

市场调查中敏感性问题的一种改进调查法*

周石泉

(广东省环境保护学校, 广东 广州 510655)

摘 要: 市场调查中敏感性问题的调查方法主要有 Greenberg 模型(1971年)及 Himmelfarb 和 Edgell(1980年)可加常量方法, 介绍在这两种方法基础上得到的能使均值的估计达到较高的精度的一种改进调查法。

关键词: 市场调查; 敏感性问题; 随机化回答调查法

中图分类号: O212.2 **文献标识码:** A **文章编号:** 0529-6579(2003)S1-0281-03

1 引 言

政府或企业进行市场调查是获取调查信息, 减少决策主观性, 增强产品竞争力, 提高企业经济效益的有效途径。所谓敏感性问题, 是指在调查中涉及政治态度或个人隐私, 企业商业秘密等被调查者往往拒绝回答或不愿真实回答的问题。理论界对于敏感性问题的界定比较有代表性的观点主要有两种: 一种观点认为, “敏感性问题等同于秘密, 是指机构、组织或个人由于经济、安全、形象等原因不宜或拒绝外部知晓的问题, 如政府秘密、军事秘密、企业商业秘密、家庭秘密和个人隐私等。”另一种观点认为, “敏感性问题是指令被调查者难堪禁忌的问题, 主要涉及各地风俗和民族习惯中忌讳, 或侵害个人利益, 隐私的尖锐问题。”

敏感性问题可分为两类: 一类是定量的敏感性问题即可用数字量化或度量的问题, 诸如对消费年龄, 体重收入等敏感问题的调查; 另一类是定性的敏感性问题, 属于性质的一面, 诸如调查消费者对某类产品的喜好等问题。

只有两种选择的定性问题, 可以通过 0-1 化处理为定量问题, 如规定回答“是”为 0, 回答“否”为 1。同样, 具有多种选择答案的定性问题, 如 k 种选择答案可以规定选择第 i 种答案变量为 i ($i=1, 2, \dots, k$)。因此, 对敏感性问题的技术处理最终都可以转化为定量的问题。一般地, 对定量问题, 我们最想得到的是对均值的估计量, 本文从均值估计量入手探讨介绍在 Greenberg 模型(1971年)及 Himmelfarb 和 Edgell(1980年)可加常量方法基础上得到的能使均值的估计达到较高的精度的一种改进调查法。

2 调查方法及均值的估计

假设我们需要调查一个定量的敏感性问题, 其真实答案为 x , 为保护被调查者, 并得到真实回答, 我们产生随机数 y 。为达到较高精度, 一般要求 y 的分布尽可能类似于 x 的分布。所以我们可以根据上次调查的结果或凭经验估计 y 的分布。假定 y 的分布密度为 $f(y)$ 。为简单起见, 这里假设已从总体中按简单随机抽样抽得 n 个被调查者, 被调查者的回答为 z 。 x 、 y 、 z 的均值为 u_x , u_y , u_z 方差为 S_x^2 , S_y^2 , S_z^2 ; z 的样本均值为 $\bar{z}$, 样本方差为 S_z^2 。

2.1 Greenberg 模型

调查方法为:

第 1 步, 产生 - 概率密度为 $f(y)$ 的随机数 y , 该过程可由计算机实现;

第 2 步, 产生 - 0,1 分布数 ϵ , $p(\epsilon=1)=p$ 。该过程容易实现, 如从 - 放置有若干红球和白球的盒中摸取一球, 摸到红球表示 $\epsilon=1$, 否则 $\epsilon=0$ 。其中盒中红球所占的比例为 p ;

第 3 步, 如果 $\epsilon=1$, 则被调查者回答敏感性问题 x , 否则要求被调查者回答看到的 y 值, x , y 与 ϵ 相互独立, 研究者只能看到被调查者给出的最终回答 z 。

将之归为模型, 得 $z = \epsilon x + (1 - \epsilon)y$

两边取期望, 由 x , Y 与 ϵ 的独立性及条件期望的概念, 得

$$\begin{aligned} u_z &= pu_x + (1-p)u_y \\ S_z^2 &= E[\epsilon x + (1-\epsilon)Y]^2 - [E(\epsilon x + (1-\epsilon)Y)]^2 = \\ &= pEx^2 + (1-p)EY^2 - [pu_x + (1-p)u_y]^2 = \end{aligned}$$

