

【统计理论与方法】

敏感性问题的统计调查新方法——网上调查

孙 蕾, 蔡 亮

(中南财经大学 统计系, 湖北 武汉 430064)

摘 要: 网上敏感性问题的调查是随着网络技术的完善而产生的一种新的抽样调查方式。文章对网上敏感性问题的调查的可行性、调查方案、特点及原则作了初步探讨。

关键词: 敏感性问题; 网上调查; 随机 IP 自动拨叫技术

中图分类号: C811 **文献标识码:** A **文章编号:** 1007-3116(2000)02-0043-02

抽样调查技术中有一项独特的调查技术,即敏感性问题的随机化回答技术。所谓敏感性问题,就是所调查内容涉及私人机密而不愿或不便公开表态或陈述的问题。这类问题涉及到被调查者的隐私,如赌博、吸毒、婚前性行为、偷税漏税等。如果直接提问,易引起反感,或者得不到真实回答,或者遭到拒绝,无法得到结果。尽管国外已设计出随机化回答技术对此类问题进行调查,但传统调查是以随机抽取已准备好的卡片或彩球来实现,其方式陈旧、不方便,在大范围情况下抽样难以实施。现代信息技术的飞速发展正好弥补了以上不足,也随之产生了新的调查方式——网上敏感性问题的调查。所谓网上敏感性问题的调查,就是通过上网方式,按照随机原则抽取部分网民作样本进行敏感性问题的调查,进而估计全部网民特征的调查方法。由于运用了信息技术这一新方式,使得网上敏感性问题的调查具有独特而鲜明的特点。

一 网上敏感性问题的调查的可行性

目前,国内国际互联网用户已达数百万人。用户的增长速度快,分布范围广,运用领域多。据国际权威组织预测,到 2003 年中国网民的数量将会达到 1 800 万。基于以上事实,用因特网做敏感性问题的调查已具备了数量基础,并且随机化回答技术所要求的特点,网上调查都能做得更好:在运用该技术前要向被调查者解释清楚,被调查者选取哪一个问题,回答什么都是保密的,不会暴露自己的隐私。但被调查者在面对调查者或其他被调查者时,总感到一种无形的压力,影响调查的真实性。而进行网上调查,根本无需与任何人见面,选择时只要轻点鼠标一下,问题就回答了,既无心理压力,又无人知晓,不会让调查者、被调查者尴尬。

接受该技术调查的群体应具备较高素质,否则无法理解该技术的保密特点。而我国当前网上群体是城市知识层次较高的年轻人群,正好适应了上述要求。只要网上说明清晰细致,被调查的网民一般都能够理解和接受,容易采取合作的态度,也就能使调查顺利进行。

二 网上敏感性问题的调查的方案设计

(一)进行网上抽样。运用随机 IP 自动拨叫技术来进行网上随机抽样。通过对一个随机 IP 发生软件设定 IP 码段取值范围来产生一批随机 IP 地址,再利用 IP 自动拨叫软件,向这些 IP 地址呼叫,并传送到类似“打扰一下,我们是××× 调查机构,现正对××× 问题进行网上调查,网址是×××。如果您有兴趣参加本次调查活动,我们将非常感谢,这是一次有奖调查活动,谢谢!”的信息,由网民自己决定是否参加调查。

(二)编制网上敏感性问题的调查软件。1. 网上敏感性问题的调查软件的编制必须由专业软件人员进行。调查网页的首页是调查名称、目的及保密性介绍。应充分发挥网络多媒体功能,如背景采用亲切可人的图案,辅之

