DOI: 10.13860/j.cnki.sltj.2000.03.017

文章编号: 1002- 1566(2000)03- 0061- 03

多项选择随机化调查的多样本模型

—— 敏感性问题调查方法(III)

孙明举[1] 段 钢[2] 孙山泽[3]

(1.信息产业部电信规划研究院,北京,100037; 2.电子科技大学,成都,610054; 3.北京大学,北京,100871)

摘要: 敏感性问题调查中常会碰到问题回答为多项选择的情况,本文介绍多样本调查模型.

中图分类号: O212 C8 S3 文献标识码: A

在实际工作中,常会碰到总体可划分多于两类的情况,如调查某厂职工对领导的满意程度,职工可分为"满意","一般","不满意"三种情况。其中"一般",与"不满意"。具有不同程度的敏感性,下面就讨论同一敏感性特征可分为 $t(t\geqslant 2)$项(其中至少有 1项至多有 $t-1$项具有不同程度的敏感性)情况的调查,本文介绍多样本模型。

先介绍 $t=3$的情况。设总体中的每一个体属于而且只能属于三个互相排斥的 A B C类之一,三类中至少有一类,至多有两类具有敏感性,我们的目的是通过抽样调查来估计三类个体在总体中所占的比例。用简单随机有放回的方式从总体中抽取两个独立但又互不相交的样本,容量分别为 n_1和 n_2(n_1、n_2不必相等)。调查中,两个样本使用类似的随机化装置,不同的只是随机化装置的参数设计。如两个装有卡片的盒子,分别用于两个样本中。每一盒子中有三种类型的卡片:一类卡片上写有问题"你属于 A吗?",第二类上写有"你属于 B吗?"第三类则为"你属于 C吗?"三类卡片按预先设计的比例混合好装入盒子中,但两个盒子中同一类型的卡片所占的比例又不相同。且每一盒子中三种类型的卡片所占的比例不是$\frac{1}{3}$。

调查中,被调查者将盒子中的卡片充分混合,然后随机地从中抽取一张,在没有被调查员看到的情况下按卡片上的问题作出回答。答案只能为"是"或"否"。调查员将其回答记录下来,而后被调查者把卡片放回盒子,混合好以后再将盒子交还调查员。

设 c_1, c_2, c_3分别是属于 A B C的人在总体中所占的真实比例。则有

$$c_1+c_2+c_3=1$$

设 p_{ij}是用于第 $i(i=1,2)$个样本的盒子中第 $j(j=1,2,3)$类卡片所占的比例。

有: $\sum_{j=1}^{3} p_{ij}=1 \qquad i=1,2$

$(p_{11}-p_{13})(p_{22}-p_{23})\neq(p_{12}-p_{13})(p_{21}-p_{23})$

令 $X_{ir}=\begin{cases}1 & \text{第 } i \text{ 个样本的第 } r \text{ 个人回答是}\\ 0 & \text{第 } i \text{ 个样本的第 } r \text{ 个人回答否}\end{cases}$

$i=1,2, r=1,2,\ldots,n_1$(对于 $i=1$) $r=1,2,\ldots n_2$(对于 $i=2$)

第一个样本中第 r个被调查者回答"是"的概率为:

$$P(x_{1r}=1)=\sum_{j=1}^{3} p_{1j}c_j=(p_{11}-p_{13})c_1+(p_{12}-p_{13})c_2+p_{13}=\lambda_1 \tag{1}$$

回答"否"的概率为:

$$P(x_{1x}=0)=1-\sum_{j=1}^{3} p_{1j}c_j=1-\lambda_1 \tag{2}$$

类似地,对于第二个样本,有:

$$P(x_{2r}=1)=\sum_{j=1}^{3} p_{2j}c_j=(p_{21}-p_{23})c_1+(p_{22}-p_{23})c_2+p_{23}=\lambda_2 \tag{3}$$

设 n_{11}是第一个样本中回答"是"的人数,则 n_1-n_{11}就是回答"否"的人数,n_{21}是第一个样本中回答"是"的人数,则 n_2-n_{21}就是回答"否"的人数。从而可以求得 c_1、c_2的极大似然估计值分别为:

$$\hat{\pi}_1 = \frac{(\hat{\lambda}_1 - p_{13})(p_{22} - p_{23}) - (\hat{\lambda}_2 - p_{23})(p_{12} - p_{13})}{(p_{11} - p_{13})(p_{22} - p_{23}) - (p_{12} - p_{13})(p_{21} - p_{23})} \tag{4a}$$

$$\hat{\pi}_2 = \frac{(\hat{\lambda}_1 - p_{13})(p_{21} - p_{23}) - (\hat{\lambda}_2 - p_{23})(p_{11} - p_{13})}{(p_{11} - p_{13})(p_{22} - p_{23}) - (p_{12} - p_{13})(p_{21} - p_{23})} \tag{4b}$$

