

【统计理论与方法】

基于随机化回答模型的最低工资敏感性问题的研究

陈光慧, 韩兆洲

(暨南大学 经济学院, 广东 广州 510632)

摘要:针对最低工资调查中出现的敏感性问题,引入随机化回答技术,并针对定性和定量两类敏感性问题,分别引入不同的随机化回答模型。同时还考虑到最低工资调查中,不同调查单位对同一问题敏感程度不同的特性,在定量问题的随机化回答模型中引入敏感性水平,对原有模型进行有效的改进,使被调查者能够更加积极配合最低工资调查,从而能够进一步减少由于各类敏感性问题造成的非抽样误差。这套随机化处理方法还可推广应用到其他类型的敏感性问题中。

关键词:最低工资调查;敏感性问题;随机化回答

中图分类号:C811 **文献标志码:**A **文章编号:**1007-3116(2012)09-0003-05

一、引言

包括中国在内的世界上很多国家都实行了最低工资保障制度,以此来保障处于最低收入的那部分劳动者能够维持其自身及其家属的基本生存能力。就中国来说,如何准确地确定某一地区的最低工资标准一直是困扰各级政府的一个难题,各地劳动和社会保障部门往往凭借各主管部门领导的经验以及社会各方面专家、学者的建议来确定一个最低工资标准^{[1]-16}。这种估计方法带有很大的主观性,是一种不科学的方法。要想确定一个恰当的最低工资标准就必须先了解这一地区实际的最低工资水平;而要想了解实际的最低工资水平,最科学的方法莫过于进行最低工资的抽样调查。但是,从国内外的发展现状来看,关于最低工资抽样调查方法研究的发展非常缓慢,目前还没有一种成熟、可用的调查方案,其主要原因在于最低工资调查的总体很难确定,从而得不到一个完整的抽样框。也就是说,不能确切地知道一个地区哪些人处于最低工资水平?因而很难构建一个关于该地区最低工资人群的抽样框。

为了解决这一难题,笔者主张把某一地区所有

存在雇佣关系的用人单位作为调查总体,即把各种工业、服务业等所有用人单位作为总体单位,从而构成一个抽样框。这样可随机抽取一个由不同的用人单位构成的样本,通过调查样本中的这些用人单位来获取最低工资的信息。在具体调查某个用人单位的最低工资水平时,最便捷有效的方法就是直接去询问用人单位的有关负责人,他们对本单位的工资水平了如指掌,但是最低工资调查中所包含的调查项目很多都是敏感性问题,比如该用人单位支付的最低工资是否高于本地区的最低工资标准?该用人单位支付的最低工资是多少?等等,对于这些问题,如果直接询问用人单位的有关负责人,通常很难取得对方的合作,因为如果如实说出本单位的最低工资低于最低工资标准,用人单位负责人担心政府有关部门会追究该单位的有关责任。所以用人单位负责人很可能会拒绝回答,或者干脆故意抬高该用人单位的最低工资水平,因而很难得到真实的回答,这样就会大大地增加非抽样误差,有时甚至会得出错误的结论。

针对这些敏感性问题,本文运用随机化回答模型来解决,并根据最低工资调查所固有的特点,引入

收稿日期:2012-03-03; 修复日期:2012-05-16

基金项目:国家社会科学基金项目《我国最低工资调查方法与统计测算模型研究》(06BTJ017);国家社会科学基金项目《我国经常性抽样调查制度与方法体系改革研究》(10CTJ006)

作者简介:陈光慧,男,安徽旌德人,经济学博士,副教授,硕士生导师,研究方向:统计调查与数据分析;

韩兆洲,男,江苏苏州人,经济学博士,教授,博士生导师,研究方向:经济预测与决策。

敏感性水平,对原有定量问题的随机化回答模型进行改进,从而更加有效地消除用人单位负责人的顾虑,使之积极配合,以进一步减少非抽样误差。

二、定性问题的随机化回答模型 在最低工资调查中的应用

随机化回答技术最早由 Warner 提出^[2],其目的是为了调查定性的敏感性问题,比如被调查者是否偷税漏税、考试是否作弊以及吸毒、婚前性行为等。在最低工资调查中,也会出现这一类定性的敏感性问题,比如被调查的用人单位支付的最低工资是否高于当地的最低工资标准? Warner 提出的随机化回答模型是最简单的一种随机化模型,即将调查的敏感性问题按照正反两方面提出,比如在最低工资调查中,调查人员可向被调查的用人单位负责人提出如下两个相反问题,这两个问题的答案都是“是”或者“否”。

问题 A:本用人单位支付的最低工资高于当地的最低工资标准吗?

