敏感性问题统计推断模型探讨

徐云庆*
（浙江经济高等专科学校科研处，浙江嘉兴 314001）

摘　要：对市场调查，特别是对社会调查中敏感性问题处理中的两种主要随机化模型进行了评判，认为在运用这两种模型时，调查人员要充分理解和掌握方法；在调查中应允许被调查者在正式调查前检查卡片，了解记录方式；在使用西蒙斯模型时，要选择无关的非敏感性问题。

关键词：敏感性问题；调查；问卷设计；随机化模型。　**中图分类号**：F222

ABSTRACT: The article makes comments on the markct investigation, especially on the two random models dealing with sensitive problems in social investigation. It gives the opinion that research workers should thoroughly understand and master the ways of them while using the two models. Before investigation those who will be investigated should be allowed to check up the cards and get to know the way of recording; when using Simmons' models, non—sensitive problems which are not relevant are to be chosen.

KEYWORD: sensitive problems; research; questionaire designing; random models
文献标识码：A.　**文章编号**：1008—6781(2000)01—0031—(04)

无论是进行社会问题研究还是进行市场需求的统计分析研究，运用得最为普遍的当属抽样调查。抽样调查包括纯随机抽样、等距抽样、分层抽样、整群抽样等多种抽样组织形式，但对于具体的某个问题而言，通常采用以下两种手段收集数据：一是面对面的实地调查，如居民入户调查等；二是利用媒介工具进行调查，如邮寄调查、电话调查等。在使用以上两种调查方法时，对于一些比较敏感的问题，常常会遇到被调查者拒绝回答或故意答错的现象，从而影响调查结果的准确性。所谓敏感性问题，指的是风俗和民族习惯中忌讳的问题、个人隐私问题、有碍声誉的问题等。例如："您家有多少存款?"，"你考试时有作弊念头吗?"等。对于这类问题，被调查者有一种本能的自我防范心理，其结果要么不予回答，要么不真实回答，有的还会引起应答者反感。所以，在一般的问卷调查中，要尽量避免社会上禁忌和敏感性的问题。但有些调查本身就是敏感性调查，这就需要研究专门的方法，这些方法包括配备合适的访问者、注意访问技巧、选择恰当的调查时间、改进问卷的设计方法等。

对于较大规模的问卷调查，一般可用改进问卷的设计方法来解决，如假定法（即用一个假定性条件句作为问题的前提，然后再询问应答者的看法）、释疑法（即在问题前面写上一段消除疑虑的文字）及转移法等。但具有普查特征的大规模访问调查要花费很多的人力、物力和时间，从而使调查在内容和深度上都受到限制。若采用邮寄问卷的方法，人、财、物的消耗有一定的下降，但问卷的回收率很低，对影响答卷的因素也难以了解和判断。因此，以样本指标推算总体相关指标

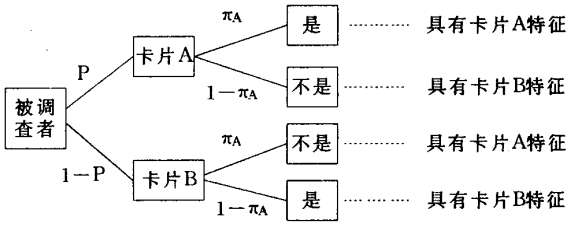
* 收稿日期：1999—08—25.　作者简介：徐云庆（1962—），男，浙江余姚人，浙江经济高等专科学校科研处副教授

的抽样技术得到了人们的重视，随机化回答模型的应用即是其中之一，它是采用一种既能保护被调查者个人秘密，又能使其讲实话，并在敏感性问题调查中获得某类人数所占比重或某敏感指标均值的估计量的方法。现有的随机化回答模型有两类：一类是用于敏感性问题的比例估计，如沃纳模型（Warner，1965）、西蒙斯模型（Simmons，1967）、双样本模型和隐含的随机化模型等；另一类是可同时用于敏感性问题的比例估计和均值估计，如非相关模型和转换模型等。这种方法在国外的社会调查中已得到广泛的应用，在我国的应用尚处在探讨阶段，其主要障碍是调查人员的素质较低和模型本身的繁琐性。下面着重探讨沃纳模型和西蒙斯模型的应用问题。

沃纳于 1965 年首先提出随机化回答模型，此后人们把此模型称为沃纳模型。例如某校欲对考试作弊学生人数的比例进行随机抽样调查，其做法是，将两个问题分别写在两叠卡片上，一叠卡片上的问题是“我在考试中作了弊”，另一叠卡片上的问题是“我在考试中没有作弊”。然后将两叠卡片混在一起，让被调查者随机从中抽取一张，并根据自己的真实情况回答卡片上的问题。调查人员不知道被调查者抽中哪张卡片，因此无法从他的回答中知其是否有作弊行为，以此来鼓励被调查者作出真实回答。

设：卡片 A：“我在考试中作弊”　　卡片 B：“我在考试中没有作弊”

