DOI: 10.13860/j.cnki.sltj.2000.01.014

文章编号: 1002-1566(2000)01-0058-07

二项选择敏感性问题调查的基本方法
——敏感性问题调查方法(Ⅰ)

孙山泽[1] 孙明举[1] 段 钢[2]

(1.北京大学概率统计系,北京,100871; 2.电子科技大学,成都,610054)

摘 要

孙山泽,孙明举.二项选择敏感性问题调查的基本方法.

本讲座介绍敏感性问题调查的随机化回答技术(RRT),本文首先介绍二项选择问题的基本方法:沃纳模型,西蒙斯模型.

关键词: 敏感性问题,随机化回答,沃纳模型,西蒙斯模型

中图分类号: O212 C8 C3 文献标识码: A

一、敏感性问题与随机化回答技术(RRT)

抽样调查现在已被广泛应用. 它省时省力,能获得较为准确的结果,这一方面是由于方法本身的科学性;但另一方面很重要的一个前提是被调查者的回答必须都是真实的. 在当今的社会经济调查等各种统计调查中,经常会遇到各种各样的敏感性问题. 所谓敏感性问题,是指与个人或单位的隐私或私人利益有关而不便向外界透露的问题. 比如,个人或单位是否偷税漏税及数额的多少;考生在考试中是否有作弊行为;吸毒、赌博;个人储蓄的多少;是否参加过走私货物的交易;是否有犯罪行为;各种类型的额外消费、公款吃喝;同性恋及类似的为社会所不赞成的各种事件. 对于这类敏感性问题,调查中若采用直接问答的方式,被调查者为了保护自己的隐私或出于其他目的,往往会拒绝回答或故意做出错误的回答. 这样就破坏了数据的真实性,而且破坏程度的大小无法度量. 如调查个体营业者偷税漏税情况,采用抽样调查的方法,若直接询问"你在上个月是否有偷税漏税行为?偷税漏税的数额是多少?"因为偷税漏税是违法行为,对于第一个问题,往往会得到否定回答. 即使有人敢于回答"是",在第二个问题中,回答的数额肯定会比其真实的偷税漏税数额低. 这样总的调查结果就没有可靠性.

总的来说,对于敏感性问题,若采用直接调查的方法,调查者将难以控制样本信息,得不到可靠的样本数据. 为了得到敏感性问题的可靠的样本数据,有必要采用一种科学的可行的技术——随机化回答技术(Randomized Response Technique简记为 RRT).

随机化回答,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. 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 m个,其中 m_1 个卡片上写

* 国家教委博士点基金资助(96000151)项目
收稿日期: 1999-10-15.

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www

上“你考试是否作过弊?”, m- m_1 个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,回答完毕再把卡片放回盒子。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。

随机化回答技术是 1965年沃纳(Warner)提出了沃纳模型(Warner model)后才发展起来的。

敏感性问题按总体的特征可分为两类: 属性特征的敏感性问题和数量特征的敏感性问题。

属性特征的敏感性问题(例如考生是否作弊,是否有吸毒行为等)是指被调查者是否具有敏感性问题的特征,一般是估计具有敏感性特征的人在总体中所占的比例,因此又可称作敏感性比例问题。

数量特征的敏感性问题是指被调查者具有敏感性问题数额的多少的特征,一般是估计敏感性数额的均值或总和,也可称作敏感性均值问题。比如个体户或企业偷税漏税数额多少的问题: 一段时间内吸毒次数的问题;职工额外收入问题等。

本文介绍二项属性特征的敏感性问题的随机化回答模型

二、沃纳模型(Warner model)

这一模型是 1965年由 Warner提出的,它的提出开创了随机化回答的先河。其设计原则是根据敏感性特征设计两个相互对立的问题,让被调查者按预定的概率从中选一个回答,调查者无权过问被调查者究竟回答的是哪一个问题,从而起到了为被调查者保密的效果。

(1)模型的设计及参数的估计:

设总体可分为互不相容的两类: 具有敏感性特征的一类 A与不具敏感性特征的一类 $\bar{A}$ 即总体中的每一个体或者具有敏感性特征(属于 A),或者不具有敏感性特征(属于 $\bar{A}$)。我们的目的是估计具有敏感性特征(属于 A)的人在总体中所占的比例 π_A。

