管　　理　　與　　系　　統
第十六卷 第三期　民國九十八年七月
397-414 頁

Journal of Management & Systems
Vol. 16, No. 3, July 2009
pp. 397-414

分層隨機作答模式於敏感性問題之一般化研究

A Generalized Study of Sensitive Problems by using Stratified Randomized Response Model

王智立　Chih-Li Wang　　　蔡宛容　Wang-Jung Tsai

銘傳大學應用統計資訊學系

Department of Applied Statistics and Information Science, Ming Chuan University

(Received June 6, 2007; Final Version March 27, 2008)

摘 要：在許多問卷訪問過程中，若探討的主題牽涉到敏感性特質時，受訪者往往會基於保護個人隱私或種種心理因素，而產生拒絕回答或不誠實回答的情形，所以，較具敏感性問題的統計資料通常是缺乏或不精確的。爲解決此一情形，Warner (1965) 首先提出隨機作答模式（Randomized Response Model），期望藉由隨機作答模式的方式來獲得可靠的資料以推論敏感性群體比例，除此之外，還可保護受訪者的隱私進而降低受訪者不誠實回答的機率。Greenberg *et al*. (1969) 提出無關聯問題模式 (Unrelated-Question Model)，改善 Warner (1965) 模式中兩個問題皆具敏感性的疑慮，而 Christofides (2003) 延伸 Warner (1965) 模式提出新的隨機作答模式，讓受訪者不需要直接回答「是」與「否」，藉以降低不誠實回答的機率。本論文的重點在於如何應用隨機實驗來協助隨機作答程序的進行，以期在實務運用上，可以降低訪問的敏感程度，減低受訪者的排拒心及保障受訪者的隱私，以獲得更好的比例估計值或訪問品質。具體的成果有：(1). 將分層隨機抽樣設計應用到 Christofides (2003) 之隨機作答模式上，並與 Warner (1965)、Greenberg *et al*. (1969) 的無關聯問題模式作一合併，進而提出兩種新的隨機作答模式，文中將可找出敏感性比例的估計量與其變異數。(2)藉由資料模擬之方式比較在不同參數設計下，本文模式與 Warner (1965)、Christofides (2003) 模式之估計量的效率比較。(3)文中亦說明在相同的參數設計下，Warner (1965) 與 Christofides (2003) 模式皆爲本文所提模式的特例 (special cases)。

本研究感謝國科會研究經費之補助（計畫編號：NSC 94-2416-H-130-004）。

本文之通訊作者爲王智立，e-mail: clwang@mcu.edu.tw。

關鍵詞：敏感性問題、隨機作答模式、比例估計

Abstract: In the process of many questionnaires, if the topic of study is involves in sensitive characteristics, the participant usually resists or reply untruthfully. It is because of hiding the actual characteristic and protecting the individual privacy or all sorts of psychological factor. Hence, static data or estimates concerning sensitive question are usually very insubstantial and inaccurate. In order to solve this problem, Warner (1965) first proposed "Randomized Response Model", and he helps to obtain the reliable data to deduce proportion of the sensitive group by executing this model. Besides, it may raise the protection of participants and reduce the probability which the participant untruthful replied. Greenberg *et al*. (1969) proposes Unrelated-Question model, to improve the two questions which both have the sensitive anxiety in Warner (1965) model. Christofides (2003) proposed a new randomized response which extended Warner (1965) model. The participant doesn't have to answer 「yes」 and 「no」 directly, and reduces the sensitive level of the participant. This article mainly focuses on how to employ random trials to assist the progress of randomized response procedure so that the rejection rate of interview can be reduced and the privacy of interviewees is well protected. Better ratio estimate and interview qualities are expected to obtain. Toward this purpose, the following works have been done: (1) We apply the stratified random sampling design to generalized Christofides (2003) randomized response model. Two new generalized models for the sensitive problems, which is an integration of Warner (1965), Greenberg *et al*. (1969) and Christofides (2003) model, is proposed. In this paper, we also can find the estimator and variance of sensitive proportion. (2) We want to compare the performance of estimation with Warner (1965) and Christofides (2003) model, by data simulating. (3) The proposed models include Warner (1965) and Christofides (2003) methods as special cases for a specific choice of the parameters.

***Key words*: Sensitive Problems, Randomized Response, Estimation of a Proportion**

1. 緒論

在問卷訪問過程中，若探討的主題牽涉到敏感性特質時，受訪者往往會基於保護個人隱私或種種心理因素，而產生拒絕回答或不誠實回答的情形，所以，這類問題的統計資料通常是缺乏及不精確的。爲了解決此情形，Warner (1965) 首先提出隨機作答模式 (Randomized Response Model)，期望藉由隨機作答模式以推論敏感性群體比例。此後許多學者也提出了其他的隨機作

答程序改良 Warner (1965) 模式，其最主要的目的皆在於能更準確的估計敏感性問題的效率；其中 Greenberg *et al*. (1969) 提出無關聯問題模式 (Unrelated-Question Model)，改善 Warner (1965) 模式中兩個問題皆具敏感性的疑慮，而 Christofides (2003) 延伸 Warner (1965) 模式提出新的隨機作答模式，讓受訪者不需要直接回答「是」與「否」，藉以降低不誠實回答的機率。接著，2004 年 Kim and Warde 將 Warner (1965) 的隨機作答程序作改良並提出「分層隨機作答模式 (Stratified Randomized Response Model)」，他們所提出的分層隨機作答程序是將母體分成若干層，在每一層中採取簡單隨機抽樣法且抽出放回的方式抽出一組樣本進行隨機作答程序。

本文的重點在於如何應用隨機實驗來協助隨機作答程序的進行，以期在實務運用上，可以降低訪問的敏感程度，減低受訪者的排拒心及保障受訪者的隱私，以獲得更好的比例估計值或訪問品質。具體的成果有：(1)將分層隨機抽樣設計應用到 Christofides (2003) 之隨機作答模式上，並與 Warner (1965)、Greenberg *et al*. (1969) 的無關聯問題模式作一合併，進而提出兩種新的廣義隨機作答模式，文中將可找出敏感性比例的估計量與其變異數。(2)藉由資料模擬之方式比較在不同參數設計下，本文模式與 Warner (1965)、Christofides (2003) 模式之估計量的效率比較。(3)文中亦說明在相同的參數設計下，Warner (1965) 與 Christofides (2003) 模式皆爲本文所提模式的特例 (special cases)。