* 收稿日期: 2003-04-28

作者简介: 周石泉(1968年生), 男, 讲师; E-mail: zhoushiquan8888@sina.com

$$S_x^2 + (1-p)(S_y^2 - S_x^2) + p(1-p)(u_y - u_x)^2$$

因为 $z_1, z_2, \dots, z_n$ 为简单随机样本, 所以我们可取估计量 $\hat{u}_z = \bar{z}$, $\hat{s}_z^2 = s_z^2$, $\hat{u}_z$ 与 $\hat{s}_z^2$ 分别为 u_z 、 s_z^2 的无偏估计

$$v_{ar}(\bar{z}) = \frac{1-f}{n} s_z^2$$

$$\text{其无偏估计量 } \hat{v}_{ar}(z) = \frac{1-f}{n} \hat{s}_z^2$$

由于 $u_x = \frac{u_z - (1-p)u_y}{p}$ 若用 $\bar{z}_1$ 替代 u_z 则得

(1) 简单估计量

$$\hat{u}_{x1} = \frac{\bar{z} - (1-p)u_y}{p}$$

(2) 估计量的性质

$$\textcircled{1} \quad E(\hat{u}_{x1}) = U_x$$

$$\textcircled{2} \quad \hat{V}_{ar}(\hat{u}_{x1}) = v_{ar}\left(\frac{\bar{z}}{p}\right) = \frac{1-f}{np^2} s_z^2$$

$$\textcircled{3} \quad \hat{v}_{ar}(\hat{u}_{x1}) = \hat{v}_{ar}\left(\frac{\bar{z}}{p}\right) = \frac{1-f}{np^2} s_z^2$$

是 $v_{ar}(\hat{u}_x)$ 的无偏估计量。

Greenberg 模型中我们看到由于有 $(1-p)$ 的概率回答与被调查者无关的随机数 Y , 使得信息大量浪费, 从而导致 U_x 的估计精度不高, $P=1$ 时, 该模型成为一般的直接提问法, 这时均值估计的方差

$$v_{ar}(\hat{u}_x) = \frac{1-f}{n} s_x^2$$

由于

$$\begin{aligned} v_{ar}(\hat{u}_x) - v_{ar}(\hat{u}_{x0}) &= \\ \frac{1-f}{np^2} \{ & (s_x^2 + (1-p)(s_y^2 - s_x^2)) + \\ & p(1-p)(u_y - u_x)^2 - s_x^2 \} = \\ \frac{1-f}{np^2} [& p(1-p)s_x^2 + (1-p)s_y^2 + \\ & p(1-p)(u_y - u_x)^2] > 0 \end{aligned}$$

通过比较两者的差可看出, 要提高精度, 一是使 y 与 x 的分布相同, 或者做到 $u_y = u_x$, $s_y^2 = s_x^2$, 但由于 x 是研究对象其分布是不可能完全已知的, 所以我们只能做到使 y 的分布与 x 的分布尽量相似。提高精度的另一途径是增大 p 值, 但增大 p 值会导致被调查者的怀疑, 使数据不可靠, 所以只能取适当的 p 值。事实上, 从调查方案的设计中, 我们看到的最终的 n 个回答中理论上只有 p 的概率与所研究问题有关。是否我们可做到被调查者的回答完全含有所研究的敏感性问题的信息, 而又使被调查者无后顾之忧地给出其回答呢? 答案是肯定的。如 Himmelfarb 和 Edgell 的可加常量方法就做到了这一点。

2.2 Himmelfarb 和 Edgell 的可加常量方法

其调查方法为:

第1步, 产生一概率密度为 $f(y)$ 的随机数 y 。

第2步, 被调查者回答 $x+y$ 的和。其中 x 与 y 相互独立。研究者只能看到被调查者给出的最终回答 z 。将之归为模型, 则得 $z = x + y$

两边取期望, 得

$$u_z = u_x + u_y, s_z^2 = s_x^2 + s_y^2$$

取估计量 $\hat{u}_z = \bar{z}$, $\hat{s}_z^2 = s_z^2$ 且 $\hat{u}_z$, $\hat{s}_z^2$ 分别为 u_z , s_z^2 的无偏估计, 又

$$v_{ar}(\bar{z}) = \frac{1-f}{n} s_z^2$$

其无偏估计量为

$$\hat{v}_{ar}(\bar{z}) = \frac{1-f}{n} \hat{s}_z^2$$

由于 $u_x = u_z - u_y$ 所以如果我们用 $\bar{z}$ 替代 u_z , 则得

(1) 简单估计量:

$$\hat{u}_{x2} = \bar{z} - u_y$$

(2) 估计量的性质:

$$\textcircled{1} \quad E(\hat{u}_{x2}) = u_x$$

$$\textcircled{2} \quad v_{ar}(\hat{u}_{x2}) = v_{ar}(\bar{z}) = \frac{1-f}{n} s_z^2 = \frac{1-f}{n} (s_x^2 + s_y^2)$$

$$\textcircled{3} \quad \hat{v}_{ar}(\hat{u}_{x2}) = \hat{v}_{ar}(\bar{z}) = \frac{1-f}{n} \hat{s}_z^2$$