在 SRSWR(简单随机有放回抽样)下从总体中抽得 n个样本,然后对这 n个样本进行随机化回答调查。所使用的随机装置描述如下: 外形相同的卡片上分别写有问题: “你属于 A吗?”与“你属于 $\bar{A}$吗?”如(你在考试中作弊了吗?”与“你在考试中没有作弊吗?”)以预定的比例 p混合后放入一盒子中,调查时,被调查者从盒子中任拿出一卡片,根据卡片上的问题进行回答。回答完后仍把卡片放回盒子,供其他被调查者使用。

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。

$$X_i=\begin{cases}1 & \text{若被调查者回答“是”}\\ 0 & \text{若被调查者回答“否”}\end{cases}$$

假设所有被调查者的回答都是真实的。则:

$$P(X_i=1)=\pi_A P+(1-\pi_A)(1-p) \qquad i=1,2,\cdots n$$

$$P(X_i=0)=(1-\pi_A)P+\pi_A(1-p) \qquad i=1,2,\cdots n$$

设调查结果中有 m_1 个人回答“是”,有 n- m_1 个人回答“否”。则 π_A 的极大似然估计为:

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

易知 $\hat{\pi}_A$ 是 π_A 的无偏估计。其方差为:

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www

$$Var(\hat{c}_A) = Var\{[\frac{n_1}{n} - (1-p)]/(2p-1)\} = \frac{nVar(X_i)}{(2p-1)^2n^2}$$
$$= \frac{[c_A p + (1-c_A)(1-p)]\cdot[(1-c_A)p + c_A(1-p)]}{(2p-1)^2 n} \quad (2)$$
$$= \frac{c_A(1-c_A)}{n} + \frac{p(1-p)}{(2p-1)^2 n}$$

令 $\lambda = \pi_A p + (1-\pi_A)(1-p)$, $1-\lambda = \pi_A(1-p) + (1-\pi_A)p$

则易知: $E(n_1) = n\lambda$, $E(n_1^2) = n\lambda + n(n-1)\lambda^2$。

又令: $\hat{\lambda} = \frac{n_1}{n}$,则有: $Var(\hat{\lambda}) = \lambda(1-\lambda)/n$

而 $E[\hat{\lambda}(1-\hat{\lambda})/(n-1)] = E[(nn_1 - n_1^2)/n^2(n-1)] = \frac{\lambda(1-\lambda)}{n}$.

从而 $Var(\pi_A)$的一个无偏估计为:

$$\hat{Var}(\hat{c}_A) = var(\hat{\lambda})/(2p-1)^2 = \hat{\lambda}(1-\hat{\lambda})/(n-1)(2p-1)^2$$
$$= \hat{c}_A(1-\hat{c}_A)/n + p(1-p)/(2p-1)^2 n \quad (3)$$

(2)参数 p的设定与样本容量 n的确定

由(2)可以看出,p越靠近$\frac{1}{2}$,则 $Var(\hat{\pi}_A)$的值越大。当 p比较靠近 0或 1时,$Var(\hat{\pi}_A)$就比较小。但另一方面,当 p比较接近 1或 0时,对被调查的保护程度就会降低,从而会降低被调查者的合作程度,使随机化回答的作用降低,增加了收集到真实的、正确的数据的困难程度。p的取值一般介于 0.7—0.8之间较适宜,当然也应根据实际调查问题的敏感程度适当选取。若敏感程度较高,则 p应取得小一点,但一般不宜低于 0.6;若敏感程度较低,则 p可取大一点,但一般也不宜高于 0.85。

从(3)中可以看出,只有$\hat{\pi}_A$是待估量,又易知 $f(x) = x(1-x)$, $x \in [0,1]$是凹函数,在 $x = \frac{1}{2}$时达到最大值。从而可知:

$$Var(\hat{c}_A) = \frac{c_A(1-c_A)}{n} + \frac{p(1-p)}{(2p-1)^2 n} \leqslant \frac{1}{4n} + \frac{p(1-p)}{(2p-1)^2 n}$$

预先给定要求精度,要求方差不超过 a,则:

$$Var(\hat{c}_A) \leqslant \frac{1}{4n} + \frac{p(1-p)}{(2p-1)^2 n} \leqslant a \qquad \text{从而得: } n \geqslant [\frac{1}{4} + \frac{p(p-1)}{(2p-1)^2}]/a \quad (4)$$