问题 B:本用人单位支付的最低工资低于当地的最低工资标准吗?

在进行调查之前,制作两种外型一致的卡片若干张,分别写上 A 和 B,其中 A 卡片所占的比例是 p ,放在一个盒子里均匀混合。调查时,可让被调查的用人单位负责人从盒子里随机抽取一张卡片,如果是 A 卡片,则如实地回答问题 A,反之则如实地回答问题 B。被调查者到底抽到哪一张卡片,调查人员无权过问。通过这种方法能够消除被调查者的顾虑,从而得出真实的回答。如果有 n 个被调查者,其中有 m 个被调查者回答了“是”,那么对于这一地区所有用人单位,支付的最低工资高于当地最低工资标准的那些用人单位所占的比例 π ,Warner 给出了一个无偏的极大似然估计量:

$$\hat{\pi} = \frac{1}{2p-1} \left[\frac{m}{n} - (1-p) \right] \quad (p \neq 0.5)$$

此估计量的方差为:

$$V(\hat{\pi}) = \frac{\pi(1-\pi)}{n} + \frac{p(1-p)}{n(2p-1)^2}$$

科克伦和张尧庭^{[3]594-600}、Chaudhuri 和 Mukerjee^{[4]23-49}、Bar-Lev 等人^[5]对此类定性的敏感性问题随机化处理进行了详细讨论。通过随机化技术,虽然使估计量的方差增大、估计精度有所下降,但是得到的估计量却是无偏的,因为通过随机化回答技术能够得到真实的回答,消除了虚报和瞒报

的现象,从而减少了非抽样误差。这种以牺牲一定的估计精度换取真实回答的做法还是值得的。

三、定量问题的随机化回答模型

以上的随机化回答模型有效地解决了定性敏感性问题,但是在很多调查中敏感性问题并不局限于定性问题,还存在定量的敏感性问题,比如在最低工资调查中,询问某用人单位的最低工资水平;在工业调查中往往想知道某企业的利润、纳税情况等,这些定量的信息往往比定性的信息更为重要。自从 Warner 提出针对定性敏感问题的 Warner 模型之后,很多学者在定性问题研究的基础上研究定量的敏感性问题。Greenberg 等人^[6-7]根据 Simmons 模型的设计原则^[8],应用两个随机样本技术类似地估计一个敏感的离散型或连续型变量的均值,从而对最初的 Warner 随机化模型进行了改进,即 Greenberg 模型。Himmelfarb 和 Edgell 提出了可加常量的方法^[9],这是处理定量敏感问题的常用方法。之后,Eichhorn 和 Hayre 提出了很多针对定量敏感问题的随机化回答模型,并给出了无偏的估计量^[10];金莹和梁小筠对 Himmelfarb 和 Edgell 提出的可加常量方法进行改进^[11],有效地降低了估计量方差并提高了估计精度。假定敏感问题的真实答案为 X ,具体做法如下:

第一步:产生一个概率密度为 $f(y)$ 的随机数 Y ,其中均值 μ_Y 和方差 S_Y^2 已知。

第二步:产生一个 $0 \sim 1$ 分布的随机数 ϵ ,且满足 $P(\epsilon = 1) = p$ 。

第三步:如果 $\epsilon = 1$,则要求被调查者回答 $X + \mu_Y$ 之和,否则回答 $X + Y$ 之和。

这里的 X, Y 与 ϵ 相互独立。调查者只能看到被调查者给出的最终回答 $Z = X + Y$,将其归纳为如下模型:

$$Z_i = \begin{cases} X_i + \mu_Y & \text{以概率 } p \text{ 回答} \\ X_i + Y_i & \text{以概率 } 1-p \text{ 回答} \end{cases}$$