2. 文獻探討

近年來許多涉及敏感性特質的研究無法得到準確率較高的資料，許多學者爲了讓受訪者誠實作答，一直不斷的研究如何獲得準確率較高的資料。因此，依母體特性將母體分爲兩個群體，一爲具有敏感性特質的群體 A，另一爲不具敏感性的群體 $\overline{A}$，利用簡單隨機抽樣且抽出不放回的方式 (simple random sampling with replacement; SRSWR) 得到一組樣本 n，藉此估計出群體 A 在母體中的比例 θ。接著本文將介紹與本文模式相關之隨機作答模式的發展過程。

Warner (1965) 提出的隨機作答模式爲「Q_{W1}：我是屬於敏感性群體 A 中的一員」;「Q_{W2}：我不是屬於敏感性群體 A 中的一員」。接下來將會提供每一位被抽中的受訪者一個隨機裝置 (random device)，在此隨機裝置中 Q_{w1} 與 Q_{w2} 被抽中的機率分別爲 P 與 $1-P$，而受訪者針對所抽中的問題只需回答「是」與「否」即可，值得注意的是受訪者不需透露給訪員知道他/她抽中的問題究竟爲何？因此受訪者的隱私將可以得到保護，Warner (1965) 希望藉由這樣的隨機作答模式可以降低受訪者的防備心而獲得較高的誠實回答率。

藉由 Warner (1965) 所提出的模式可以估計敏感性群體的比例 θ，假設受訪者皆是誠實作答的情形下，受訪者回答「是」的機率爲

$$\lambda_W = \theta P + (1-\theta)(1-P) = (1-P) + (2P-1)\theta \tag{1}$$

而 n' 爲回答「是」的人數，n 爲樣本總數，故 λ_W 的不偏估計量爲 $\hat{\lambda}_W = n'/n$ ，所以 θ 的不偏估計量爲

$$\hat{\theta}_W = \frac{P-1}{2P-1} + \frac{n'}{n(2P-1)},\ \ P \neq \frac{1}{2} \tag{2}$$

而且可以得到 $\hat{\theta}_W$ 的變異數爲

$$Var\,(\hat{\theta}_W) = \frac{\theta(1-\theta)}{n} + \frac{P(1-P)}{n(2P-1)^2},\ \ P \neq \frac{1}{2} \tag{3}$$

在此研究中機率 P 是先行決定的，然而 P 的大小會影響變異數的大小，也會影響到受訪者與訪員之間合作的程度。當選擇 $P=1$ 或 $P=0$ 時變異數會最小值，也就是說 P 選擇越靠近 0 或 1 時，$\hat{\theta}_W$ 會得到較高的效率，然而這樣的設計相當於直接詢問法，故受訪者誠實回答率將不高。若 P 愈接近 0.5 時變異數愈大，且受訪者的隱私會得到較高的保護，但是當 $P=0.5$ 時根本無法提供任何訊息。

Greenberg *et al*. (1969) 提出若受訪者不誠實回答則 Warner (1965) 所提出的 $\hat{\theta}_W$ 將不再是不偏估計量 (unbiased estimator)，因此提出了一個新的程序，稱爲「無關聯隨機作答程序」，Greenberg *et al*. (1969) 利用敏感性群體 A ，並且將隨機器中的第二個問題以不相關且不具有敏感性的群體 Y 來替代 Warner (1965) 程序中的非敏感性群體 $\overline{A}$ ，因爲基本上 Warner (1965) 模式的 Q_{W1} 和 Q_{W2} 爲兩個相關聯的敏感性問題，有可能會引起受訪者的不信任感。Greenberg *et al*. (1969) 的執行步驟和 Warner (1965) 相似，其問題如下：「 Q_{G1} : 我是屬於敏感性群體的一員」;「 Q_{G2} : 我是屬於群體 Y （和敏感性群體 A 無關）的一員」。舉例來說，研究者想知道墮胎的真正比例爲何，則受訪者有可能會抽中的問題一爲“我曾經墮過胎”，另一問題爲“我是五月份出生的人”。受訪者針對所抽中的問題只需回答「是」與「否」即可，且受訪者不需透露訪員他/她抽中的問題究竟爲何？Greenberg *et al*. (1969) 模式在非敏感性群體 Y 的比例 π_y 已知的情況下，令 P_1 爲選擇問題一 Q_{G1} 作答的機率。則敏感性群體的母體比例 θ 的估計量及其變異數分別爲：

$$\hat{\theta}_G = \frac{\frac{n'}{n} - (1-P_1)\pi_y}{P_1} \tag{4}$$

$$Var\left(\hat{\theta}_G\right) = \frac{\lambda(1-\lambda)}{nP_1^{\ 2}} \tag{5}$$

其中 λ 爲受訪者中回答「是」的比例，且 $\lambda = P_1\theta + (1-P_1)\pi_y$ 。

Christofides (2003) 假設欲估計具有敏感性特質的母體比例 θ ，並藉由 SRSWR 從母體抽出

一組樣本 n ，提供每一位被抽出的受訪者一個隨機裝置，而這個隨機裝置會產生相對機率爲 $p_1,\cdots,p_L$ 的整數 $1,\cdots,L$ ，每一位受訪者在使用完此隨機裝置之後回答出一個數字：若他/她具有此敏感性特質，則回答所抽出的數字到 $L+1$ 的距離；反之，則回答抽出的數字到 0 的距離。若第 i 個受訪者對於敏感性問題的答案爲「是」，則令 x_i 爲 $L+1$ ，反之，則令 x_i 爲 0，因此，$P(x_i=L+1)=\theta$ 且 $P(x_i=0)=1-\theta$ 。 y_i 爲第 i 個受訪者使用隨機器產生的數字，且受訪者所回答的數字爲 d_i ，則 $d_i=|x_i-y_i|$ ，因此，第 i 位受訪者回答數字爲 k 的機率爲

$$P(d_i=k)=(1-\theta)P_k+\theta P_{L+1-k},\quad k=1,\cdots,L \tag{6}$$

令 $\bar{d}$ 爲 $d_1,\cdots,d_n$ 的樣本平均數，因此可以得到敏感性群體比例 θ 的估計式爲

$$\hat{\theta}_C=\left(\bar{d}-E(y)\right)\left(L+1-2E(y)\right)^{-1},L+1-2E(y)\neq 0 \tag{7}$$

