

# 敏感性问题调查方法探讨

郑俊池 秦亦民

(华北煤炭医学院)

人们为研究某一问题往往需要进行调查.但有些问题采用通常使用的口头询问、开调查会或信访的方式,则不是无法调查就是难以获得真实情况.这类问题即所谓的“敏感性问题”.对这种敏感性问题

的调查可采用“随机化选答”的方法.随机化选答方法的最大优点应是能为调查对象保守秘密.但目前的教科书或期刊中所介绍的随机化选答方法有的实际上并不能为全部调查对象保守秘密.如文献[1]、[2]的方法;有的虽能保密,但由于方法本身的缺陷,使调查对象若无相当高的文化水平,则不易正确回答,如文献[3]所介绍的另一种方法.有关敏感性问题的调查方法的保密性问题未见有文献报道,本文就其保密性问题进行探讨,并对两种方法的优劣进行比较后,提出了一种改良的方法供参考.下面仅以《数理统计与管理》及高等医药院校教材《卫生统计学》中所介绍的方法为例进行探讨.

文献[1]与[2]所介绍的方法是通过设计一组相互无关的问题,让调查对象从随机选择回答(以下简称为无关问题的随机化选答).相互无关的问题的模型如:

{ 您这次流产是未婚人工流产吗?  
您是单月份出生的吗?

这里,  $A$  是调查者所关心的敏感问题,  $B$  是与  $A$  无关的非敏感问题.让每个调查对象自己用二项随机试验的方法(如抓阄法、掷币法等)确定回答哪一个问题.要求遵守约定.如任意取出一张人民币,号码末

位为奇者必须答  $A$ , 为偶数或必须答  $B$ , 抽定题后据实回答“是”或“否”,但不告知调查员自己回答的是哪一个问题.然后按式(1)计算对问题  $A$  回答“是”的样本率 ( $p_A$ ).

$$p_A = \frac{\lambda - (1-p)p_B}{p} \quad (1)$$

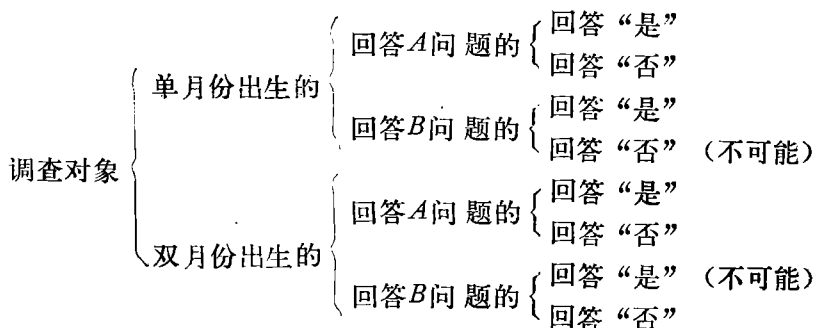
$p_A$  的标准误可按下式计算

$$S_{p_A} = \sqrt{\frac{\lambda(1-\lambda)}{np^2}} \quad (2)$$

式中  $\lambda$  为所有调查对象对  $A$ 、 $B$  两个问题回答“是”的总率,  $p$  与  $(1-p)$  分别为每个调查对象抽中  $A$  题、 $B$  题的概率(可由所选用的随机方法得知),  $p_B$  为调查对象对非敏感问题回答“是”的样本率(可另行组织调查或由样本得知,例如可从样本中的每个个体所在单位查出单、双月出生的各为多少人,然后即可计算出  $p_B$ ),  $n$  为样本容量.

从表面上看,该方法是让调查对象自己用随机的方法确定回答哪个问题,也即其它人并不知道调查对象回答的是哪一个问题,故可为调查对象保守秘密.实际上,一部分调查对象回答的是哪一个问题,其它人是很容易由他(她)是否具有问题  $B$  的某种特征推知的,因为这种非敏感性问题的特征事先或事后很容易了解清楚,有的甚至是不言自明的,如文献[3]方法2中介绍的非敏感性问题“你是女生吗?”.对部分调查对象来说,该方法也就无保密的作用了.

为了叙述的方便,我们可按上文所列出的相互无关问题的模型的内容,将调查对象按出生的单、双月份分为两类,每类



由上述的分类可见,对于单月份出生的人而言,回答“是”的,其它人确实无法得知她回答的是哪一问题,也即可以保密;回答“否”的,其它人很容易确定她回答的是敏感问题(假设调查对象都能据实回答的话,单月份出生的人不会对B问题做出“否”的回答),但这种情况无需保密.对于双月份出生的人而言,回答“否”的,其它人无法得知她回答的是哪一问题,但这种情况亦无需保密;回答“是”的,则其它人很容易确定她回答的是“A”问题,这时也即不能为其保密.如果对敏感问题A的回答“否”是不愿告知它人的,例如敏感问题“您拥护社会主义吗?”,则情形刚好与上述的分析相反.由以上分析可知,该方法并不能为全部调查对象保守秘密.随着敏感问题的调查技术的推广应用,一旦有的调查对象经独立思考看出了该方法的破绽,就可能误认为这是一种欺骗,这种看法流传开来,则可能产生不良后果.