其中 $i = 1, 2, \dots, n$ 。基于此随机化回答模型,提出关于研究变量 X 的总体均值 μ_X 的估计量为:

$$\hat{\mu}_X = \bar{z} - \mu_Y$$

其中 $\bar{z} = \sum_{i=1}^n Z_i / n$

由 X, Y 与 ϵ 的独立性及条件期望的概念,可得如下性质一:

$$\textcircled{1} E(\hat{\mu}_X) = \mu_X$$

$$\textcircled{2} V(\hat{\mu}_X) = V(\bar{z}) = \frac{1-f}{n} S_Z^2$$

$$= \frac{1-f}{n} [S_X^2 + (1-p)S_Y^2]$$

$$\textcircled{3} \hat{V}(\hat{\mu}_X) = \hat{V}(\bar{z}) = \frac{1-f}{n} s_z^2$$

是 $V(\hat{\mu}_X)$ 的一个无偏估计量。

这种改进方法对 Greenberg 模型以及 Himmelfarb 和 Edgell 模型的可加常量方法均作出了改进, 使所有回答均含有敏感性问题的信息, 同时又只有 $1-p$ 的概率引入无关的随机数, 从而减少了估计量方差并提高了估计精度。Gupta 等人在此基础上提出了一套基于无放回简单随机抽样设计的最优随机化回答技术^[12], Arnab 把这种最优随机化回答技术扩展到了任意抽样设计下^[13]。不过以上这些文献都存在一个缺陷, 即都假设不同的回答者对同一问题的敏感程度是相同的, 这一假设在很多实际问题中并不能得到满足。

四、定量问题的随机化回答模型 在最低工资调查中的应用

在最低工资调查中, 调查者想知道某一地区实际的最低工资水平, 这是一个定量的敏感性问题, 因为各个地区都规定了法定的最低工资标准, 如果没有达到最低工资标准, 用人单位的负责人可能会担心受到有关部门的追究, 因而对这个问题非常敏感。如果直接询问该问题, 用人单位负责人很可能会故意隐瞒或者给出不真实的工资水平, 这将大大地增加调查的非抽样误差。为此, 调查者就有必要引入定量的随机化回答模型来尽量减少这种非抽样误差。

前文提到的改进的随机化回答模型^[11-13], 在一定程度上减少了估计量方差, 并提高了估计精度, 是一种较好的处理定量问题的随机化回答模型, 但是最低工资调查具有不同于其它调查的特殊性, 用人单位的最低工资水平这一类敏感性问题也不同于一般的敏感性问题。因为各地都规定了最低工资标准, 对于那些最低工资水平高于当地最低工资标准的用人单位来说, 这并不是什么敏感性问题, 用人单位很可能会毫无顾忌地直接回答, 并不需要那些繁琐的随机化技术; 而只有那些最低工资水平低于当地最低工资标准的用人单位才会把这一问题当作比较严重的敏感性问题。简言之, 不同的用人单位对同一个问题的态度或者敏感程度是不一样的, 因此本文根据最低工资调查所固有的特点, 对原有的随机化回答模型进一步加以改进。

假设用人单位最低工资的真实水平为 X , 该地区的法定最低工资标准为 μ_Y , 可进行如下步骤:

第一步: 调查者询问被调查的用人单位负责人是否愿意直接回答问题。

第二步: 如果愿意直接回答, 则直接回答 $X + \mu_Y$ 。

第三步: 如果不愿意直接回答, 则产生一个概率密度为 $f(y)$ 的随机数 Y , 最终回答 $X+Y$ 之和, 调查者只能看到被调查者给出的最终答案 $Z = X+Y$ 。

这里 $f(y)$ 是以 μ_Y 为均值的对称分布, 如均匀分布、正态分布等, 且方差 σ_Y^2 已知, 将其归纳为如下模型:

$$Z_i = \begin{cases} X_i + \mu_Y & \text{如果愿意直接回答} \\ X_i + Y_i & \text{否则随机化回答} \end{cases}$$