則 $\hat{\theta}_C$ 的變異數爲

$$Var(\hat{\theta}_C)=\frac{1}{n}\theta(1-\theta)+\frac{1}{n}Var(y)(L+1-2E(y))^{-2} \tag{8}$$

式(8)中等號右邊的第一項是由隨機抽樣 (random sampling) 所產生的變異，而第二項爲隨機化程序 (randomized procedure) 所產生的變異。

Kim and Warde (2004) 利用 Warner (1965) 所提出的隨機作答模式爲基本概念，提出分層隨機作答模式 (Stratified Randomized Response Model)，所謂的「層」是指將母體分成多個不重複的群體，分層隨機作答模式是指在每一層當中皆使用相同的隨機器，並且皆採用 SRSWR 的方法。Kim and Warde (2004) 模型的設計如下：「Q_{S1}：我是屬於敏感性群體的一員」；「Q_{S2}：我不是屬於敏感性群體的一員」。在第 i 層中問題 Q_{S1} 與問題 Q_{S2} 的機率分別爲 P_i 與 $1-P_i$ ，而受訪者針對抽中的問題只需回答「是」或「否」即可，並且受訪者不須讓訪員知道自己抽中的問題爲何。

令 N 爲母體總人數， N_i 爲第 i 層 $(i=1,\cdots,m)$ 的母體人數。 n 爲總樣本數且 n_i 爲第 i 層中樣本數，因此 $n=\sum_{i=1}^{m}n_i$ 。假設受訪者皆誠實回答的情形下，Kim and Warde (2004) 模式之第 i 層受訪者中回答「是」的機率爲

$$\lambda_{Si}=P_i\theta_i+(1-P_i)(1-\theta_i),\quad i=1,2,\cdots,m,\quad P_i\neq 0.5 \tag{9}$$

此處 θ_i 爲第 i 層中敏感性群體的比例，在 Kim and Warde (2004) 模式中，參數 λ_i 之估計量 $\hat{\lambda}_{Si}$ 屬於二項分配 (n_i,λ_i) ， λ_i 爲第 i 層受訪者中回答「是」的比例，且各層中取樣的方式爲獨立，因此可得到第 i 層的敏感性群體比例 θ_i 如下

$$\hat{\theta}_i = \frac{\hat{\lambda}_{Si} - (1 - P_i)}{2P_i - 1}, \quad i = 1,2,\cdots,m \tag{10}$$

令 w_i 爲第 i 層中的權重，故 $w_i = \frac{N_i}{N}$ ， N_i 爲第 i 層中母體人數 $i = 1,2,\cdots,m$ ，且 $\sum_{i=1}^{m} w_i = 1$ ，因此可得到敏感性比例 θ 之估計量爲

$$\hat{\theta}_{KM} = \sum_{i=1}^{m} w_i \hat{\theta}_i = \sum_{i=1}^{m} w_i \left[\frac{\hat{\lambda}_{Si} - (1 - P_i)}{2P_i - 1} \right] \tag{11}$$

且 $\hat{\theta}_{KM}$ 爲 θ 的不偏估計量。而 $\hat{\theta}_{KM}$ 的變異數爲

$$Var(\hat{\theta}_{KM}) = Var\left(\sum_{i=1}^{m} w_i \hat{\theta}_i \right) = \sum_{i=1}^{m} w_i^2 Var(\hat{\theta}_i) = \sum_{i=1}^{m} \frac{w_i^2}{n_i} \left[\theta_i (1 - \theta_i) + \frac{P_i (1 - P_i)}{(2P_i - 1)^2} \right] \tag{12}$$

Christofides (2005) 假設母體可分割爲 K 層，且 w_i 爲第 i 層的權重， n_i 爲第 i 層中抽出的樣本數，並利用其 Christofides (2003) 之隨機作答模式，估計其第 i 層的敏感性群體比例 θ_i 及其變異數分別爲式(7)之 $\hat{\theta}_C$ 及式(8)之 $Var(\hat{\theta}_C)$ ，因此可得敏感性比例 θ 之估計量及其變異數分別爲

$$\hat{\theta}_{CH} = \sum_{i=1}^{m} w_i \hat{\theta}_C = \sum_{i=1}^{m} w_i \left[\left(\bar{d} - E(y) \right) \left(L + 1 - 2E(y) \right)^{-1} \right], L + 1 - 2E(y) \neq 0 \tag{13}$$

$$Var\left(\hat{\theta}_{CH}\right) = \sum_{i=1}^{m} w_i^2 Var\left(\hat{\theta}_C\right) = \sum_{i=1}^{m} \frac{w_i^2}{n_i} \left[\theta(1 - \theta) \right] + \frac{Var(y)}{\left[L + 1 - 2E(y) \right]^2} \sum_{i=1}^{m} \frac{w_i^2}{n_i} \tag{14}$$

王智立、蔡宛容 (民96) 延伸Christofides (2003) 所提出的隨機作答模式，並與Warner (1965)模式、Greenberg *et al*. (1969) 的無關聯問題模式作合併，進而提出一般化 Greenberg 無關聯隨機作答模式，此模式之敏感性比例 θ 的估計量與其變異數分別爲

$$\hat{\theta}_{WT} = \frac{\bar{d} - E(y)}{\pi \left[L + 1 - 2E(y) \right]}, L + 1 - 2E(y) \neq 0 \tag{15}$$

$$Var(\hat{\theta}_{WT}) = \frac{1}{n\pi} \theta (1 - \pi\theta) + \frac{1}{n} Var(y) \left\{ \pi [L + 1 - 2E(y)] \right\}^{-2} \tag{16}$$

3. 研究方法

在 3.1 節中，將介紹本文第一個新的「分層廣義隨機作答模式」的執行方式與推導過程，進而可推知敏感性群體比例的估計量及其變異數；而在 3.2 節中，將介紹第二個新的「分層一般隨

機作答模式」的執行方式與推導過程，並推導出敏感性群體比例的估計量及其變異數。

3.1 分層廣義隨機作答模式

本節中吾人先提出第一個新的隨機作答模式，謂之爲「分層廣義隨機作答模式」，其精神係考慮 Christofides (2003) 所提出的模式，吾人希望藉由無關聯的問題可以降低受訪者的不安全感而獲得較高的誠實作答率，因此基於 Christofides (2003) 所提出的隨機作答模式，Warner (1965) 模式及 Greenberg *et al.* (1969) 的無關聯問題模式，本文將上述三篇文章提出的隨機作答模式作一延伸。首先，吾人將母體切割爲 m 層，並假設第 i 層的樣本人數爲 n_i，提供第 i 層 $(i=1,2,\cdots,m)$ 中的每一位受訪者一個隨機器 R_i，此隨機器中有 $1,2,\cdots,L_i$ 個整數且相對機率爲 $p_{i1},p_{i2},\cdots,p_{iL_i}$，接著在每一層中訪員將同時提出兩個問題，