在制作卡片时，两种卡片的比例是已知的，设卡片 A 的比例为 p，则卡片 B 的比例为 1−p。一个有作弊行为的学生，如果抽中的卡片是 A，其真实回答应该是“是”，如果抽中卡片 B，其真实回答应该是“不是”；一个没有作弊行为的学生，他的回答则正好相反。如果将抽取卡片和进行回答视作两个步骤，可以用下图表示这个程序。



设具有卡片 A 特征的人数比例为 π_A，样本容量为 n，其中回答“是”的人数为 n_1，则回答“是”的人数比例为：

$$\frac{n_1}{n}=p(\pi_A)+(1-p)(1-\pi_A)=\pi_A(2p-1)+(1-p)$$

于是，移项求出 π_A 的估计值为：

$$\hat{\pi}_A=\frac{1}{2p-1}\left(\frac{n_1}{n}\right)-\left(\frac{1-p}{2p-1}\right)\qquad p\neq\frac{1}{2}$$

根据数理统计原理，π_A的方差估计量为：

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

置信区间为：$\hat{\pi}_A\pm t\sqrt{V(\hat{\pi}_A)}$

假如在某次对学生调查中我们已取得如下数据：A 卡片的比例 p=3/4，样本容量 n=200，调查中回答“是”的共有 60 人，试用 95%的概率计算学生中考试作弊人数比例的置信区间。计算结果如下：

$$\hat{\pi}_A=\frac{1}{2p-1}\left(\frac{n_1}{n}\right)-\left(\frac{1-p}{2p-1}\right)=\frac{1}{2\times\frac{3}{4}-1}\left(\frac{60}{200}\right)-\left(\frac{1-\frac{3}{4}}{2\times\frac{3}{4}-1}\right)=0.6-0.5=0.1$$

∵F（t）=95%，　∴t=1.96（查表得到）

考试作弊人数比例的点估计值为 10%

$$V(\hat{\pi}_A)=\frac{\hat{\pi}_A(1-\hat{\pi}_A)}{n}+\frac{p(1-p)}{n(2p-1)^2}=\frac{0.1(1-0.1)}{200}+\frac{\frac{3}{4}(1-\frac{3}{4})}{200(2\times\frac{3}{4}-1)^2}=0.0042$$

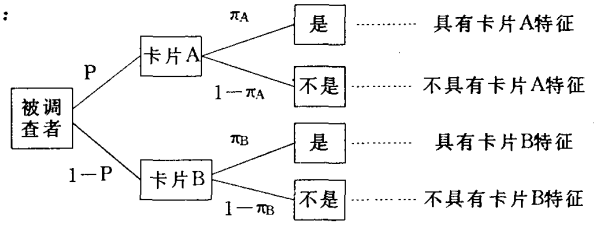
$\therefore \hat{\pi}_A \pm t\sqrt{V(\hat{\pi}_A)}=0.1\pm 1.96\sqrt{0.0042}=0.1\pm 0.127$

总体 $\hat{\pi}_A$ 的置信区间为：$0\leqslant\pi_A\leqslant 0.227$　　（π_A 不能出现负值）

也即，以 95%的把握程度（概率）推断，学生考试作弊的人数比例在 0 至 22.7%之间。

沃纳的随机化回答模型有两个缺点：一是要求被调查者可能回答的两个问题存在相关关系，如上例中卡片 A“我在考试中作弊”和卡片 B“我在考试中没有作弊”是相互关联的两个问题，对此，被调查者仍有可能因怀疑而不予合作；二是两种卡片比例 p 不能各 1/2，否则公式不成立。但是从消除被调查者顾虑的角度看，两种卡片最好各占 1/2，这样抽中两类卡片的机会是均等的。西蒙斯建立的模型对沃纳模型进行了改进，较好地克服了以上两点不足。它的核心是：若在卡片 A 上是敏感性问题，则卡片 B 上是无关的非敏感性问题。例如卡片 A：“我在考试中作弊”；卡片 B：“我上星期在图书馆借了书”。调查的过程和前面相同，被调查者从两种卡片中随机抽取一张，并根据卡片中的问题回答“是”或“不是”。由于使用了无关联的两个问题，故 p 可以等于 1/2，而不会影响估计过程的展开。需要特别指出的是，在采用西蒙斯模型时，必须知道具有非敏感性问题特征的人数比例。因此，在设计时就要考虑这个无关问题是否可以通过其他渠道获得或利用概率推算得到。如“我上星期在图书馆借了书”的学生人数比例可以通过查阅图书馆的借书纪录得到。类似的无关问题还有“我是上半年出生的”，“我身份证号码最后一位是奇数”等。

设卡片 A 的比例为 p（可以是 1/2），卡片 B 的比例为 1−p，π_A 为总体中具有特征 A 人数的比例，π_B 为总体中具有特征 B 人数的比例。则西蒙斯模型中抽取卡片和进行回答的程序如下图所示：



n 为样本容量，n_1 为回答“是”的人数，则调查中回答“是”的人数比例为：

$$\frac{n_1}{n}=p\cdot\pi_A+(1-p)\cdot\pi_B$$

移项后得到 π_A 的估计值：

$$\hat{\pi}_A=\frac{\frac{n_1}{n}-(1-p)\pi_B}{p}$$

π_A 的方差估计量为：$V(\hat{\pi}_A)=\frac{1}{np^2}(\frac{n_1}{n})\cdot(1-\frac{n_1}{n})$

假如在对学生的某次调查后得到如下数据：卡片 A 的比例 p=1/2，样本容量 n=200，调查结果回答“是”的人数比例 n_1 为 30 人。试以 95%的把握程度对作弊学生的比例进行区间估计。