文献〔3〕所介绍的第一种方法\*与上述无关问题的随机化选答的主要区别点在于对某一敏感问题向调查对象提出相关的正、反两个问题,让调查对象从中随机选

又可按回答A或B问题分为两类,最后再分别回答“是”或“否”.

择回答(简称相关问题的随机化选答).正、反问题的模型如:

{	A: 您对马列主义课
	感兴趣吗?
{	B: 您对马列主义课
	不感兴趣吗?

这里所采用的二项随机试验的方法要求相互对立的两种结果中,出现某一结果的概率 $p$ 不得为0.5.这可通过设计一个随机箱或抓阄的方法来实现.例如箱内放四个红球,一个白球,约定抓到红球回答A问题,抓到白球回答B问题.这时回答A问题的概率 $p = 4/5$ ,回答B问题的概率 $1 - p = 1/5$ .根据全概率公式,可得到如下关系式:

$$\lambda = p p_A + (1 - p)(1 - p_A) \quad (3)$$

式中各符号的意义同(1)式.因 $p$ 已知,解得

$$p_A = \frac{\lambda - (1 - p)}{2p - 1} \quad (p \neq \frac{1}{2}) \quad (3)$$

$p_A$ 的标准误可按式计算

$$S_{p_A} = \sqrt{\frac{\lambda(1 - \lambda)}{n(2p - 1)^2}} \quad (5)$$

相关问题的随机化选答法与无关问题的随机化选答法相比,优点为:①由于这方法对某一敏感问题向调查对象提出了有关该问

\*: 文献〔3〕共介绍了两种方法,第二种方法同文献〔1〕,〔2〕.

题的正、反两个问题,故不论调查对象回答“是”或“否”,其它人都无法得知其回答的究竟是哪一个问题,从而真正起到了保密的作用.②完全符合二项分布原理, $p + (1-p) = 1$ ,  $p_A + p_B = 1$ ,易于理解.③调查方案实施简便,不需另组织调查 $p_B$ 的值,  $p_B = 1 - p_A$ .④不必考虑无关问题选择恰当与否的问题.故总的来讲,相关问题的随机化选答法从理论上完全克服了无关问题随机化选答法的泄密问题,且方法简便,更易于实施.但该法的最大缺陷是没有相当高的文化程度的调查对象难以正确回答.尽管在调查之前即对如何回答问题作了规定,但由于所提问题之一是一般疑问句的否定式,如“您不拥护社会主义吗?”,调查对象往往搞不清该回答“是”或“否”,故限制了该方法在文化程度较低(如仅具有初中或小学文化程度)人群中的应用.为避免调查对象因对正、反问题何时该回答“是”、何时该回答“否”的费解可能带来的错答,我们将之改良为“正、反命题的随机化选择判断”.例如:

{ 您拥护社会主义.  
 { 您不拥护社会主义.

答案:符合、不符合

然后随机选择对 $A$ 或 $B$ 作出判断.符合自己情况的在答案“符合”上划“√”,不符合自己情况的在答案“不符合”上划“√”

(上接第14页)

根据以上统计分析结论,对秦皇岛港口引进出口煤炭机械化采样成套设备作出肯定答案是确信无疑的.

计算方法同相关问题的随机化选答,式中 $\lambda$ 为所有调查对象对 $A$ 、 $B$ 两个命题判断为符合的总率,  $p_A$ 为调查对象承认命题 $A$ 的样本率,余同前.

综上所述可知,目前国内教科书或期刊中介绍了二种敏感性问题的调查方法,其中无关问题的随机化选答法保密性能差,无关问题不易设置.相关问题的随机化选答法从根本上克服了前者保密性能差的弱点,构思巧妙,完全符合二项分布的原理,不需再另外组织辅助调查以计算 $p_B$ ,调查方案实施简便.但该方法的最大缺陷是调查对象因对正、反问题何时该回答“是”,何时该回答“否”的费解可能造成错答.我们将之改良为正、反命题的随机化选择判断,使之更适合我国的国情,从而弥补了该法之不足.故建议采用本文改良的“正、反命题的随机化选择判断法”取代“无关问题的随机化选答法”及“相关问题的随机化选答法”.

### 参考文献

1. 杨树勤主编(1986),卫生统计学(第2版),人民卫生出版社.
2. 常兆丰(1988),一种能为回答问题者保守秘密的调查方法,数理统计与管理.
3. 李从珠、杜凤超(1988),关于大学生某些敏感性问题的调查与统计分析,数理统计与管理.

### 参 考 资 料

1. ISO—1988 HARD COAL—SAMPLING
2. ASTM 05.05 D2234—86 Collection of a Gross Sample of coal
3. BC1017 Methodo for Sampling of Coal & Coke
4. AS 2646.2—1984 Sampling of Solid Mineral Fuels