Q_1 : 我是屬於敏感性群體 A 的一員？

Q_2 : 我是屬於群體 Y（不具敏感性且和敏感性群體 A 無關）的一員？

第一個問題 Q_1 爲具有敏感性特質 A 的問題，其對應之相對機率爲 θ_i，而第二個問題 Q_2 爲與 A 無關聯的問題，其對應的機率爲 π_i，且假設無關聯特質之母體比例 π_i 爲已知，受訪者在使用此隨機器之後會回答一個「數字」，若他/她對於這兩個問題的答案皆爲「是」或皆爲「否」，則回答隨機器所抽出的數字到 L_i+1 的距離；若他/她對於這兩個問題的答案其中有一個爲「是」一個爲「否」，則回答隨機器所抽出的數字到 0 的距離。因此，受訪者並不會洩漏他/她究竟真正的特質爲何。

若第 i 層中第 j 位 $(j=1,2,\cdots,n_i)$ 受訪者對於兩個問題的答案皆爲「是」或皆爲「否」，則令 $x_{ij}=L_i+1$，反之，若其答案爲一個「是」與一個「否」，則令 $x_{ij}=0$。因此，第 i 層中第 j 位受訪者 x_{ij} 值爲 L_i+1 的機率與 x_{ij} 值爲 0 的機率分別爲 $\pi_i\theta_i+(1-\theta_i)(1-\pi_i)$ 與 $\theta_i(1-\pi_i)+\pi_i(1-\theta_i)$，即 $P(x_{ij}=L_i+1)=\pi_i\theta_i+(1-\theta_i)(1-\pi_i)$ 與 $P(x_{ij}=0)=\theta_i(1-\pi_i)+\pi_i(1-\theta_i)$。再令 y_{ij} 爲第 i 層中第 j 位受訪者使用隨機器抽出的數字，且受訪者所回答的數字爲 d_{ij}，則 $d_{ij}=\left|x_{ij}-y_{ij}\right|$，因此，第 i 層中第 j 位受訪者會回答數字爲 k 的機率爲

$$P(d_{ij}=k)=\left[\theta_i(1-\pi_i)+\pi_i(1-\theta_i)\right]P_k+\left[\pi_i\theta_i+(1-\theta_i)(1-\pi_i)\right]P_{L_i+1-k}, \tag{17}$$

其中 $i=1,2,\cdots,m$，$j=1,2,\cdots,n_i$，$k=1,\cdots,L_i$ 且 P_k、P_{L_i+1-k} 分別爲隨機裝置中數字爲「k」及「L_i+1-k」之機率。

接著藉由以下的引理 3.1.1 和 3.1.2，可推導出第 i 層敏感性群體比例 θ_i 的估計量及其變異數。

引理 3.1.1：可以得到 d_{ij} 的期望值爲

$$E(d_{ij})=\left[\theta_i(1-2\pi_i)+\pi_i\right]E(y_i)+\left[\theta_i(2\pi_i-1)+(1-\pi_i)\right]\left[L_i+1-E(y_i)\right] \tag{18}$$

證明過程如下：

$$E(d_{ij})=\sum_{k=1}^{L_i} kP(d_{ij}=k)$$

$$=\sum_{k=1}^{L_i} k\left[\theta_i(1-2\pi_i)+\pi_i\right]P_k+\sum_{k=1}^{L_i} k\left[\theta_i(2\pi_i-1)+(1-\pi_i)\right]P_{L_i+1-k}$$

$$=\left[\theta_i(1-2\pi_i)+\pi_i\right]E(y_i)+\left[\theta_i(2\pi_i-1)+(1-\pi_i)\right]\left[(L_i+1)-E(y_i)\right]$$

此處，$\sum_{k=1}^{L_i} kP_k=E(y_i)$ 。

引理 3.1.2：d_{ij} 的變異數爲

$$Var(d_{ij})=Var(y_i)+\left[\theta_i(1-2\pi_i)+\pi_i\right]\left[\theta_i(2\pi_i-1)+(1-\pi_i)\right](L_i+1-2E(y_i))^2 \tag{19}$$

證明過程如下：

$$Var(d_{ij})=E(d_{ij}^2)-\left[E(d_{ij})\right]^2$$

$$=\left[\theta_i(1-2\pi_i)+\pi_i\right]Var(y_i)+\left[\theta_i(2\pi_i-1)+(1-\pi_i)\right]Var(y_i)$$

$$+\left[\theta_i(1-2\pi_i)+\pi_i\right]E(y_i)\times B_i+\left[\theta_i(2\pi_i-1)+(1-\pi_i)\right]\times C_i$$

此處，$B_i=\left[\theta_i(2\pi_i\text{ -}1)+(1+\pi_i)\right]\left[2E(y_i)-L_i-1\right]$

$C_i=(L_i+1-E(y_i))(L_i+1-2E(y_i))\left[\theta_i(1-2\pi_i)+\pi_i\right]$

經合併化簡後可以得到 d_{ij} 的變異數爲

$$Var(d_{ij})=Var(y_i)+\left[\theta_i(1-2\pi_i)+\pi_i\right]\left[\theta_i(2\pi_i-1)+(1-\pi_i)\right](L_i+1-2E(y_i))^2$$

經由上述引理之證明，且再令 $\bar{d}_i$ 爲 $d_{i1},d_{i2},\cdots,d_{in_i}$ 的第 i 層樣本平均數後，可由(18)式可以得到 θ_i 的估計量爲

$$\hat{\theta}_{iI}=\frac{\bar{d}_i-\left[L_i+1-E(y_i)-\pi_i(L_i+1-2E(y_i))\right]}{(L_i+1-2E(y_i))(2\pi_i-1)} \tag{20}$$

其中，$L_i+1-2E(y_i)\neq 0,\ \pi_i\neq 0.5$ 。

定理 3.1.1：$\hat{\theta}_{iI}$ 爲第 i 層敏感性群體比例 θ_i 的不偏估計量。

證明：首先因爲

$$E(\bar{d}_i)=\frac{1}{n_i}\sum_{j=1}^{n_i}E(d_{ij})$$
$$=\left\{\left[\theta_i(1-2\pi_i)+\pi_i\right]E(y_i)+\left[\theta_i(2\pi_i-1)+(1-\pi_i)\right]\right\}(L_i+1-2E(y_i))$$
$$=E(d_{ij})$$