其中: $\hat{\lambda}_1 = \frac{n_{11}}{n_1}$, $\hat{\lambda}_2 = \frac{n_{21}}{n_2}$

则 π_3 的一个估计量为:

$$\hat{\pi}_3 = 1 - \hat{\pi}_1 - \hat{\pi}_2 \tag{4c}$$

又因为 $\hat{\lambda}_1$,$\hat{\lambda}_2$ 分别服从参数为 (λ_1, n_1), (λ_2, n_2)的二项分布,故易知有:

$$E(\hat{\pi}_1) = \pi_1, E(\hat{\pi}_2) = \pi_2, E(\hat{\pi}_3) = \pi_3$$

即 π_1, π_2, π_3 的极大似然估计是无偏的,且有:

$$Var(\hat{\pi}_1) = \frac{1}{K^2}\left[(p_{22} - p_{23})^2 \frac{\lambda_1(1-\lambda_1)}{n_1} + (p_{12} - p_{13})^2 \frac{\lambda_2(1-\lambda_2)}{n_2}\right]$$

$$Var(\hat{\pi}_2) = \frac{1}{K^2}\left[(P_{21} - P_{23})^2 \frac{\lambda_1(1-\lambda_1)}{n_1} + (P_{11} - P_{13})^2 \frac{\lambda_2(1-\lambda_2)}{n_2}\right]$$

$$Var(\hat{\pi}_3) = \frac{1}{K^2}\left[(P_{22} - P_{21})^2 \frac{\lambda_1(1-\lambda_1)}{n_1} + (P_{12} - P_{11})^2 \frac{\lambda_2(1-\lambda_2)}{n_2}\right]$$

其中: $K = (P_{11} - P_{13})(P_{22} - P_{23}) - (P_{12} - {}_{13})(P_{21} - P_{23})$

易知 $Var(\hat{\pi}_1), Var(\hat{\pi}_2), Var(\hat{\pi}_3)$的一个无偏估计量分别是:

$$\hat{Var}(\hat{\pi}_1) = \frac{1}{K^2}\left[(p_{22} - p_{23})^2 \frac{\lambda_1(1-\lambda_1)}{n_1} + (p_{12} - p_{13})^2 \frac{\hat{\lambda}_2(1-\hat{\lambda}_2)}{n_2}\right]$$

$$\hat{Var}(\hat{\pi}_2) = \frac{1}{K^2}\left[(p_{21} - p_{23})^2 \frac{\hat{\lambda}_1(1-\hat{\lambda}_1)}{n_1 - 1} + (p_{11} - p_{13})^2 \frac{\hat{\lambda}_2(1-\hat{\lambda}_2)}{n_2 - 1}\right] \tag{6}$$

$$\hat{Var}(\hat{\pi}_3) = \frac{1}{K^2}\left[(p_{22} - p_{21})^2 \frac{\hat{\lambda}_1(1-\hat{\lambda}_1)}{n_1 - 1} + (p_{12} - p_{11})^2 \frac{\hat{\lambda}_2(1-\hat{\lambda}_2)}{n_2 - 1}\right]$$

考虑 $t > 3$的情况, π_i $(i = 1, 2, \cdots, t)$为具有第 i类敏感性特征的人在总体中所占的真实比例,则有:$\sum_{i=1}^{t} \pi_i = 1$我们的目的是估计 π_i $(i = 1, 2, \ldots, t)$的值。

由 $\pi_t = 1 - \sum_{i=1}^{t-1} \pi_i$,故只需估计出前 $t - 1$个即可,令 $s = t - 1$,抽取 s 个简单随机有放回的样本,令 P_{ij}是用第 i $(i = 1, 2, \ldots, s)$个样本的盒子中第 j $(j = 1, 2, \ldots t)$类卡片所占的比例,有:

$$\sum_{j=1}^{t} p_{ij} = 1 \qquad i = 1, 2, \ldots, s$$

行列式 $|P| \neq 0$

$$P = \begin{bmatrix} p_{11} - p_{1t} & p_{12} - p_{1t} & \ldots & p_{1s} - p_{1t} \\ \text{行列式 } |P| \neq 0 & & & \\ p_{21} - p_{2t} & p_{22} - p_{2t} & \ldots & p_{2s} - p_{2t} \\ \vdots & \vdots & \vdots & \vdots \\ p_{s1} - p_{st} & p_{s2} - p_{st} & \ldots & p_{ss} - p_{st} \end{bmatrix} \tag{7}$$