不同用人单位的负责人对此问题的敏感程度是不一样的, 假定用 W 表示某个用人单位希望用随机化方式回答此敏感性问题的概率, 那么 W 就可以称之为敏感性水平。如果某个问题非常敏感, 那么就会有更多的用人单位选择随机化回答, 因而 W 就会接近于 1, 否则接近于 0。但是, 在调查之前, W 是未知的, 还需要通过样本进行估计。由 X 、 Y 的独立性及条件期望的概念可得:

$$\begin{aligned} E(Z_i) &= (1-W)E(X_i + \mu_Y) + WE(X_i + Y_i) \\ &= E(X_i) + (1-W)\mu_Y + WE(Y_i) \\ &= \mu_X + \mu_Y \end{aligned} \quad (1)$$

$$\begin{aligned} E(Z_i^2) &= (1-W)E(X_i + \mu_Y)^2 + WE(X_i + Y_i)^2 \\ &= (1-W)[E(X_i)^2 + 2\mu_X\mu_Y + \mu_Y^2] + \\ &\quad W[E(X_i)^2 + E(Y_i)^2 + 2\mu_X\mu_Y] \\ &= E(X_i)^2 + 2\mu_X\mu_Y + (1-W)\mu_Y^2 + \\ &\quad WE(Y_i)^2 \\ &= \mu_X^2 + S_X^2 + 2\mu_X\mu_Y + \mu_Y^2 + WS_Y^2 \end{aligned} \quad (2)$$

因而 Z 的方差为:

$$\begin{aligned} S_Z^2 &= E(Z_i^2) - E(Z_i)^2 \\ &= \mu_X^2 + S_X^2 + 2\mu_X\mu_Y + \mu_Y^2 + WS_Y^2 - (\mu_X + \mu_Y)^2 \\ &= S_X^2 + WS_Y^2 \end{aligned} \quad (3)$$

基于此随机化回答模型, 提出关于研究变量 X 总体均值的估计量为:

$$\hat{\mu}_X^* = \bar{z} - \mu_Y$$

$$\text{其中 } \bar{z} = \sum_{i=1}^n Z_i / n$$

可得如下性质二:

$$\textcircled{1} E(\hat{\mu}_X^*) = \mu_X$$

$$\textcircled{2} V(\hat{\mu}_X^*) = V(\bar{z}) = \frac{1-f}{n} S_Z^2$$

$$= \frac{1-f}{n} [S_x^2 + WS_y^2]$$

$$\textcircled{3} \hat{V}(\hat{\mu}_x^*) = V(\bar{z})$$

$$= \frac{1-f}{n} s_z^2 = \frac{1-f}{n} [s_x^2 + \hat{W}S_y^2]$$

是 $V(\hat{\mu}_x^*)$ 的一个无偏估计量。

证明 通过式(1)、(2)、(3)容易推出性质二中的①、②。下面证明性质二中的③。未知的敏感性水平 W 表示某个用人单位希望用随机化方式回答此敏感性问题的概率,则 W 的样本估计量为:

$$\hat{W} = \frac{\text{运用随机化方式回答的用人单位数}}{\text{样本量 } n} \quad (4)$$

易知 \hat{W} 是 W 的无偏估计量。通过式(3)可得 $S_x^2 = S_z^2 - WS_y^2$ 。那么可给出 S_x^2 的一个无偏估计量为:

$$s_x^2 = s_z^2 - \hat{W}S_y^2$$

其中 $s_z^2 = \sum_{i=1}^n (Z_i - \bar{Z})^2 / (n-1)$, \hat{W} 由式(4)给定,这时可得 $V(\hat{\mu}_x^*)$ 的一个无偏估计量为:

$$\hat{V}(\hat{\mu}_x^*) = \hat{V}(\bar{z}) = \frac{1-f}{n} s_z^2 = \frac{1-f}{n} [s_x^2 + \hat{W}S_y^2]$$