故可以得知

$$E(\hat{\theta}_{iI})=\frac{\theta_i(L_i+1)(2\pi_i-1)-2\theta_iE(y_i)(2\pi_i-1)}{(L_i+1-2E(y_i))(2\pi_i-1)}=\theta_i$$

因此可以得到敏感性群體比例θ的不偏估計量$\hat{\theta}_I$如下：

$$\hat{\theta}_I=\sum_{i=1}^{m}w_i\hat{\theta}_{iI}=\sum_{i=1}^{m}w_i\left\{\frac{\bar{d}_i-\left[L_i+1-E(y_i)-\pi_i(L_i+1-2E(y_i))\right]}{(L_i+1-2E(y_i))(2\pi_i-1)}\right\} \tag{21}$$

令N爲母體總數，N_i爲第i層的母體人數，則$w_i=\left(\frac{N_i}{N}\right),\ i=1,2,\cdots,m$，且$\sum_{i=1}^{m}w_i=1$。

定理 3.1.2：$\hat{\theta}_I$爲敏感性群體比例θ的不偏估計量。

證明：

$$E\left(\hat{\theta}_I\right)=E\left(\sum_{i=1}^{m}w_i\hat{\theta}_{iI}\right)=\sum_{i=1}^{m}w_iE\left(\hat{\theta}_{iI}\right)=\sum_{i=1}^{m}w_i\theta_i=\theta$$

因此，可得定理 3.1.1 中第i層敏感性群體比例θ_i的不偏估計量$\hat{\theta}_{iI}$的變異數如定理 3.1.3 所示。

定理 3.1.3：估計量$\hat{\theta}_{iI}$的變異數爲

$$Var(\hat{\theta}_{iI})=\frac{1}{n_i}Var(y_i)\left[(L_i+1-2E(y_i))(2\pi_i-1)\right]^{-2}$$
$$+\frac{1}{n_i}\left[\theta_i(1-2\pi_i)+\pi_i\right]\left[\theta_i(2\pi_i-1)+(1-\pi_i)\right](2\pi_i-1)^{-2} \tag{22}$$

證明：利用(20)式所得到之不偏估計量$\hat{\theta}_{iI}$，計算其變異數爲

$$Var(\hat{\theta}_{iI})=Var\left(\frac{\bar{d}_i-\left[L_i+1-E(y_i)-\pi_i(L_i+1-2E(y_i))\right]}{(L_i+1-2E(y_i))(2\pi_i-1)}\right)$$

$$=\frac{1}{n_i}Var(y_i)\left[(L_i+1-2E(y_i))(2\pi_i-1)\right]^{-2}$$

$$+\frac{1}{n_i}\left[\theta_i(1-2\pi_i)+\pi_i\right]\left[\theta_i(2\pi_i-1)+(1-\pi_i)\right](2\pi_i-1)^{-2}$$

定理 3.1.4：估計量 $\hat{\theta}_I$ 的變異數爲

$$Var(\hat{\theta}_I)=\sum_{i=1}^{m}\frac{w_i^2}{n_i}\Big\{Var(y_i)\left[(L_i+1-2E(y_i))(2\pi_i-1)\right]^{-2} \\ +\left[\theta_i(1-2\pi_i)+\pi_i\right]\left[\theta_i(2\pi_i-1)+(1-\pi_i)\right](2\pi_i-1)^{-2}\Big\} \quad (23)$$

證明：利用(21)式所得到之不偏估計量 $\hat{\theta}_I$ 及定理 3.1.3 之結果，計算 $\hat{\theta}_I$ 變異數爲

$$Var(\hat{\theta}_I)=Var\left(\sum_{i=1}^{m}w_i\hat{\theta}_I\right)=\sum_{i=1}^{m}w_i^2Var\left(\hat{\theta}_{iI}\right) \\ =\sum_{i=1}^{m}\frac{w_i^2}{n_i}\Big\{Var(y_i)\left[(L_i+1-2E(y_i))(2\pi_i-1)\right]^{-2} \\ +\left[\theta_i(1-2\pi_i)+\pi_i\right]\left[\theta_i(2\pi_i-1)+(1-\pi_i)\right](2\pi_i-1)^{-2}\Big\}$$

值得注意的是，當 $\pi_i=0$ 或 1，分層廣義隨機作答模式之估計量 $\hat{\theta}_I$ 與 Christofides (2005)的估計量 $\hat{\theta}_{CH}$ 相同且其變異數亦相同，因此，Christofides (2005) 的模式可視爲本文分層廣義隨機作答模式之特例。當 $\pi_i=0$ 或 1 時且 $m=1$，分層廣義隨機作答模式之估計量 $\hat{\theta}_I$ 與 Christofides (2003) 的估計量 $\hat{\theta}_C$ 相同且其變異數亦相同，因此，Christofides (2003) 的模式可視爲本文分層廣義隨機作答模式之特例。當 $\pi_i=1$ 且 $m=1$時，分層廣義隨機作答模式之估計量 $\hat{\theta}_I$ 與王智立、蔡宛容 (民 96) 的估計量 $\hat{\theta}_{WT}$ 相同且其變異數亦相同，因此，王智立、蔡宛容 (民 96) 的模式可視爲本文分層廣義隨機作答模式之特例。另外，當 $L_i=2$ 、 $\pi_i=1$ 且 $m=1$時，分層廣義隨機作答模式之估計量 $\hat{\theta}_I$ 與 Warner (1965) 的估計量 $\hat{\theta}_W$ 相同且其變異數亦相同，因此，Warner (1965) 的模式亦可視爲本文分層廣義隨機作答模式之特例。

3.2 分層一般隨機作答模式

本節吾人將提出另一個新的隨機作答模式，謂之爲「分層一般隨機作答模式」，其執行的方法與本文 3.1 節之分層廣義隨機作答模式大致相同，不同的地方在於若受訪者對於這兩個問題的答案皆爲「是」，則回答隨機器所抽出的數字到 L_i+1 的距離；若他/她對於這兩個問題的答案中有一個爲「否」或皆爲「否」，則回答隨機器所抽出的數字到 0 的距離。接著若第 i 層中第 j 位 $(j=1,2,\cdots,n_i)$ 受訪者對於兩個問題的答案皆爲「是」，則令 $x_{ij}=L_i+1$ ，反之，若其答案中有一個爲「否」或皆爲「否」，則令 $x_{ij}=0$ 。因此，第 i 層中第 j 位受訪者 x_{ij} 值爲 L_i+1 與 0 的機率