设

$$X_{ij} = \begin{cases} 1 & \text{第 } i \text{ 个样本中第 } r \text{ 个人回答“是”} \\ 0 & \text{第 } i \text{ 个样本中第 } r \text{ 个人回答“否”} \end{cases}$$

假设所有被调查者的回答都是真实的,第 i个样本中第 r个人回答“是”的概率为:

$$P(X_{ir} = 1) = p_{i1}\pi_1 + p_{i2}\pi_2 + \cdots + p_{it}\pi_t = \lambda_i$$

第 i 个样本中第 r 个人回答“否”的概率为:

$$P(X_{ir}=0)=1-\sum_{j=1}^{t}p_{ij}\pi_j=1-\lambda_i$$

设第 i 个样本的容量为 n_i, n_{i1} 为第 i 个样本中回答“是”的人数, n_i-n_{i1} 为回答“否”的人数,则有:

$$\hat{\pi}=P^{-1}c \qquad \hat{\pi}_t=1-\sum_{i=1}^{s}\hat{\pi}_i$$

其中:矩阵 P 的表示式见(7)

$$\hat{\pi}=\begin{bmatrix}\hat{\pi}_1\\ \hat{\pi}_2\\ \cdot\\ \cdot\\ \hat{\pi}_s\end{bmatrix} \qquad c=\begin{bmatrix}\frac{n_{11}}{n_1}-p_{1t}\\ \frac{n_{21}}{n_2}-p_{2t}\\ \cdot\\ \cdot\\ \cdot\\ \frac{n_{s1}}{n_s}-p_{st}\end{bmatrix}$$

且易知: $E(\hat{\pi})P^{-1}E(c)=\pi$

$$Var(\hat{\pi})=Var(P^{-1}c)=P^{-1}\begin{bmatrix}V_{1,1} & & \\ & \ddots & \\ & & V_{t-1\ t-1}\end{bmatrix}(P^{-1})'$$

其中 $v_{ii}=\lambda_i(1-\lambda_i)/n_i$

例:在一次婚前怀孕流行情况的调查中,调查对象被分为互相排斥的三类 A B C,描述如下:

A:结婚时受孕的妇女

B:怀孕期间结婚的妇女

C:婚前有生育的妇女

她们在总体中所占比例分别为 π_1、π_2、π_3 为估计 π_1、π_2、π_3 的值,从女性中抽取了两个容量分别为 180 200的独立的简单随机样本,随机化装置是装有 100张卡片的盒子,用于第 1、2个样本的盒子中卡片的比例分别为: 40 30 30与 30 30 40,要求被调查者在调查者没有观察到的情况下从盒子中抽一卡片,并根据自己的真实情况对卡片上的问题回答“是”或“不是”,再将卡片放回盒中摇匀,两个样本中回答“是”的人数分别为 70, 61.

注意到此时有: $t=3, s=2, n_1=180, n_2=200$

$p_{11}=0.4,\quad p_{12}=p_{22}=0.3,\quad p_{23}=0.4$

$$P=\begin{bmatrix}0.1 & 0\\ -0.1 & -0.1\end{bmatrix} \qquad P^{-1}=\begin{bmatrix}10 & 0\\ -10 & -10\end{bmatrix}$$

$n_{11}=70, n_{12}=61, \hat{\lambda}_1=0.3889, \hat{\lambda}_2=0.3050, c=(0.0889,-0.0950)'$

则 $\pi=(\pi_1,\pi_2)'$ 的一个无偏估计为:

$$\hat{\pi}=(0.889, 0.061)'$$

从而知 $\hat{\pi}_3=1-\hat{\pi}_1-\hat{\pi}_2=0.050$

$Var(\hat{\pi}_1)=0.133 \qquad Var(\hat{\pi}_2)=0.239$

$$Var(\hat{\pi}_3)=Var(1-\hat{\pi}_1-\hat{\pi}_2)=0.106$$