是 $v_{ar}(\hat{u}_{x2})$ 的无偏估计。

可加常量方法的一个最大特点是操作简单, 估计量的计算简单。但在每个回答的所含有敏感性问题的信息的同时, 又使所有回答引入与调查问题无关的随机数, 使估计量的精度在一方面提高的同时又有了下降。 $Y=0$ 时, 该模型成为一般的直接调查法。这时该估计量的方差 $v_{ar}(\hat{u}_{x0}) = \frac{1-f}{n} s_x^2$ 由于

$$\begin{aligned} v_{ar}(\hat{u}_{x2}) - v_{ar}(\hat{u}_{x0}) &= \\ \frac{1-f}{n} (s_x^2 + s_y^2 - s_x^2) &= \frac{1-f}{n} s_y^2 \geq 0 \end{aligned}$$

所以通过比较两者的差同样可看出, 要使精度提高, 在尽可能也含较多有用信息的同时, 最好尽可能少地引入无关信息。结合上述两种方法优点改进的一个方案是, 使每个被调查者的回答都含有敏感性问题的信息, 又使随机数仅以一定的概率影响被调查者的最终回答。

2.3 改进的调查方法

第1步, 产生一概率密度为 $f(y)$ 的随机数 y 。

第2步, 产生 $-0, 1$ 分布的随机数 ϵ , $p(\epsilon=1)=p$

第3步, 如果 $\epsilon=1$, 则要求被调查者回答 $x+u_y$ 的和; 否则回答 $x+y$ 的和。 x, y 与 ϵ 相线独立。研究者只能看到被调查者给出的最终回答 z 。将之归为

模型,则得

$$z = \varepsilon(x + u_y) + (1 + \varepsilon)(x + y)$$

两边取期望,得

$$u_z = u_x + u_y$$

$$\begin{aligned} s_z^2 &= E[\varepsilon(x + u_y) + (1 - \varepsilon)(x + y)]^2 - \\ &\quad [E(\varepsilon(x + u_y) + (1 - \varepsilon)(x + y))]^2 = \\ &= pE(x + u_y)^2 + (1 - p)E(x + y)^2 - (u_x + u_y)^2 = \\ &\quad s_x^2 + (1 - p)s_y^2 \end{aligned}$$

因为 $z_1, z_2, \dots, z_n$ 为简单随机样本,所以我们可取估计量 $\hat{u} = \bar{z}, \hat{\sigma}_z^2 = s_z^2$, 且 $\hat{u}_z, \hat{\sigma}_z^2$ 分别为 u_z, s_z^2 的无偏估计。又

$$v_{ar}(\bar{z}) = \frac{1-f}{n} s_z^2$$

其无偏估计量为

$$\hat{v}_{ar}(\bar{z}) = \frac{1-f}{n} \hat{\sigma}_z^2$$

由于 $u_x = u_z - u_y$
用它替代 u_x , 则得

(1) 简单估计量:

$$\hat{u}_{x3} = \bar{z} - u_y$$

(2) 估计量的性质:

$$\textcircled{1} \quad E(\hat{u}_{x3}) = u_x$$

$$\textcircled{2} \quad v_{ar}(\hat{u}_{x3}) = v_{ar}(\bar{z}) = \frac{1-f}{n} s_z^2 = \frac{1-f}{n} [s_x^2 + (1-p)s_y^2]$$

$$\textcircled{3} \quad \hat{v}_{ar}(\hat{u}_{x3}) = \hat{v}_{ar}(\bar{z}) = \frac{1-f}{n} \hat{\sigma}_z^2$$

是 $v_{ar}(\hat{u}_{x3})$ 的无偏估计。改进方法对 Greenberg 模型以及 Himmelfarb 和 Edgell 的可加常量方法均作出了改进。它能使所有回答均含敏感性问题信息, 同时, 引入无关随机数的概率只有 $1-p$, 从而提高了精度。

参考文献:

- [1] 柯惠新, 丁立宏. 市场调查与分析[M]. 北京: 中国统计出版社, 2000.

An Improved Way of Investigation on Sensitive Problem in Market Survey

ZHOU Shi-quan

(Guangdong Province Environmental Protection School, Guangzhou 510655, China)

Abstract: There are two major ways of investigation on sensitive problems in market survey: Greenberg (1971), Himmelfarb and Edgell (1980). In this article an improved way of investigation based on the above two is introduced, which enables the evaluation of average value to reach a higher accuracy.

Key words: Market survey; sensitive problem; random reply survey method