分別為 $\pi_i\theta_i$ 與 $1-\pi_i\theta_i$ ，即 $P(x_{ij}=L_i+1)=\pi_i\theta_i$ 與 $P(x_{ij}=0)=1-\pi_i\theta_i$ 。再令 y_{ij} 為第 i 層中第 j 位受訪者使用隨機器抽出的數字，且令受訪者所回答的數字為 d_{ij} ，則 $d_{ij}=\left|x_{ij}-y_{ij}\right|$ ，因此，第 i 層中第 j 位受訪者會回答數字為 k 的機率為

$$P(d_{ij}=k)=(1-\pi_i\theta_i)P_k+(\pi_i\theta_i)P_{L_i+1-k}\text{ ， }k=1,\cdots,L_i \tag{24}$$

其中 P_k 、 P_{L_i+1-k} 分別為隨機器中數字為「 k 」及「 L_i+1-k 」之機率。

接著藉由以下的引理 3.2.1 和 3.2.2，可推導出第 i 層敏感性群體比例 θ_i 的估計量及其變異數。

引理 3.2.1：可以得到 d_{ij} 的期望值為

$$E(d_{ij})=E(y_i)+(\pi_i\theta_i)[L_i+1-2E(y_i)] \tag{25}$$

證明過程如下：

$$\begin{aligned}E(d_{ij})&=\sum_{k=1}^{L_i}kP(d_{ij}=k)=\sum_{i=1}^{L_i}(1-\pi_i\theta_i)\times kP_k+\sum_{i=1}^{L_i}(\pi_i\theta_i)\times kP_{L_i+1-k}\\&=E(y_i)+(\pi_i\theta_i)[L_i+1-2E(y_i)]\end{aligned}$$

引理 3.2.2： d_{ij} 的變異數為

$$Var(d_{ij})=Var(y_i)+\pi_i\theta_i(1-\pi_i\theta_i)[L_i+1-2E(y_i)]^2 \tag{26}$$

證明過程如下：

$$\begin{aligned}Var(d_{ij})&=E(d_{ij}^2)-[E(d_{ij})]^2\\&=E(y_i^2)+\pi_i\theta_i\{(L_i+1)^2-2(L_i+1)E(y_i)\}-\{E(y_i)+(\pi_i\theta_i)[L_i+1-2E(y_i)]\}^2\\&=Var(y_i)+\pi_i\theta_i(1-\pi_i\theta_i)[L_i+1-2E(y_i)]^2\end{aligned}$$

經由上述引理之證明，再令 $\bar{d}_i$ 為 $d_{i1},d_{i2},\cdots,d_{in_i}$ 的第 i 層樣本平均數後，可由(25)式可以得到 θ 的估計量為

$$\hat{\theta}_{iII}=\frac{\bar{d}_i-E(y_i)}{\pi_i[L_i+1-2E(y_i)]}\text{ ， }\quad L_i+1-2E(y_i)\neq 0,\pi_i\neq 0\text{ 。} \tag{27}$$

定理 3.2.1： $\hat{\theta}_{iII}$ 為第 i 層敏感性群體比例 θ_i 的不偏估計量。

證明：首先因為

$$E(\bar{d}_i)=\frac{1}{n_i}\sum_{j=1}^{n_i}E(d_{ij})=E(y_i)+(\pi_i\theta_i)[L_i+1-2E(y_i)]=E(d_{ij})$$

故可以得知

$$E(\hat{\theta}_{iII})=E\left(\frac{\bar{d}_i-E(y_i)}{\pi_i[L_i+1-2E(y_i)]}\right)=\frac{E(\bar{d}_i)-E(y_i)}{\pi_i[L_i+1-2E(y_i)]}=\theta_i$$

因此可以得到敏感性群體比例 θ 的不偏估計量如下：

$$\hat{\theta}_{II}=\sum_{i=1}^{m}w_i\hat{\theta}_i=\sum_{i=1}^{m}w_i\left\{\frac{\bar{d}_i-E(y_i)}{\pi_i[L_i+1-2E(y_i)]}\right\} \tag{28}$$

令 N 爲母體總數，N_i 爲第 i 層的母體人數，則 $w_i=\left(\frac{N_i}{N}\right),\ i=1,2,\cdots,m$，且 $\sum_{i=1}^{m}w_i=1$。

定理 3.2.2：$\hat{\theta}_{II}$ 爲敏感性群體比例 θ 的不偏估計量。

證明：

$$E(\hat{\theta}_{II})=E\left(\sum_{i=1}^{m}w_i\hat{\theta}_{iII}\right)=\sum_{i=1}^{m}w_iE(\hat{\theta}_{iII})=\sum_{i=1}^{m}w_i\theta_i=\theta$$

接著，可得第 i 層敏感性群體比例 θ_i 之不偏估計量 $\hat{\theta}_{iII}$ 的變異數如定理 3.2.3 所示。

定理 3.2.3：估計量 $\hat{\theta}_{iII}$ 的變異數爲

$$Var(\hat{\theta}_{iII})=\frac{1}{n_i}\left\{\frac{\theta_i(1-\pi_i\theta_i)}{\pi_i}+Var(y_i)\{\pi_i[L_i+1-2E(y_i)]\}^{-2}\right\} \tag{29}$$

證明：利用(27)式所得到之不偏估計量 $\hat{\theta}_{iII}$，計算其變異數爲

$$Var(\hat{\theta}_{iII})=Var\left(\frac{\bar{d}_i-E(y_i)}{\pi_i[L_i+1-2E(y_i)]}\right)=\frac{1}{n_i}Var(d_{ij})\{\pi_i[L_i+1-2E(y_i)]\}^{-2}$$
$$=\frac{1}{n_i}\left\{\frac{\theta_i(1-\pi_i\theta_i)}{\pi_i}+Var(y_i)\{\pi_i[L_i+1-2E(y_i)]\}^{-2}\right\}$$

其中，$L_i+1-2E(y_i)\neq 0$ 且 $\pi_i\neq 0$

定理 3.2.4：估計量 $\hat{\theta}_{II}$ 的變異數爲

$$Var(\hat{\theta}_{II})=\sum_{i=1}^{m}\frac{w_i^2}{n_i}\left\{\frac{\theta_i(1-\pi_i\theta_i)}{\pi_i}+Var(y_i)\{\pi_i[L_i+1-2E(y_i)]\}^{-2}\right\} \tag{30}$$

