四、敏感性问题的处理

在社会经济调查中经常会出现一些敏感性的或高度私人绝密的问题。譬如：收入、吸毒、作弊等敏感性问题，直接调查难以得到真实答案。调研工作者很希望设计一种办法既使被调查者不担心暴露隐私，又使调查者获得正确的资料。

处理方法：

- 第一，随机化回答。沃纳法
- 第二，提出无关的第二个问题，进行随机化回答。西蒙斯的改进

四、敏感性问题的处理

随机化回答技术 (Randomized Response Technique，RRT)

对敏感性问题的调查方案，关键要使被调查者愿意做出真实回答，又能保守个人秘密。一旦调查方案设计有误，被调查者就会拒绝配合，所得调查数据将会失去真实性。心理学家与统计学家为此设计了一种调查方法，一些统计分析方法——随机化回答技术，也应运而生。

随机化回答技术是指在调查中使用特定的随机化装置，使得被调查者以预定的概率P来回答敏感性问题。

这一技术的宗旨就是最大限度地为被调查者保守秘密，从而取得被调查者的信任。

四、敏感性问题的处理

随机化回答技术

比如在调查学生考试作弊的问题中，设计外形完全一样的卡片n张，其中n_1张卡片上写上“你考试是否作过弊?”，$n-n_1$张卡片上写上另外的问题。然后放在一盒子里。调查时，由被调查者从盒子里任抽一卡片，根据卡片上的问题做出回答，至于卡片上具体是什么问题，调查者无权过问。这样就起到了为被调查者保密的作用。因而相对于直接问答调查，易于得到被调查者的合作。

四、敏感性问题的处理

- 两种处理方法：

  1) 随机化的回答。假如要调查对浙工商教学改革的看法，提的问题可以是：

  A：您赞成工商大学提出的教学改革方案吗？

  1. 是 ____________ 2. 不是___________

  B：您不赞成工商大学提出的教学改革方案吗？

  1. 是 ____________ 2. 不是__________

  2）提出无关的第二个问题，进行随机化回答。例如，上例中问题 A 仍为原来的敏感性问题，把问题 B 换成与问题 A 无关的、毫无敏感性的问题。如：“您是一九六四五月份生的吗？”等等。

  这样，被访者的合作态度可能会有所改进。

- 沃纳模型 (Warner Model)
- 西蒙斯模型（Simmons Model）

四、敏感性问题的处理

1.沃纳随机化回答模型

由美国统计学家沃纳提出，具体的做法是：要调查的敏感性问题，列出正反两个问题。如调查考试作弊问题，就作成两种卡片：

A、你在考试中作了弊吗？

B、你在考试中没有作弊吗？

然后由被调查者随机抽取一张来回答“是”或“否”，至于卡片上具体是什么问题，调查者无权过问。因此，调查人员并不知道被调查者在回答哪一个问题，而达到对被调查者个人秘密的保密作用。

四、敏感性问题的处理

1.沃纳随机化回答模型

要求被调查者从中随机抽取一个回答而调查人员不知道其具体抽中的是哪一个问题，但问题A的比例P是确定的。如果他所抽到的问题与自己情况一致则回答“是”，否则回答“不是”。

设π_A是具有敏感性特征的人所占的比例，p是写有问题“你属于A吗?”的卡片所占的比例。如果对n人进行调查，调查结果中有n_1个人回答“是”，有$n-n_1$个人回答“否”，统计结果中回答“是”的人的比例$\phi=n_1/n$，对问题A回答“是”的人数比例为π_A。