由(4)知,只要样本容量 n的取值大于$[\frac{1}{4} + \frac{p(p-1)}{(2p-1)^2}]/a$就可达到预定的精度。一般只需取:

$n = [\frac{1}{4a} + \frac{p(p-1)}{(2p-1)^2 a}] + 1$即可满足要求。

其中 $[x]$表示不大于 x的最大整数。

(3)无放回方式抽样下的沃纳模型

设总体容量为 N,用无放回简单随机抽样方法抽取容量为 n的样本,则沃纳模型估计为:

$$\hat{c}_A = (\hat{\lambda} - (1-p))/(2p-1) \quad (5)$$

方差为:

$$Var(\hat{c}_A) = \frac{N-n}{N-1}\cdot\frac{c_A(1-c_A)}{n} + \frac{p(1-p)}{n(2p-1)^2}$$

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www

$$\approx (1-f)\cdot\frac{c_A(1-c_A)}{n}+\frac{p(1-p)}{n(2p-1)^2} \quad (6)$$

方差的无偏估计量为:

$$\hat{Var}(\hat{c}_A)\approx (1-f)\cdot\frac{\hat{c}_A(1-\hat{c}_A)}{n}+\frac{p(1-p)}{n(2p-1)^2} \quad (7)$$

其中 $\hat{\lambda}=\frac{n_1}{n}$, n_1 是回答“是”的人数, $f=\frac{n}{N}$.

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的一个月内饮酒至少 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有放回地抽取了 100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。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?”占全部卡片的比例 $p=0.75$,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,被调查者把盒子中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”。调查结果是有 28个人回答了“是”, 72个人回答“不是”。

本例为—— 沃纳模型,有: $n=100$, $n_1=28$, $p=0.75$,因此有: $\hat{\lambda}=n_1/n=0.28$根据 (1)式可得 π_A 的沃纳估计值为:

$$\hat{c}_A=[\frac{n_1}{n}-(1-p)]/(2p-1)=[0.28-0.25]/(0.5)=0.06$$

也即有 6% 的人是酗酒者。

根据 (3), $\hat{\pi}_A$ 的方差一个无偏估计值为: $\hat{Var}(\hat{c}_A)=0.008145$

三、西蒙斯模型 (Simmons model)

这一模型是 1967年由西蒙斯提出的。其设计思想仍是基于沃纳的随机化回答思想,只是在设计中,用无关的问题 Y代替了沃纳模型中的敏感性问题 A的对立问题。比如敏感性问题为“你在考试中作弊了吗?”沃纳模型中的对立问题是“你在考试中没有作弊吗?”在西蒙斯模型中,用一与敏感性问题无关的问题来代替这一问题,比如“你是四月份出生的吗?”

(1)模型的设计与参数的估计

模型的基本设计为: 制作一个能产生两种实验结果的随机化装置,如两套外形一样的卡片,一套卡上写有敏感性问题“你属于 A吗?” (比如“你在考试中作弊了吗?”)不妨称为 1号卡片。另一套片上写有无关问题“你属于 Y吗?”其中 Y是与 A无关的非敏感性问题,如你是四月份出生的吗?”称此卡片为 2号卡片。将 1号卡片与 2号卡片按预定比例混合后,放入一盒子中,调查时,被调查者只需从盒子中任意抽取一张卡片,根据卡片上的问题做出真实的回答,当然调查员无权知道卡片上写的究竟是哪一个问题

(a) π_y 已知的情况:

设抽样方式是简单随机有放回的,样本容量为 n, π_A 是具有敏感性特征 A的人所占比例。π_y 是具有无关特性 Y的人所占比例。 p是 1号卡片出现的概率。

$$X_i=\begin{cases}1 & \text{第 } i\text{个样本回答“是”}\\ 0 & \text{第 } i\text{个样本回答“否”}\end{cases} \qquad i=1,2,\cdots n$$

则有:

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www

$P(X_i = 1) = p\,C_{AU} + (1-p)\,C_y \triangleq \lambda \quad P(X_i = 1) = p\,C_{AU} + (1-p)\,C_y \triangleq 1-\lambda$