證明：利用(28)式所得到之不偏估計量 $\hat{\theta}_{iII}$ 及定理 3.2.3 之結果，計算 $\hat{\theta}_{II}$ 變異數為

$$
\begin{aligned}
Var(\hat{\theta}_{II}) &= Var\left(\sum_{i=1}^{m} w_i \hat{\theta}_{iII}\right) = \sum_{i=1}^{m} w_i^2 Var\left(\hat{\theta}_{iII}\right) \\
&= \sum_{i=1}^{m} \frac{w_i^2}{n_i}\left\{\frac{\theta_i(1-\pi_i\theta_i)}{\pi_i} + Var(y_i)\{\pi_i[L_i + 1 - 2E(y_i)]\}^{-2}\right\}
\end{aligned}
$$

值得注意的是，當 $\pi_i = 1$ 時，分層一般隨機作答模式之估計量 $\hat{\theta}_{II}$ 與 Christofides (2005) 的估計量 $\hat{\theta}_{CH}$ 相同且其變異數亦相同，因此，Christofides (2005) 的模式亦可視為本文分層一般隨機作答模式之特例。當 $\pi_i = 1$ 且 $m = 1$ 時，分層一般隨機作答模式之估計量 $\hat{\theta}_{II}$ 與 Christofides (2003) 的估計量 $\hat{\theta}_{C}$ 相同且其變異數亦相同，因此，Christofides (2003) 的模式亦可視為本文分層一般隨機作答模式之特例。當 $m = 1$ 時，分層一般隨機作答模式之估計量 $\hat{\theta}_{II}$ 與王智立、蔡宛容（民 96）的估計量 $\hat{\theta}_{WT}$ 相同且其變異數亦相同，因此，王智立、蔡宛容 (民 96) 的模式可視為本文分層一般隨機作答模式之特例。另外，當 $L_i = 2$、$\pi_i = 1$ 且 $m = 1$ 時，分層一般隨機作答模式之估計量 $\hat{\theta}_{II}$ 與 Warner (1965) 的估計量 $\hat{\theta}_{W}$ 相同且其變異數亦相同，因此，Warner (1965) 的模式亦可視為本文分層一般隨機作答模式之特例。

4. 數值模擬與效率比較

在 4.1 節中，將介紹本文分層廣義隨機作答模式與 Christofides (2003) 模式之效率比較。4.2 節為本文分層一般隨機作答模式與 Christofides (2003) 模式之效率比較。而 4.3 節將進行本文所提出的兩個新的分層隨機作答模式之效率比較。

4.1 分層廣義隨機作答模式與 Christofides (2003) 模式之效率比較

為比較上述兩種模式之相對效率，參數需建構在 $\pi_i = 0$ 或 1 時才可進行比較，因此，在不失一般性情況下，吾人考慮在兩分層($m = 2$)、 π_i $(i = 1,2)$ 值皆為 1 且 P_k 值為遞增 $(P_{k+1} = 2P_k, k = 1,\cdots,L)$ 的條件下，吾人選擇 θ_1 與 θ_2 值介在 0.1 到 0.9 之間，而 L 值為 2 至 10 之間的情形。因此定義分層廣義隨機作答模式與 Christofides (2003) 模式之效率比較為

$$
RE_1 = \frac{Var\left(\hat{\theta}_C\right)}{Var\left(\hat{\theta}_I\right)}
$$

表 1 為當 L 在 2 至 10 之間的兩分層廣義隨機作答模式與 Christofides (2003) 模式之效率比較 (因其效率比較對稱於 $\theta_1 = \theta_2 = 0.5$ ，故僅列出部分結果)。

表 1　當 $\pi_i = 1$ 時兩分層廣義隨機作答模式與 Christofides (2003) 模式之效率比較

θ_1	θ_2	w_1	w_2	L	2	3	4	5	6	7	8	9	10
				θ	RE_1	RE_1	RE_1	RE_1	RE_1	RE_1	RE_1	RE_1	RE_1
0.1	0.9	0.9	0.1	0.18	1.2531	1.3060	1.3733	1.4487	1.5256	1.5985	1.6635	1.7193	1.7657
		0.7	0.3	0.34	1.8350	2.0094	2.2315	2.4803	2.7340	2.9742	3.1889	3.3729	3.5260
		0.5	0.5	0.50	2.1531	2.3940	2.7006	3.0443	3.3946	3.7263	4.0228	4.2768	4.4883
		0.3	0.7	0.66	1.8350	2.0094	2.2315	2.4803	2.7340	2.9742	3.1889	3.3729	3.5260
		0.1	0.9	0.82	1.2531	1.3060	1.3733	1.4487	1.5256	1.5985	1.6635	1.7193	1.7657
0.2	0.8	0.9	0.1	0.26	1.2378	1.2643	1.2945	1.3245	1.3515	1.3742	1.3927	1.4072	1.4185
		0.7	0.3	0.38	1.7845	1.8719	1.9716	2.0704	2.1594	2.2346	2.2953	2.3432	2.3804
		0.5	0.5	0.50	2.0833	2.2040	2.3417	2.4782	2.6011	2.7049	2.7888	2.8549	2.9063
		0.3	0.7	0.62	1.7845	1.8719	1.9716	2.0704	2.1594	2.2346	2.2953	2.3432	2.3804
		0.1	0.9	0.74	1.2378	1.2643	1.2945	1.3245	1.3515	1.3742	1.3927	1.4072	1.4185
0.3	0.7	0.9	0.1	0.34	1.2275	1.2383	1.2500	1.2607	1.2698	1.2770	1.2826	1.2869	1.2902
		0.7	0.3	0.42	1.7504	1.7863	1.8246	1.8600	1.8899	1.9138	1.9324	1.9465	1.9572
		0.5	0.5	0.50	2.0362	2.0858	2.1387	2.1876	2.2289	2.2620	2.2876	2.3070	2.3218
		0.3	0.7	0.58	1.7504	1.7863	1.8246	1.8600	1.8899	1.9138	1.9324	1.9465	1.9572
		0.1	0.9	0.66	1.2275	1.2383	1.2500	1.2607	1.2698	1.2770	1.2826	1.2869	1.2902
0.4	0.6	0.9	0.1	0.42	1.2215	1.2241	1.2267	1.2291	1.2311	1.2326	1.2338	1.2346	1.2353
		0.7	0.3	0.46	1.7306	1.7392	1.7480	1.7559	1.7623	1.7673	1.7711	1.7740	1.7761
		0.5	0.5	0.50	2.0089	2.0208	2.0330	2.0438	2.0527	2.0596	2.0649	2.0688	2.0718
		0.3	0.7	0.54	1.7306	1.7392	1.7480	1.7559	1.7623	1.7673	1.7711	1.7740	1.7761
		0.1	0.9	0.58	1.2215	1.2241	1.2267	1.2291	1.2311	1.2326	1.2338	1.2346	1.2353
0.5	0.5	0.9	0.1	0.50	1.2195	1.2195	1.2195	1.2195	1.2195	1.2195	1.2195	1.2195	1.2195
		0.7	0.3	0.50	1.7241	1.7241	1.7241	1.7241	1.7241	1.7241	1.7241	1.7241	1.7241
		0.5	0.5	0.50	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000
		0.3	0.7	0.50	1.7241	1.7241	1.7241	1.7241	1.7241	1.7241	1.7241	1.7241	1.7241
		0.1	0.9	0.50	1.2195	1.2195	1.2195	1.2195	1.2195	1.2195	1.2195	1.2195	1.2195