收稿日期: 1999-10-22

作者简介: 孙蕾(1979-),女,江苏徐州人,现是中南财经大学统计系三年级学生。

以轻音乐及“您无需担心”等语音。2. 网页的第二页为记录被调查者的有关情况,如工作、年龄、性别、身份证号码等等,根据调查目的不同具体确定需要记录的内容,并可作为抽样框的控制项,将不符合抽样范围的被调查者排除,如可通过年龄项禁止未满 16 岁的儿童进入调查网页,以保护上网儿童。另外,还可以检查被调查者是否已回答过问题,并加以措施防范重复回答。3. 第三页要随机抽取问题,由被调查者自己用 0~1 随机函数控制,先让随机函数开始运行,用户根据意愿随机结束运行,此时屏幕上将出现 0~1 之间的一个数。由于总共只有两个问题,可以将 0.5 作为分界点,小于 0.5 时,选择问题 I;大于等于 0.5 时,选择问题 II。确定了要回答的问题后,再直接进入下一页。4. 最后一页就是所提问题,以调查上网人群婚前性行为为例^①,准备问题 I: 你有婚前性行为,对吗? 问题 II: 你是八月份出生的,对吗? 如果上页被调查者抽到问题 I,本页则只出现问题 I 和“对”、“不对”两个选项。被调查者只需要用鼠标选择选项,通过因特网传输回调查者主机,自动录入最后结果。由于被调查者首先要将上述内容下载至自己电脑上,所以选择哪个问题只有本人清楚,而调查者仅录入“对”或“不对”的回答,不用关心是哪一个问题的答案,因此保护了被调查者的隐私。

(三)分析调查结果。在记录了被调查者的答案后,由电脑自动计算出调查结果。仍以调查上网人群婚前性行为为例,其基本原理如下: 设 π_A 是有婚前性行为的人的概率,样本总量为 n ,回答“对”的人数为 m ; π_B 为八月份出生的人的比例,它近似等于 $1/12$,由于用 0.5 作为分界点确定回答的问题,所以 P 值等于 $1/2$,则回答“对”的比例为: $\frac{m}{n} = P \cdot \pi_A + (1-P) \pi_B$, π_A 的总体估计值为: $\hat{\pi}_A = \frac{1}{P} [\frac{m}{n} - (1-P)\pi_B]$, π_A 的方差估计量为 $V(\hat{\pi}_A) = \frac{1}{(n-1)P^2} \cdot \frac{m}{n} (1 - \frac{m}{n})$ 。将这三个数学模型输入电脑,自动得到 $\hat{\pi}_A$ 和 $V(\hat{\pi}_A)$,即有婚前性行为的人的总体估计量及其方差。

三、网上敏感性调查的特点

通过因特网进行敏感性调查,传播媒介发生了变化,又加入了多媒体技术,使之具有了其它调查方式不可比拟的优势: 1 客观性。与传统调查方式相比,网上调查中的被调查者不用面对任何人,回答不真实的情况相应减少,使调查的可靠性大大提高。2 简单性。一方面网上调查组织简单,费用低廉,不需要派出调查人员,不受天气和距离限制,不需要印刷调查问卷,减少大量人力物力;另一方面,回答问题的方式和录入方式简单,仅仅按动鼠标就回答了问题并记录在案,无需人工收集整理。3 广泛性。网络覆盖面广,调查范围足够大。只要是因特网用户,都有可能接受调查,即使远在千里之外也毫无阻隔。4 保护性。敏感性问题的关键在于保护被调查者,通过调查方案设计,可以看出设计者处处为被调查者着想,使他们无后顾之忧。同时,网络两端的调查者与被调查者互不相识,也保护了被调查者的利益。

四、网上敏感性调查应遵循的原则

我国很少进行敏感性调查,主要是文化传统因素和调查方式的落后。然而有些敏感性问题又是应该进行调查并掌握的基本情况,故应遵循以下原则: 1 亲切性原则。网上敏感性调查可充分利用多媒体功能,运用多种形式的背景资料,文字、图形、声音共同发挥作用,营造出亲切的氛围,这是传统调查方式所无法比拟的。2 质量控制原则。采用网上敏感性调查,应防止发生一个被调查者多次回答问题给结果带来不可靠性偏差的问题。具体方法可用“IP 码+若干特征标志”的方法判断被调查者是否已回答,在调查软件中提醒已回答过问题的网友退出该调查软件,并禁止其进入回答问题的网页。3 答谢原则。为调动网友参与网上调查的积极性和主动性,适当给予奖励和答谢,十分必要。调查组织者可以用 E-mail 向被调查者寄去电子贺卡,或者以身份证号码为依据进行抽奖,获奖者的名单及时公布在调查站点上。尽量扩大中奖面,但奖品的价值可适当小一些。

[责任编辑: 郭诗梦]

^① 该方案所使用的随机化技术是西蒙斯模型,因而有一个问题是无关性问题。