敏感性问题调查的一种定量研究方法——谈随机乘法模型

李红梅
（北方工业大学理学院，北京，100041）

摘要：在敏感性问题的市场调查中，人们对随机化回答技术的要求越来越高，本文在乘法模型的基础上提出一改进模型——随机乘法模型，不仅使估计量的精度有一定的保证，而且操作简便，估计量形式简单，同时又对被调查者有一定的保护，该模型在敏感性问题数量特征的调查中，是一种比较理想的方法。

关键词：敏感性问题，随机化回答技术，乘法模型，随机乘法模型

1. 引言

随着社会经济的发展，人们对信息范围与质量的要求逐步提高，在市场调查中，涉及到的敏感性问题也越来越多。在敏感性问题的调查中，如果采用一般的直接询问方式进行调查，是根本无法获得真实数据。拒绝回答和故意给出不真实信息是产生样本估计偏差的非抽样误差的两种主要来源。为了克服这种搜集数据的困难，1965 年沃纳提出沃纳模型，从而开创了随机化回答技术的先河。之后的 40 年，随机化回答技术日臻完善，已经广泛地应用于各种敏感性问题的调查研究中，该领域的研究学者不断推陈出新，产生了一系列的模型。然而，在这众多的随机化回答模型中，并非都是十分完善的，也存在优劣之分。综合来看，一个优良的随机化回答模型，应满足以下几点要求：

1、随机化装置操作简便

2、估计量形式简单

3、估计精度有一定保证

4、能够最大限度地保护被调查者隐私

基于以上条件，本文在介绍乘法模型的基础上，提出一个改进模型，可以称之为随机乘法模型，是一种操作简单，行之有效的方法，具有一定的推广应用价值。

2. 乘法模型介绍

2.1 模型的设计及参数的估计

在乘法模型中，要求回答者将其敏感特征数值(X)与一个来自已知分布的随机数值(Y)相乘。观测值 Z 可被表示为：Z=XY （X、Y 相互独立），故

其均值为：$\mu_Z = \mu_X \mu_Y$，方差为：$\sigma_Z^2 = \mu_Y^2\sigma_X^2 + \mu_X^2\sigma_Y^2 + \sigma_X^2\sigma_Y^2$。

于是 μ_X 的估计量为：$\hat{\mu}_{X0} = \hat{\mu}_Z / \mu_Y$。

在无放回简单随机抽样条件下，估计量方差为：

$$Var(\hat{\mu}_{X0}) = \frac{\sigma_Z^2}{n\mu_Y^2} \times \frac{N-n}{N-1} = \frac{(1-f)S_Z^2}{n\mu_Y^2}.$$

因为 $Z_1, Z_2, \cdots, Z_n$ 为简单随机样本，所以我们可取估计量：

$\hat{\mu}_Z = \bar{z}, \hat{S}_Z^2 = s_z^2$，且容易证明 $\hat{\mu}_Z, \hat{S}_Z^2$ 分别为 μ_Z, S_Z^2 的无偏估计

以样本均值 $\bar{z}$ 代替 $\hat{\mu}_Z$ 得：$\hat{\mu}_{X0} = \bar{z} / \mu_Y$

估计量方差的无偏估计为：$\hat{Var}(\hat{\mu}_{X0})=\frac{(1-f)s_z^2}{n\mu_Y^2}$

2.2 乘法模型评价

乘法模型通过引入另一无关变量Y使回答结果Z与敏感性问题答案X产生较大差异，从而较大限度地保护了被调查者的隐私。同时该模型随机化装置及估计量的形式都比较简单，这是该模型的优点。但是，一般情况下，该模型精度较差，这点通过与一般简单估计的比较可以发现。

当Y=1时，乘法模型即为一般的简单估计，其方差为：$Var(\hat{\mu}_X)=\frac{\sigma_X^2}{n}\times\frac{N-n}{N-1}$

乘法模型与一般简单估计方差之比为：

$$\frac{Var(\hat{\mu}_{X0})}{Var(\hat{\mu}_X)}=1+\frac{\mu_X^2\sigma_Y^2}{\mu_Y^2\sigma_X^2}+\frac{\sigma_Y^2}{\mu_Y^2}=1+(\frac{\mu_X^2}{\sigma_X^2}+1)(\frac{\sigma_Y}{\mu_Y})^2$$

显然，乘法模型估计精度低于一般简单估计，通常情况下，乘法模型的抽样估计误差会因引入了无关变量而显著地增加。提高乘法模型估计精度的有效办法是降低随机变量Y的变异系数，然而这又与容易引起被调查者怀疑产生矛盾。能否在不改变无关问题的前提下，降低引入无关问题的影响程度呢？下面介绍的随机乘法模型便很好地解决了这个问题。

3. 随机乘法模型介绍

3.1 模型的设计与参数的估计

随机乘法模型是在乘法模型的基础上提出的一个改进模型，它的基本设计思路如下：

1）通过计算机产生一均值为μ_Y的随机数Y，可以适当界定Y值的范围。

2）在一个容器内放置若干个外观完全相同的小球，小球上面分别标有数字1或0。其中标有数字1的小球所占的比例为P。

3）被调查者从容器内随机抓取一小球，如果小球号码为1,则被调查者回答敏感性问题X与μ_Y的乘积，即$X\mu_Y$；号码为0时，要求被调查者回答敏感性问题X与随机产生的Y值的乘积，即XY。