四、敏感性问题的处理

公式：

π_A 的极大似然估计为：

$$\hat{\pi}_A = \frac{\frac{n1}{n} - (1-P)}{2P-1} \quad (P \neq 1/2)$$

其方差为：

$$Var(\hat{\pi}_A) = \frac{\frac{n1}{n}(1-\frac{n1}{n})}{n(2P-1)^2}$$

四、敏感性问题的处理

举例：

印度教育当局研究大学生中酗酒的流行程度。如果一个学生在调查前一个月内饮酒至少1250毫升，则称他（她）是一个酗酒者。

在这个定义下，从加尔各答市大学生中简单随机地抽取了若干名大学生，目标是估计加尔各答大学中酗酒者所占的比例 π_A。

四、敏感性问题的处理

举例：

写有敏感问题A的卡片占全部卡片的比例为 p=0.75，回答“是”的人数占总人数的比例 $\phi=n_1/n=0.28$

随机抽取了100名大学生，所用随机化装置为一装有60个卡片的盒子。盒子中有45张卡片上写有问题“在上一个月你是否至少饮酒1250毫升?”（问题A），剩余的15张卡片上写有问题“在上一个月内你是否饮酒少于1250毫升?”

调查时，在没有调查员观察的情况下，被调查者把盒子的卡片摇匀后从中随机抽取一张，而后根据所抽到的卡片上的问题如实地回答“是”或“不是”。调查结果为：有28个人回答了“是”，72个人回答“不是”。

四、敏感性问题的处理

有：n=100，n_1=28，p=0.75，因此有：

$$\hat{\pi}_A = \frac{\frac{n1}{n} - (1 - P)}{2P - 1}$$

=[0.28-0.25]/0.5=0.06

也即有6%的人是酗酒者。

根据

$$Var(\hat{\pi}_A) = \frac{\frac{n1}{n}(1 - \frac{n1}{n})}{n(2P-1)^2}$$ =0.008145

这样应用第一种方法估计统计对问题一回答“是”的比例为：**6%**。标准差的估计值为：$\sqrt{0.008145}$ 即**0.09**。

四、敏感性问题的处理

2.西蒙斯模型（Simmons model）

沃纳的方法虽然比直接提出敏感性问题要好，但所提的两个问题都还具有敏感性。而且，该方法中回答A的人数比例不能为1/2。1967年西蒙斯对沃纳模型进行了改进。

他所建立的模型与沃纳模型最大的不同点：在于调查人员提出的随机化问题是两个不相关的问题，其中一个为敏感性问题，另一个为非敏感性问题B，这样的处理使被调查者的合作态度进一步提高。

四、敏感性问题的处理

西蒙斯模型（Simmons model）

设样本中对问题B（无关问题）回答“是”的人数比例为 π_B

ϕ 和 π_A 的含义仍同上。

对问题**A**回答“是”的人数比例

统计结果中回答“是”的人的比例，也就是对问题**A**或**B**回答“是”的人数比例

四、敏感性问题的处理

2.西蒙斯模型（Simmons model）

(1) π_B **已知的情况。**设抽样方式是简单随机有放回的

π_A 是具有敏感性特征A的人所占的比例。

设总体为n的简单随机样本中，有n_1人回答“是”，则

$\hat{\phi} = n_1/n$ ，从而得到则 π_A 的极大似然估计为：

$$\hat{\pi}_A = \frac{\hat{\phi} - \pi_B(1-P)}{P}$$

其方差为：

$$Var(\hat{\pi}_A) = \frac{\phi(1-\phi)}{nP^2}$$

四、敏感性问题的处理

实践中，π_B并不总是已知的，例如对于无关问题“你是四月份出生的吗?”我们可以通过查有关资料来获得 π_B 的值，而对于无关问题“你喜欢蓝色吗?”我们就无法预知 π_B的值，此时 π_B就是未知的。因此有必要对

π_B 未知的情况进行讨论。

四、敏感性问题的处理

(2) π_B 未知的情况

这时需要抽取两个随机样本进行调查。设这两个样本的容量分别为 n_1 和 n_2

敏感性问题占的比例分别为 P_1 和 P_2

假设总体1中回答敏感性问题的人占的比例为 P_1 时

对问题A或B作出“是”的答复者所占的比例为 ϕ_1

假设总体2中回答敏感性问题的人占的比例为 P_2

对这两个问题作出“是”的答复者的比例为 ϕ_2

四、敏感性问题的处理

从而得到该敏感问题的估计回答：

$$\hat{\pi}_A = \frac{\phi_1(1-P_2)-\phi_2(1-P_1)}{P_1-P_2}$$

其方差为：

$$Var(\hat{\pi}_A) = \frac{1}{(P_1-P_2)^2}\left[\frac{\phi_1(1-\phi_1)(1-P_2)^2}{n_1}+\frac{\phi_2(1-\phi_2)(1-P_1)^2}{n_2}\right]$$