解：已知 p=1/2，n=200，n_1=30，通过查阅图书馆记录，得知上星期中借过书的学生比例约为 1/4，即 π_B=1/4。则利用西蒙斯随机化模型得到：

$$\hat{\pi}_A=\frac{\frac{n_1}{n}-(1-p)\pi_B}{p}=\frac{\frac{30}{200}-(1-\frac{1}{2})\times\frac{1}{4}}{\frac{1}{2}}=0.05$$

π_A 的方差估计量为：

$$V(\hat{\pi}_A)=\frac{1}{np^2}(\frac{n_1}{n})(1-\frac{n_1}{n})=\frac{1}{200(\frac{1}{2})^2}(\frac{30}{200})(1-\frac{30}{200})=0.00255$$

（下转第 38 页）

不大，交易也不够活跃。以 1996 年为例，法国债券成交额达 115 723.96 亿美元，与法国 GDP 的比率为了 752.8%；德国为 21 727.29 亿美元，与 GDP 的比率为 91.8%；英国 16 218.70 亿美元，与 GDP 的比率为 146.2%；美国 58 588.5 亿美元，与 GDP 的比率为 78.9%，[12]而同时期上海证交所国债与企业债成交额为 17 517.54 亿元；深圳证交所成交 636.7 亿元，两地合计 18 154.24 亿元，与当年我国 GDP（68 593.8 亿元）的比率仅为 26%，今后必须大力发展。

5. 加强对地方政府发行债券募集资金使用方面的监管工作。地方政府筹措到的资金必须用于拟投资领域，不得随意更改用途。每隔一定时间（如一季、半年或一年），要向社会公布投资收益情况，接受社会监督。总之，笔者认为地方政府发行债券的时机已经来到，相信随着地方政府债券的陆续发行，地方政府财力的增加，必定会促进地方经济社会文化事业的发展，也有利于整个经济的启动。

注：

①上海财经大学公共政策研究中心：《1999 中国财政发展报告》，上海财经大学出版社，1999 年 2 月第 430 页.

②（美）劳埃德日托马斯：《货币、银行与金融市场》，机械工业出版社 1999 年 5 月第 37 页.

③周正庆主编：《证券知识读本》，中国金融出版社，1998 年 7 月，第 37—38 页.

④⑤⑥上海财经大学公共政策研究中心：《1999 中国财政发展报告》，上海财经大学出版社，1999 年 2 月第 416、355—356、125 页.

⑦陈继兴：《建国以来的广东财政发展》，《财政研究》，1999 年第 4 期.

⑧周正庆主编：《证券知识读本》，1998 年 7 月第 1 版，中国金融出版社第 61 页.

⑨陈景耀：《政府债务对国民经济的作用与影响》，《新华文摘》1999 年第 5 期.

⑩《中国统计年鉴》1998 年.

⑪周正庆主编《证券知识读本》1998 年 7 月第 1 版中国金融出版社第 34 页。

（责任编辑　丁　火）

（上接第 33 页）

置信区间为：$\hat{\pi}_A \pm t\sqrt{V(\hat{\pi}_A)} = 0.05 \pm 1.96\sqrt{0.00255} = 0.05 \pm 0.099$

由此可认为约有 5%的学生在考试中作弊了，若以 95%的概率估计，则置信区间为 0.05±0.099。由于比例不可能为负值，故可以认为作弊学生的比例在 14.9%以下。

使用随机化回答技术的目的，是为了减小或消除被调查者在回答敏感性问题时可能存在的顾虑，完成调查任务。所以，我们在应用这种方法时要注意以下几个问题：首先，调查人员要充分理解和掌握这种方法，这样才能很好地向被调查者作出解释，消除其顾虑；其次，应当允许被调查者在正式调查前检查卡片，了解调查员的记录方式，使其进一步消除疑虑；最后，如果使用西蒙斯模型，要选择好无关的非敏感性问题。做到既要能知道无关的非敏感性问题的总体比例，又要使调查人员无法判别被调查者是回答哪个问题。如前面调查学生考试是否作弊的例子中，如果非敏感性问题为“我是三年级学生”。虽然三年级学生所占比例是可以知道的，但如果调查人员为该校老师，知道被调查学生是否为三年级学生，这个非敏感性问题就起不到掩护的作用。

参考文献：

[1] 隗斌贤、李金昌、徐云庆：《调查统计学》，中国统计出版社，1997 年 2 月.

[2] 赵彦云、贾俊平、金勇进：《社会经济调查方法与应用》，中国统计出版社，1994 年 10 月.

[3] 李金昌、葛菁：《敏感性问题无回答及故意错答的控制研究》，《浙江省第八次统计科学讨论会论文集》，浙江省统计科学研究所，1998 年 8 月。

（责任编辑　丁　火）