抓取小球与产生随机数的过程是隐蔽的，研究者只能看到被调查者给出的最终回答Z。

抓取小球的过程可以用符合0，1分布的随机变量ε来表示，抽到1号小球，表示为$\varepsilon=1$，抽到0号小球，表示为$\varepsilon=0$，$p(\varepsilon=1)=P$，并且X,Y与ε相互独立。

用模型表示,则得：$Z=\varepsilon X\mu_Y+(1-\varepsilon)XY$

两边取期望,由X,Y与ε的独立性及条件期望的概念,得：

$$\mu_z=P\mu_X\mu_Y+(1-P)\mu_X\mu_Y=\mu_X\mu_Y$$

$$\begin{aligned}\sigma_z^2&=E[\varepsilon X\mu_Y+(1-\varepsilon)XY]^2-[E(\varepsilon X\mu_Y+(1-\varepsilon)XY)]^2\\&=P\mu_Y^2E(X^2)+(1-P)E(X^2)E(Y^2)-(\mu_X\mu_Y)^2\\&=\mu_Y^2\sigma_X^2+(1-P)(\sigma_X^2\sigma_Y^2+\mu_X^2\sigma_Y^2)\end{aligned}$$

于是μ_X的估计量为：$\hat{\mu}_{X1}=\hat{\mu}_Z/\mu_Y$

以样本均值 $\bar{z}$ 代替 $\hat{\mu}_Z$ 得：$\hat{\mu}_{X1}=\bar{z}/\mu_Y$，

在无放回简单随机抽样条件下，估计量方差为：

$$Var(\hat{\mu}_{X1})=\frac{\sigma_Z^2}{n\mu_Y^2}\times\frac{N-n}{N-1}=\frac{(1-f)S_Z^2}{n\mu_Y^2}$$

估计量方差的无偏估计为：$\hat{Var}(\hat{\mu}_{X1})=\dfrac{(1-f)s_Z^2}{n\mu_Y^2}$

3.2 随机乘法模型评价

随机乘法模型吸收了乘法模型的优点，模型随机化装置及估计量的形式都比较简单，同时还能够较大限度地保护被调查者的隐私。然而，在估计精度方面，是否有显著提高呢？我们先进行以下比较：

同样，当 Y=1 时，随机乘法模型即为一般的简单估计

随机乘法模型与一般简单估计方差之比为：

$$\frac{Var(\hat{\mu}_{X1})}{Var(\hat{\mu}_X)}=1+(1-P)(\frac{\mu_X^2\sigma_Y^2}{\mu_Y^2\sigma_X^2}+\frac{\sigma_Y^2}{\mu_Y^2})=1+(1-P)(\frac{\mu_X^2}{\sigma_X^2}+1)(\frac{\sigma_Y}{\mu_Y})^2$$

显然，由于无关问题的引入，随机乘法模型估计精度仍然会低于一般简单估计，但通过与乘法模型方差的对比可以看出，比值大于 1 的部分多出一个系数（1-P），从而使随机乘法模型的估计精度比乘法模型有了显著的提高。提高随机乘法模型估计精度不仅可以通过降低随机变量 Y 的变异系数而且可以提高 P 的值来实现。可见，随机乘法模型不仅使估计量的精度有一定的保证,而且操作简便，估计量形式简单，同时又对被调查者有一定的保护，所以该模型在敏感性问题数量特征的调查中，是一种比较理想的方法。

4. 随机乘法模型的应用

为调查某地区高校教师隐性收入状况，设计敏感性问题 X：“你平均每月的隐性收入数额大概是多少？”用随机乘法模型进行估计。首先通过计算机编程，设计 $\mu_Y=68$ 的随机数 Y，为保证一定精度，可以界定 Y 值的范围，比如 Y 在[0，136]之间，并确定标有数字 1 的小球所占的比例为 P=0.7。然后让被调查者从容器内随机抓取小球，并按上述规则回答问题。

已知在该地区范围内采用无放回的简单随机抽样抽选了 1000 名高校教师，调查结果如下：$\bar{z}=53175$，$s_z^2=9.2965\times10^8$。所以，$\hat{\mu}_{X1}=\dfrac{\bar{z}}{\mu_Y}=\dfrac{53175}{68}\approx782$（元），

$$\hat{Var}(\hat{\mu}_{X1})=\frac{(1-f)s_Z^2}{n\mu_Y^2}\approx\frac{9.2965\times10^8}{1000\times68^2}\approx201.0489,\quad\sqrt{\hat{Var}(\hat{\mu}_{X1})}\approx14.1792\text{（元）}$$

取 $\alpha=0.05$，即 F(t)=95%时，t=1.96，可确定隐性收入的区间范围为：

[782-1.96×14.1792，782+1.96×14.1792]，即[754，810]元

即，可以以 95%的概率把握程度保证该地区高校教师的隐性收入平均每人每月在[754，810]元范围之内。