

# 如何調查敏感性的問題

鄧進財 · 張春桃

## 主旨

直接詢問受訪者「你是否逃稅過?」、「你是否做過不道德的事?」等敏感性的問題,受訪者大多不會坦誠地回答。本文將簡述一種隨機性的問答,來消除受訪者的疑慮,進而提高所收集之資料的可信度,使得估計結果能更真切反應現實的狀況。

如果我們所要調查的問題涉及個人的名譽道德、私人秘密及其他有利害關係的事項,受訪者多半不會忠實回答所問的敏感性問題。因此,直接詢問受訪者敏感性的問題是較難獲得可靠的資料。

例如:直接問學生一年內是否作弊過?或直接詢問工商業公司老板是否逃稅過或直接調查已婚者有無出軌過?...等等,都是很難獲得受訪者坦誠相告。因此,直接調查所得的資料與事實往往相差很大,而無參考價值。

既然直接詢問敏感性的問題難以獲得可信的資料,我們就想出其他的調查方式,以獲取受訪者信賴,進而使調查結果具有可靠性。在此我們將簡潔地談談如何利用「隨機問答」的技巧,來消除受訪者疑慮,使他們能坦誠地回答問題,從而獲取與事實較接近之資料。那麼,什麼是隨機問答呢?我們提供受訪者兩個或多個問題,一個是敏感性的問題,另一個則是不傷大雅的是非題。接著,請受訪者隨機性的回答其中的一個問題,例如:受訪者自行丟銅板,若得正面就回答敏感性的問題,得反面就回答不傷大雅的是非題。受訪者祇須回答「是」或「否」,而不須告訴訪員所回答的是何種問題,如此訪員即無法確知受訪者所回答的是何種問題,因此受訪者便可以毫無顧忌的回答問題。這種調查方式,雖然我們並不知個別受訪者所回答的是何種問題,但是從整體上,我們確可輕易地估計出所要調查問題的結果。當然,估計的結果與事實常有誤差。例如:利用丟一個公平的銅板奇數次來估計出現正面的比例,估計的結果是無法獲取正確的比例0.5。但是,我們丟的次數愈多,所得的估計值與事實也愈一致。

假設我們要調查高中生手淫的比例,我們可設計兩個問題:

問題 1, 為敏感性的問題 —「你是否曾經手淫過?」,

問題 2, 為無關痛癢的問題 —「你的學號是否為奇數?」。

對每一位受訪的學生, 我們可請他或她依自己身分證字號的尾數回答問題: 若身分證字號的尾數大於或等於 7, 則回答問題 2, 若身分證字號的尾數小於 7, 則回答問題 1。

假設調查了一百位學生有四十四位回答「是」, 則我們獲得了表一的結果。任一身身分證字號的尾數大於或等於 7 的機會是十分之三。因此, 我們估計一百位受訪的學生當中約有三十位回答問題 2, 其餘均回答問題 1。接著, 從三十位回答問題 2 的學生中, 我們估計約有一半將會回答「是」, 這是因為學號有一半是奇數。因此, 回答問題 2 的三十位學生中估計有十五位會說「是」, 另外十五位會回答「否」。再由總人數中扣除回答問題 2 的人數, 即求得回答問題 1「是」與「否」的人數分別為 29 (即  $44 - 15$ ) 與 41 (即  $56 - 15$ )。最後, 利用回答問題 1 的 70 位受訪學生中, 得知有 29 位會回答「是」, 因此我們估計高中學生手淫的比率約為百分之四十一 (即  $29 \div 70 = 0.414$ ), 如表二所示。

表一、敏感性問題的調查結果

	是	否	小計
你會否手淫過?			
你學號是奇數?			
總人數	44	56	100

表二、敏感性問題的估計結果

	是	否	小計
你會否手淫過?	29	41	70
你學號是奇數?	15	15	30
總人數	44	56	100

此估計方法可以下列符號及式子表示之:

$n$ : 調查總人數,

$r$ : 受訪者回答敏感性問題的機率,

$1 - r$ : 受訪者回答非敏感性問題的機率,

$X$ : 受訪者回答「是」的隨機人數,

$n - X$ : 受訪者回答「否」的隨機人數,

$q$ : 對非敏感性問題回答「是」的機率,

$1 - q$ : 對非敏感性問題回答「否」的機率,

$p$ : 對敏感性問題回答「是」的比例,

則我們可得各項結果如表三所示, 而且得知  $p$  的估計量如下列公式

$$\hat{p} = (1/r)[X/n - (1 - r)q]. \quad (\text{公式1})$$

此外, 此估計量的變異數表示如下:

$$\hat{V}(\hat{p}) = (1/n)(1/r)^2(X/n)(1 - X/n). \quad (\text{公式2})$$

表三、敏感性問題的估計結果

	是	否	小計
敏感性問題	$X - n(1 - r)q$	$n - X - n(1 - r)(1 - q)$	$nr$
非敏感性問題	$n(1 - r)q$	$n(1 - r)(1 - q)$	$n(1 - r)$
總人數	$X$	$n - X$	$n$

同樣地，依表二的資料及公式1及公式2，我們得知  $r = 0.7$ ， $X = 44$ ， $n = 100$  及  $q = 0.5$ ，故高中學生手淫比率的估計值

$$\hat{p} = (1/0.7)[44/100 - 0.3 * 0.5] = 0.414.$$

而其變異數的估計值

$$\hat{V}(\hat{p}) = (1/100)(1/0.7)^2(44/100)(1 - 44/100) = 0.005.$$

利用此結果可得知高中學生手淫比率  $p$  的95%信賴區間值為

$$0.414 \pm 1.96 * \sqrt{0.005} = (0.275, 0.553).$$

上面我們所用的方法，若碰到一位多疑，而學號又不是奇數的同學，儘管事實上我們不知道他或她的學號，而他或她依然可能會害怕，假如他或她回答「是」，而我們若查出他或她的學號，那不就證明他或她曾手淫嗎？我們爲了避免因此而產生的資料偏差，可以應用下面的例子。

例如我們想研究青少年墮胎的比率，我們可請他或她依自己身分證字號的尾數回答問題：

若身分證字號的尾數大於或等於4：回答「你是否墮胎過？」，

若身分證字號的尾數等於2或3：回答「是」，

若身分證字號的尾數等於0或1：回答「否」。

這種方式的隨機問題，可以完全消除受訪者的疑慮。因爲受訪者回答「是」可能是身分證字號的尾數等於2或3，而必須回答「是」。接著，根據調查後「是」與「否」的總人數及類似表三的估計方式，我們就可估計青少年墮胎的比率。譬如，調查了一百位高中女生，有三十二位回答「是」，則我們獲得了表四的結果。類似表三的估計方式，我們可得表五。依表五的資料及公式1及公式2，我們得知  $r = 0.6$ （因爲100位高中女生中約有60位受訪者身分證字號的尾數大於或等於4）， $X = 32$ ， $n = 100$  及  $q = 0.5$ （因爲身分證字號的尾數等於2或3的女生人數約有20位，而且與身分證字號的尾數等於0或1的女生人數相同），故青少年墮胎率的估計值

$$\hat{p} = (1/0.6)[32/100 - 0.4 * 0.5] = 0.2.$$

而其變異數的估計值

$$\hat{V}(\hat{p}) = (1/100)(1/0.6)^2(32/100)(1 - 32/100) = 0.006.$$

利用此結果可得知青少年墮胎率  $p$  的 95% 信賴區間值為

$$0.2 \pm 1.96 * \sqrt{0.006} = (0.048, 0.352).$$

表四、青少年墮胎率的調查結果

	是	否	小計
你會否墮胎過?			
請回答「是」			
請回答「否」			
總人數	32	68	100

表五、青少年墮胎率的估計結果

	是	否	小計
你會否墮胎過?	12	48	60
請回答「是」	20	0	20
請回答「否」	0	20	20
總人數	32	68	100

—本文作者鄧進財現任教美國紐澤西州立 William Paterson 大學, 張春桃現任教淡江大學統計系—