,用 $F(t)$ 表示。

(3) 概率度。在实际调查中,可以根据专门的标准正态

分布表,由给定的置信度 $F(t)$,在表中查出相应的概率度,用 t 表示,此时,

允许误差的计算公式为:

$$\Delta X = t \cdot \mu$$

样本容量的计算公式为:

$$n = \left(\frac{t \cdot \delta}{\Delta X} \right)^2 = \left(\frac{t \cdot s}{\Delta X} \right)^2$$

从公式中可以看出,影响样本容量的因素有三个:

(1) 总体中各单位之间的差异程度。样本标准差 S 是在计算抽样误差(μ)时代替总体标准差(δ)的,而总体标准差(δ)正是用来说明总体各单位之间差异程度的,因此,总体各单位之间差异程度越小,所需样本数就越多。

(2) 概率度 t 的大小。即对抽样误差范围估计可靠程度的大小,可以看出,对可靠程度要求越高,所需样本数就越多,反之,对可靠程度要求越低,所需样本数就越多。

(3) 允许误差 ΔX 的大小。即对抽样误差范围估计精确度需求。对精确度要求越高,即可允许的抽样误差范围越小,需要的样本数就越多;反之,对精确度要求越低,所需样本数就越多。

二、不同性质的调查误差

在实际调查中,由于各方面因素的影响,调查结果总会存在误差。调查误差按性质可以分为两类,一类是抽样误差,另一类是非抽样误差。

(一) 抽样误差

抽样误差是由样本的代表性引起的,样本仅仅是总体的一个缩影,有许多方面与总体不一致,因此用样本信息推断总体信息就必然要带来相应的误差。导致抽样误差的原因主要有以下几个方面:

(1) 抽样方法。各种抽样方法都有对应的误差计算公式,采用不同的抽样方法会导致不同的抽样误差。

(2) 是否重复抽样。重复抽样有可能使同一单位多次被选中,从而增大误差;不重复抽样每一个单位最多只能被选中一次,误差略小于前者。

(3) 总体方差大小。在样本量一定的情况下,总体的差

异越大,抽样误差也越大,反之,总体的差异越小,抽样误差也越小。

(4) 样本量大小。是影响抽样误差的根本因素,样本量越大,抽样误差越小。

(二) 非抽样误差

是指在所有调查活动中所产生的除抽样误差以外的所有误差。其形成原因十分复杂,主要是由于人为的差错造成的,它的计算比较困难,多数情况下只能借助经验进行估计,无法获得精确的判断。通常非抽样误差有以下几类:

(1) 抽样框误差。抽样框是记录总体单位信息的清单,如果抽样框中的信息存在错误,在抽样时就可能产生误差。例如,由调研人员定义的总体与所使用的抽样框隐含的总体之间的变异,常见的如使用电话号码簿作为抽样框代表所有消费者,但实际上有些号码未登记或已迁移,这部分消费者实际并不包括在原定抽样框中。

(2) 不回答误差。不回答是指调查时未能从指定的受访者处获得有效的回答,当样本中的一些被调查对象不回答时,就产生了不回答误差。主要原因有拒访、联系不上、地址错误和受访者不在家等。不回答的结果使得最后的样本在大小与组成上与原始样本不同,不回答误差实质就是原始样本中变量的真正平均值与结果样本(纯样本)的真正平均值之间的差异。

(3) 回答误差。回答误差是由于受访者给出了不正确的回答,或者是由于将回答错误录入或分析不正确而引起的误差,实质就是结果样本(纯样本)中变量的真正平均值与调查结果中得到的观测平均值之间的差异。通常由调研人员、访问员和受访者三方面引起的。

由调研人员引起的误差:

① 代用误差。调研问题所需信息与调研人员所收集的信息之间的偏差,如将希望收集如何选择品牌方面的信息误为对品牌的偏好信息。② 测量误差。调研问题所收集的信息与调研人员所采用的测量过程所生成的信息之间的偏差,如测量消费者的偏好误用了测量概念的量表。③ 总体定义误差。调研问题相关的真正总体与调研定义的总体的偏差,如希望了解某品牌在其使用者心目中的地位,误定义为同类产品所有使用者都要调查。④ 数据分析误差。由问卷的原始数据转换成调查结果时产生的误差,如错用统计方法等。

由访问员引起的误差:

① 回答误差。提问或追问不当引起的误差。② 记录误差。在听、理解、记录被访者答案时因曲解原意而产生的误差。③ 欺骗误差。访员有意伪造部分或全部答案产生的误差。

由被访者引起的误差:

① 不能正确回答误差。由于被访者不熟悉、劳累、厌烦

或忘记问卷内容而引起的误差。② 不愿正确回答误差。由于被访者为了给出一个社会能接受的答案、避免出麻烦或取悦访问员等因素而有意答错造成。

针对上述调查误差,在调查方案设计中应注意总误差的降低,而不必过分关注某种类型误差的大小,如果一味为了降低抽样误差而加大样本量,则忽视了样本增大后同样增加了访问员误差,从而令总体误差增大。

在实际调查中,非抽样误差比抽样误差危害更大,抽样误差是可以计算的,研究表明,在总误差中往往非抽样误差占了主要部分。对于非抽样误差,可以采用以下一些控制措施:

(1) 建立标准化系统。规范调查活动是控制非抽样误差的重要手段,首先要建立标准化系统,包括标准化术语(调查术语和项目术语)和标准化工具(抽样表、问卷、卡片等)。标准化工作的意义在于使调查各环节能够在统一的要求下操作,各环节之间的工作交流和信息交流能够准确、有效。建立一套标准化的操作系统,才有可能保证调查质量的稳定。

(2) 加强文档管理。文档具有信息传递的职能,文档管理属于信息管理活动,调查所获得的数据资料都是信息产品,文档是这些信息产品的载体,调查活动涉及许多文档,如各种设计方案、抽样资料、问卷、访问记录、数据处理记录等。加强文档管理是明确责任的重要手段。

(3) 完善内部计量工作。在调查实施过程中,有大量的内部计量工作需要完成,如常见的一些计量指标包括平均访问时间、拒访率等,这些计量指标一方面是评价访问工作质量的标准,另一方面也是进行调查设计的基础数据。

(4) 建立质量责任制。建立质量责任制是进行质量控制的前提,对于调查活动的各环节的操作规范,只有在相应人员各自都了解自身责任的前提下,才有可能保证实际操作是按控制要求规范进行的。

(5) 搞好质量教育。质量教育是一种积极的质量控制手段,一方面,通过对质量意识的教育,提高调查人员对控制调查质量的意义与要求的认识,加深对各项规范要求的理解,自觉提高工作的责任心;另一方面,加强对相关业务知识的学习,可以进一步提高工作能力与水平。

参 考 文 献

1. 胡玉立、李东贤等,《市场预测与管理决策》,中国人民大学出版社
2. 柯惠新、刘红鹰 编著,《民意调查实务》,中国经济出版社
3. 龚江辉 著,《商业调查实务》,经济科学出版社
(作者单位:兰州商学院工商管理系)