四、敏感性问题的处理

敏感性问题调查方法的应用

某高校在开展关于严肃考风考纪的活动中，要求对学校的学生是否有过作弊行为这一问题进行抽样调查。该调查问题具有敏感性，运用通常采取的调查方式，调查根本无法进行，因此运用了敏感性问题抽样调查方法，该高校在校生人数为6000人，随机抽取1500名学生进行抽样调查，且分别运用了以上的两种方法，比较统计结果。

四、敏感性问题的处理

第一种方法：

提出两个都具有敏感性相关问题

采用随机化的回答技术设计了两种用信封封装比例一定的问卷，一种问题为：“你有过作弊行为吗？”；另一种问题为：“你没有过作弊行为吗？”。在调查时，让同学任意选取一个信封并回答上面的问题，当然调查人员是不知道该同学回答的是哪一个问题。

四、敏感性问题的处理

第一种方法：

提出两个都具有敏感性相关问题

这样同学们根据他们的实际情况回答抽到的问题，与自己的情况一致的则回答“是”；否则回答“不是”。研究者在设计问卷时，设计第一种问题占60%，这样两个问题所占的比例比较接近，有助于让被调查者消除顾虑，我们对收回的问卷进行统计，结果对两种问题回答“是”的有638人，占样本的比例为：

$$\phi = 638/1500 = 0.4253$$

四、敏感性问题的处理

已知 ϕ **=638/1500=0.4253，*P*=60%**

将它代入 $$\hat{\pi}_A = \frac{\hat{\phi} - (1 - P)}{2P - 1} \quad (P \neq 1/2)$$

得到回答第一种问题为“是”的人数估计比例：

$$\hat{\pi}_A = \frac{0.4253 - (1 - 0.6)}{2 \times 0.6 - 1} = 0.1265$$

其方差为：

$$Var(\hat{\pi}_A) = \frac{\phi(1-\phi)}{n(2P-1)^2} = \frac{0.4253 \times (1-0.4253)}{1500(2 \times 0.6 - 1)^2} = 0.0041$$

这样应用第一种方法估计统计对问题一回答“是”的比例为：**12.65%**。标准差的估计值为：$\sqrt{0.0041}$ 即 **0.064**。

四、敏感性问题的处理

第二种方法：提出的两个问题，一个为敏感性问题，另一个为与调查内容无关的非敏感性问题。

我们同样采用随机化的回答技术设计了两种用信封封装比例一定的问卷，然而一种问题为：“你有过作弊行为吗？”；为了统计的方便我们选择了π_B 已知的情况,即另一种问题设计为：“你是四月份出生的吗？”。

π_B　　　　p

显然，第二个问题与我们所要调查的问题无关，而且被调查同学当中是四月份出生的比例我们可以很容易从学校教务处学生信息中心收集到，经统计该校学生中四月份出生者所占的比例为15.38%。其中设计的问卷中第一种问题同样占60%，统计结果为对两种问题回答“是”的有206人，占样本的比例为ϕ

四、敏感性问题的处理

ϕ =**206/1500=0.1373**，将它代入 $\hat{\pi}_A = \dfrac{\hat{\phi} - \pi_B(1-P)}{P}$

得到回答第一种问题为“是”的人数估计比例：

$$\hat{\pi}_A = \frac{0.1373 - 0.1538(1-0.6)}{0.6} = 0.1263$$

其方差为：

$$Var(\hat{\pi}_A) = \frac{\phi(1-\phi)}{nP^2} = \frac{0.1373 \times (1-0.1373)}{1500 \times (0.6)^2} = 0.0002$$

采用第二种方法得出的对问题一回答为“是”的估计比例为：**12.63%**。标准差的估计值为 $\sqrt{0.0002}$ 即**0.014**

可以看出采用两种调查技术，最终得出的结果是接近的