五、结 论

前文研究的随机化回答方法初看起来与原有的方法无异,但实际上两者有本质区别。在金莹和梁小筠、Gupta 等人以及 Arnab 传统的随机化方法中,概率 p 是事先主观确定的,任何被调查者都以一定的概率来随机确定回答 $X_i + \mu_y$ 或者 $X_i + Y_i$ 。这实际上是事先假定了所有被调查者对敏感问题的敏感程度是一样的,这样往往会抹杀不同被调查者之间的差异。具体到最低工资调查,很多用人单位尤其是那些最低工资高于当地最低工资标准的用人单位,对最低工资问题并不敏感,而且有些用人单位的负责人考虑到这种随机化装置的复杂性,为了省事也

愿意直接回答。所以对于这些人,如果强迫他们使用这种较为复杂的随机化装置,可能会引起抵触情绪,继而不愿意积极配合,在这种情况下用人单位很可能会随意给出一个不真实的答案。实际上只有那些对最低工资调查非常敏感的用人单位,才愿意并积极配合使用这种复杂的随机化装置。前文提出的方法正是满足了不同用人单位的意愿,赋予用人单位选择的权利,这样不管用人单位对最低工资调查的敏感程度如何,一般都会积极配合,从而得出真实的答案。

在最低工资调查中引入较为复杂的随机化回答技术而不使用直接询问的方式,其目的就是为了解除用人单位的顾虑,以给出真实的回答。因此,为了让随机化回答技术充分发挥其作用,在调查之前还有必要做好以下几方面工作:

第一,必须对调查人员进行相应的培训,使之充分了解随机化方法的原理和实质,这样才能很好地向被调查的用人单位进行解释^[14]。

第二,调查之前,调查人员应向被调查的用人单位负责人解释这种调查方法的原理和实质,同时对于那些愿意直接回答最低工资的用人单位,尤其应该说明此次最低工资调查的真正目的是为了了解总体的情况,而并不是为了了解某个用人单位的具体情况,并会对所有用人单位的具体信息保密。

第三,随机化回答的设计要尽可能简单明了,调查人员可向被调查者演示几次,以了解这种随机化的真实性,从而彻底消除被调查者的顾虑。

在最低工资调查中引入随机化回答技术,虽会增加一部分工作量,但是相比直接询问的调查方式而言,随机化回答技术能够有效地对调查结果进行“脱水”,在很大程度上减少了回答误差,因而还是值得使用的一种调查方法。

参考文献:

- [1] 韩兆洲. 劳动工资与社会保障[M]. 北京:经济科学出版社,2006.
- [2] Warner S L. Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias [J]. Journal of the American Statistical Association, 1965(1).
- [3] 科克伦,张尧庭. 抽样技术[M]. 吴辉,译. 北京:中国统计出版社,1985.
- [4] Chaudhuri A, Mukerjee R. Randomized Response Theory and Technique[M]. New York:Marcel Dekker, 1988.
- [5] Bar-Lev S K, Bobovich E, Boukai B. A Two-Stage Sequential Scheme for Warner's Response Model[J]. Communications in Statistics - Theory and Methods, 2003(12).
- [6] Greenberg B G, Abulel A, Simmon W R, Horvit D G. The Unrelated Question Randomized Response Model: Theoretical Framework[J]. Journal of the American Statistical Association, 1969(6).
- [7] Greenberg B G, Kuebler R R, Abernathy J R, Horvitz D G. Application of the Randomized Response Technique in Obtaining Quantitative Data[J]. Journal of the American Statistical Association, 1971(1).

- [8] Horvitz D G, Shah B V, Simmons W R. The Unrelated Randomized Response Model[J]. Proceedings of Social Statistics Section, American Statistical Association, 1967(1).
- [9] Himmelfarb S, Edgell S. Additive Constants Models: A Randomized Response Technique for Eliminating Evasiveness to Quantitative Response Questions[J]. Psychological Bulletin, 1980(1).
- [10] Eichhorn B H, Hayre L S. Scrambled Randomized Response Methods for Obtaining Sensitive Quantitative Data[J]. Journal of Statistical Planning and Inference, 1983(1).
- [11] 金莹,梁小筠. 对定量的敏感性问题的一种改进调查法及其估计量[J]. 统计研究, 2000(11).
- [12] Gupta S, Gupta B, Singh S. Estimation of Sensitivity Level of Personal Interview Survey Questions[J]. Journal of Statistical Planning and Inference, 2002(1).
- [13] Arnab R. Optional Randomized Techniques for Complex Survey Designs[J]. Biometrical Journal, 2004(1).
- [14] 蒋萍,田成诗. 全方位、立体性数据质量概念的建立与实施[J]. 统计研究, 2010(12).