令 $n_1 = \sum_{i=1}^{n} X_i$,即回答"是"的人数;令 $\hat{\lambda} = n_1/n$,则 π_{AU} 的一个极大似然无偏估计为:

$$\hat{C}_{AU1} = (\hat{\lambda} - (1-p)\,C_y)/p \tag{8}$$

其方差为: $Var(\hat{C}_{AU1}) = \lambda(1-\lambda)/np^2$ (9)

它的一个无偏估计为: $\hat{Var}(\hat{C}_{AU1}) = \hat{\lambda}(1-\hat{\lambda})/(n-1)p^2$ (10)

实践中 π_y 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_y 的值,而对于无关问题"你喜欢蓝色吗?"我们就无法预知 π_y 的值,此时 π_y 就是未知的。因此有必要对 π_y 未知的情况进行讨论

(b) π_y 未知的情况:

抽取两个相互独立的有放回的而且是互不相交的简单随机样本,样本量分别为 n_1、n_2。对于第一个样本,随机化装置出现 1号卡片的概率为 p_1, 2号卡片出现的概率为 $1-p_1$,第二个样本,1号卡片出现的概率为 p_2($p_1 \neq p_2$), 2号卡片出现的概率为 $1-p_2$。

π_{AU} 是具有敏感性特征的人所占比例。

$$X_{ij} = \begin{cases} 1 & \text{第 } i \text{ 个样本的第 } j \text{ 个人回答"是"} \\ 0 & \text{第 } i \text{ 个样本的第 } j \text{ 个人回答"否"} \end{cases} \qquad i = 1, 2; \qquad j = 1, 2, \cdots n_i$$

则: $P(X_{ij} = 1) = p_i\,C_{AU} + (1-p_i)\,C_y \triangleq \lambda_i$ (11a)

$P(X_{ij} = 0) = (1-p_i)\,C_{AU} + p_i(1-C_y) \triangleq 1-\lambda_i$ (11b)

令 n_{i1} 是第 i 个样本中回答"是"的人数。$\hat{\lambda}_i = n_{i1}/n_i \quad (i = 1, 2)$

由(11a)与(11b)可得 π_{AU} 的一个估计量为:

$$\hat{C}_{AU2} = [\hat{\lambda}_1(1-p_2) - \hat{\lambda}_2(1-p_1)]/(p_1 - p_2) \tag{12}$$

注意到: $E(\hat{\lambda}_i) = \lambda_i = p_i\pi_{AU} + (1-p_i)\pi_y$, $i = 1, 2$

可知 $\hat{\pi}_{AU2}$ 是 π_{AU} 的一个无偏估计量。其方差为:

$$Var(\hat{C}_{AU2}) = \left[\frac{(1-p_2)^2\lambda_1(1-\lambda_1)}{n_1} + \frac{(1-p_1)^2\lambda_2(1-\lambda_2)}{n_2}\right]/(p_1 - p_2)^2 \tag{13}$$

类似地,可得到 $Var(\hat{\pi}_{AU2})$ 的一个无偏估计为:

$$\hat{Var}(\hat{C}_{AU2}) = \left[\frac{(1-p_2)^2\hat{\lambda}_1(1-\hat{\lambda}_1)}{n_1 - 1} + \frac{(1-p_1)^2\hat{\lambda}_2(1-\hat{\lambda}_2)}{n_2 - 1}\right]/(p_1 - p_2)^2 \tag{14}$$

(2)设计参数的选择

π_y 已知的情况下, p 一般取值于 0.7–0.8之间即可。由(9)可知:

$$Var(\hat{C}_{AU1}) = \lambda(1-\lambda)/np^2 \leqslant \frac{1}{4np^2}$$

给定精度 α,则

$$Var(\hat{C}_{AU1}) \leqslant \frac{1}{4np^2} \leqslant T$$

$$\therefore n \geqslant \frac{1}{4Tp^2}$$

从而样本容量只须取 $n = [\frac{1}{4ap^2}] + 1$即可。

π_y 未知时,由于模型中有四个参数 n_1, n_2, p_1, p_2,因而需找到一组较优的参数,使得 $Var(\hat{\pi}_{AU2})$ 较小。由(13)并运用柯西—施瓦兹不等式可得:

$$\left[\frac{(1-p_2)^2\lambda_1(1-\lambda_1)}{n_1} + \frac{(1-p_1)\lambda_2(1-\lambda_2)}{n_2}\right](n_1 + n_2) \geqslant$$