The Study of Sensitive Questions in the Minimum Wage Based on Randomized Response Models

CHEN Guang-hui, HAN Zhao-zhou

(School of Economics, Jinan University, Guangzhou 510632, China)

Abstract: In order to resolve sensitive questions in the survey of minimum wage, randomized response techniques are introduced. Concretely, different randomized response models are applied to solve qualitative and quantitative sensitive questions. According to the particularity of the survey of minimum wage, different sample units have different sensitivity to the same sensitive question. Therefore, the sensitivity level of different units is introduced to the conventional quantitative randomized model and the interviewee would cooperate with the survey of minimum wage much more. As a result, non-sampling error would be reduced much more.

Key words: minimum wage survey; sensitive question; randomized response

(责任编辑:郭诗梦)

敬告作者

——关于投稿中几个必须引起注意的事项

在长期的编辑实践中,我们发现相当一部分作者在来稿中出现一些带有共同性、普遍性的问题。现就这些问题提出如下要求,请作者在投稿时参照执行。

1. 摘要的撰写。中华人民共和国国家标准 GB6447—86《文摘编写规则》对摘要的定义、类型、要素以及撰写文摘的注意事项等都有详细的说明。但我们收到的稿件中相当一部分摘要书写不规范,摘而不要、摘而无要、语言逻辑不连贯、不严谨等现象十分严重,这给编辑工作带来很多不便。这里敬请广大作者认真研读《文摘编写规则》,知道写什么和该如何写,确保写出高质量的摘要。同时,撰写摘要也是学术写作与学术训练的基本功,我们希望作者朋友能以严谨认真的态度对待摘要的撰写。另外,高质量的摘要对文章发表后能否被学术类文摘报刊转载也起很大作用,敬请作者格外注意。

2. 参考文献著录规则。本刊按照中华人民共和国国家标准 GB/T 7714—2005《文后参考文献著录规则》编辑本刊。在长期编辑工作中,发现来稿在这方面存在的问题较多:一是在文中具体引用、参考处未在右上角加方括号标注;二是文中与文后序号错乱;三是文后参考文献不规范。敬请作者认真研读此标准,这也是做学问的基本功之一,同时反映出作者的严谨态度和学术精神。本刊来稿中文后参考文献常见问题有如下各项:①缺文献分类号,如[J]、[N]、[D]等。②著录格式不合规范,或前后项位置次序不合要求,或有缺项。③最常见的是外文参考文献中将期刊的卷次、期次混淆,常缺期次,如“1985(88)”,应为“1985, 88(3)”,表明是某期刊 1985 年第 3 期那一期,而这一期的卷次按该刊连续编排应是第 88 卷。外文文献中的作者也要遵从姓前名后的原则。④书籍缺出版地、具体页码。⑤文后参考文献的顺序应依文中参考、引用的顺序为准排序,不应错乱。

3. 中图分类号。许多来稿中缺少中图分类号这一项。添加这一项的主要目的是为了对文献进行分类,便于读者检索。本刊现采用北京图书馆出版社出版的《中国图书馆分类法》(第四版)对学术论文进行文献分类。敬请作者在来稿时注意添加这一项,因为作者对所研究的问题,包括历史与现状相对熟悉,选择分类号时更不容易出错,这对该文献的检索将起到很大的帮助。

4. 基金项目的标注。本刊来稿中各类基金项目稿件很多,但经常出现项目具体名称遗漏现象。请参照本刊格式将基金项目类别、具体名称、编号等书写齐全。

本刊编辑部