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www

$$(1-p_2)\sqrt{\lambda_1(1-\lambda_1)}+(1-p_1)\sqrt{\lambda_2(1-\lambda_2)} \tag{15}$$

等号成立时 $Var(\hat{\pi}_{AU2})$到最小 (此时 p_1, p_2固定)。此时有:

$$\frac{n_1}{n_2}=\left[\frac{(1-p_2)^2\lambda_1(1-\lambda_1)}{(1-p_1)^2\lambda_2(1-\lambda_2)}\right]^{\frac{1}{2}} \tag{16}$$

且 (16)是 (15)等号成立的充要条件。(16)成立时有:

$$Var(\hat{c}_{AU2})=\left[\frac{(1-p_2)\sqrt{\lambda_1(1-\lambda_1)}+(1-p_1)\sqrt{\lambda_2(1-\lambda_1)}}{n^{\frac{1}{2}}(p_1-p_2)}\right]^2 \tag{17}$$

由上式,当 p_1, p_2确定时又有:

$$Var(\hat{c}_{AU2})=\left[\frac{(1-p_2)\sqrt{\lambda_1(1-\lambda_1)}+(1-p_1)\sqrt{\lambda_2(1-\lambda_2)}}{n^{\frac{1}{2}}(p_1-p_2)}\right]^2\leqslant\left[\frac{(2-p_1-p_2)}{2\sqrt{n}(p_1-p_2)}\right]^2$$

给定精度 α,则有:

$$Var(\hat{c}_{AU2})\leqslant\frac{(2-p_1-p_2)^2}{4n(p_1-p_2)^2}\leqslant T$$

$$\therefore n\geqslant\frac{(2-p_1-p_2)^2}{4T(p_1-p_2)^2}$$

从而取 $n=\left[\frac{(2-p_1-p_2)^2}{4\alpha(p_1-p_2)^2}\right]+1$即可。

下面再优化 p_1, p_2的取值,使得 (17)达到最小。

(17)对 p_1求偏导,由于 $p_1\neq p_2$,故不妨设 $p_1>p_2$,从而经过整理可知 $\partial Var(\hat{\pi}_{AU2})/\partial p_1$的符号与

$$\left(\frac{1}{2}-\lambda_1\right)(p_1-p_2)(c_{AU}-c_y)[\lambda_1(1-\lambda_1)]^{-\frac{1}{2}}-[\lambda_1(1-\lambda_1)]^{\frac{1}{2}}-[\lambda_2(1-\lambda_2)]^{\frac{1}{2}} \quad (*)$$

的符号相同,(11a)知: $(p_1-p_2)(c_{AU}-c_y)=\lambda_1-\lambda_2$

故 (*)又可化为:

$$-\frac{1}{2}[\lambda_1(1-\lambda_1)]^{-\frac{1}{2}}\{[\lambda_1(1-\lambda_2)]^{\frac{1}{2}}+[\lambda_2(1-\lambda_1)]^{\frac{1}{2}}\}^2$$

从而可知: $$\frac{\partial Var(\hat{c}_{AU2})}{\partial p_1}\leqslant 0$$

同理可知: $$\frac{\partial Var(\hat{c}_{AU2})}{\partial p_2}\geqslant 0$$

由此可知为了最小化 $Var(\hat{\pi}_{AU2})$,应取 $p_2=0$, p_1尽可能地大。也就是说,在第二个样本所用的随机化装置中,敏感性问题出现的概率为 0,而在第一样中,敏感性问题出现的概率应尽可能地大。此时 $Var(\hat{\pi}_{AU2})$又简化为:

$$Var(\hat{c}_{AU2})=\frac{\{[\lambda_1(1-\lambda_1)]^{\frac{1}{2}}+(1-p_1)[c_y(1-c_y)]^{\frac{1}{2}}\}^2}{np_1^2} \tag{18}$$

在实践中, p_1一般仍不宜大于 0.85, p_2的取值应根据实际问题,或者取 $p_2=0$,或者取 p_2尽可能的小。若取 $p_2=0$,则实际上是用第二个样本估计出 π_y,再用第一个样本估计 π_{AU}。

(3)无放回方式下的参数估计

对于有限总体,且抽样方式是无放回简单随机抽样,对于西蒙斯模型, π_y已知,从容量为 N 的总体中,无放回地抽取容量为 n 的简单随机样本,则 π_A的无偏估计量为:

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www

$$\hat{c}_A = \frac{\hat{\lambda} - (1-p)c_y}{p} \tag{19}$$

方差为: $Var(\hat{c}_A) = (1-f)\frac{\hat{c}_A(1-\hat{c}_A)}{n} + (1-f)\frac{(1-p)^2 c_y(1-c_y)}{np^2}$

$$+ \frac{p(1-p)(c_A + c_y - 2c_A c_y)}{np^2} \tag{20}$$

其无偏估计量为:

$$\hat{Var}(\hat{c}_A) = (1-f)\cdot\frac{\hat{\lambda}(1-\hat{\lambda})}{np^2} + \frac{1}{N}\cdot p(1-p)\cdot\frac{(\hat{c}_A + c_y - 2\hat{c}_A c_y)}{p^2}$$

$$= (1-f)\cdot\frac{\hat{c}_A(1-\hat{c}_A)}{n} + (1-f)\cdot\frac{(1-p)^2 c_y(1-c_y)}{np^2}$$

$$+ \frac{p(1-p)(\hat{c}_A + c_y - 2\hat{c}_A c_y)}{np^2}$$

其中,$\hat{\lambda} = \frac{n_1}{n}$, n_1 为回答"是"的人数。 $f = \frac{n}{N}$。

对于西蒙斯模型 π_y 未知,从容量为 N 的总体中,无放回地抽取容量分别为 n_1, n_2 的两个简单随机样本,则 π_A 的无偏估计量为:

$$\hat{c}_A = \frac{(1-p_2)\hat{\lambda}_1 - (1-p_1)\hat{\lambda}_2}{p_1 - p_2} \tag{22}$$

方差为: $Var(\hat{c}_A) = \frac{1}{(p_1-p_2)^2}[(1-p_2)^2 Var(\hat{\lambda}_1) + (1-p_1)^2 Var(\hat{\lambda}_2)]$

其中:

$$Var(\hat{\lambda}_1) = (1-f_1)\cdot\frac{c_A(1-c_A)}{n_1}\cdot p_1^2 + (1-f_1)\cdot\frac{(1-p_1)^2 c_y(1-c_y)}{n_1}$$

$$+ \frac{p_1(1-p_1)(c_A + c_y - 2c_A c_y)}{n_1}$$

$$Var(\hat{\lambda}_2) = (1-f_2)\cdot\frac{c_A(1-c_A)}{n_2}\cdot p_2^2 + (1-f_2)\cdot\frac{(1-p_2)^2 c_y(1-c_y)}{n_2}$$

$$+ \frac{p_2(1-p_2)(c_A + c_y - 2c_A c_y)}{n_2}$$

令 $q_1 = p_1(1-p_2)$ $q_2 = p_2(1-p_1)$ $q_3 = (1-p_1)(1-p_2)$, 则有;

$$Var(\hat{c}_A) = \left[\frac{1-f_1}{n_1}\cdot q_1^2 + \frac{1-f_2}{n_2}\cdot q_2^2\right]c_A(1-c_A) + \left(\frac{1-f_1}{n_1} + \frac{1-f_2}{n_2}\right)\cdot q_3^2 c_y(1-c_y)$$

$$+ \left(\frac{q_1}{n_1} + \frac{q_2}{n_2}\right)\cdot q_3\cdot(c_A + c_y - 2c_A c_y) \tag{23}$$

$Var(\hat{\pi}_A)$ 的一个估计量为:

$$\hat{Var}(\hat{c}_A) = \left[\frac{1-f_1}{n_1}\cdot q_1^2 + \frac{1-f_2}{n_2}\cdot q_2^2\right]\hat{c}_A(1-\hat{c}_A) + \left[\frac{1-f_1}{n_1} + \frac{1-f_2}{n_2}\right]\cdot q_3^2\hat{c}_y(1-\hat{c}_y)$$

$$+ \left(\frac{q_1}{n_1} + \frac{q_2}{n_2}\right)\cdot q_3\cdot(\hat{c}_A + \hat{c}_y - 2\hat{c}_A\hat{c}_y) \tag{24}$$

其中 $\hat{\pi}_y = \frac{p_2\hat{\lambda}_1 - p_1\hat{\lambda}_2}{p_2 - p_1}$

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www