表 1 爲 θ 值從 0.18 到 0.82，w_1 值從 0.9 到 0.1 且 w_2 值從 0.1 到 0.9 的部份做效率比較。由表中的 RE_1 值可知，隨著 L 值的增加 RE_1 值也跟著增加。此外，兩種模式比較之 RE_1 值呈現對稱的情形，其對稱係以 $w_1 = w_2 = 0.5$ 爲中心。在效率比較的部分，RE_1 值皆大於 1，表示本節所提之分層廣義隨機作答模式確實比 Christofides (2003) 所提出的模式效率更佳。

4.2 分層一般隨機作答模式與 Christofides (2003) 模式之效率比較

考慮在兩分層($m=2$)、 π_i $(i=1,2)$ 值皆爲 1 且 P_k 值爲遞增 $\left(P_{k+1}=2P_k, k=1,\cdots,L\right)$的條件下，吾人選擇$\theta_1$與$\theta_2$值介在 0.1 到 0.9 之間，而 L 值爲 2 至 10 之間的情形。因此定義分層一般隨機作答模式與 Christofides (2003) 模式之效率比較爲

$$RE_2=\frac{Var\left(\hat{\theta}_C\right)}{Var\left(\hat{\theta}_{II}\right)}$$

表 2 爲當 L 在 2 至 10 間的兩分層一般隨機作答模式與 Christofides (2003) 模式之效率比較(其效率比較對稱於$\theta_1=\theta_2=0.5$，故僅列出部分結果)。

表 2　當$\pi_i=1$時，兩分層一般隨機作答模式與 Christofides (2003) 模式之效率比較

θ_1	θ_2	w_1	w_2	L	2	3	4	5	6	7	8	9	10
				θ	RE_2	RE_2	RE_2	RE_2	RE_2	RE_2	RE_2	RE_2	RE_2
0.1	0.9	0.9	0.1	0.18	1.2531	1.3060	1.3733	1.4487	1.5256	1.5985	1.6635	1.7193	1.7657
		0.7	0.3	0.34	1.8350	2.0094	2.2315	2.4803	2.7340	2.9742	3.1889	3.3729	3.5260
		0.5	0.5	0.50	2.1531	2.3940	2.7006	3.0443	3.3946	3.7263	4.0228	4.2768	4.4883
		0.3	0.7	0.66	1.8350	2.0094	2.2315	2.4803	2.7340	2.9742	3.1889	3.3729	3.5260
		0.1	0.9	0.82	1.2531	1.3060	1.3733	1.4487	1.5256	1.5985	1.6635	1.7193	1.7657
0.2	0.8	0.9	0.1	0.26	1.2378	1.2643	1.2945	1.3245	1.3515	1.3742	1.3927	1.4072	1.4185
		0.7	0.3	0.38	1.7845	1.8719	1.9716	2.0704	2.1594	2.2346	2.2953	2.3432	2.3804
		0.5	0.5	0.50	2.0833	2.2040	2.3417	2.4782	2.6011	2.7049	2.7888	2.8549	2.9063
		0.3	0.7	0.62	1.7845	1.8719	1.9716	2.0704	2.1594	2.2346	2.2953	2.3432	2.3804
		0.1	0.9	0.74	1.2378	1.2643	1.2945	1.3245	1.3515	1.3742	1.3927	1.4072	1.4185
0.3	0.7	0.9	0.1	0.34	1.2275	1.2383	1.2500	1.2607	1.2698	1.2770	1.2826	1.2869	1.2902
		0.7	0.3	0.42	1.7504	1.7863	1.8246	1.8600	1.8899	1.9138	1.9324	1.9465	1.9572
		0.5	0.5	0.50	2.0362	2.0858	2.1387	2.1876	2.2289	2.2620	2.2876	2.3070	2.3218
		0.3	0.7	0.58	1.7504	1.7863	1.8246	1.8600	1.8899	1.9138	1.9324	1.9465	1.9572
		0.1	0.9	0.66	1.2275	1.2383	1.2500	1.2607	1.2698	1.2770	1.2826	1.2869	1.2902
0.4	0.6	0.9	0.1	0.42	1.2215	1.2241	1.2267	1.2291	1.2311	1.2326	1.2338	1.2346	1.2353
		0.7	0.3	0.46	1.7306	1.7392	1.7480	1.7559	1.7623	1.7673	1.7711	1.7740	1.7761
		0.5	0.5	0.50	2.0089	2.0208	2.0330	2.0438	2.0527	2.0596	2.0649	2.0688	2.0718
		0.3	0.7	0.54	1.7306	1.7392	1.7480	1.7559	1.7623	1.7673	1.7711	1.7740	1.7761
		0.1	0.9	0.58	1.2215	1.2241	1.2267	1.2291	1.2311	1.2326	1.2338	1.2346	1.2353
0.5	0.5	0.9	0.1	0.50	1.2195	1.2195	1.2195	1.2195	1.2195	1.2195	1.2195	1.2195	1.2195
		0.7	0.3	0.50	1.7241	1.7241	1.7241	1.7241	1.7241	1.7241	1.7241	1.7241	1.7241
		0.5	0.5	0.50	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000
		0.3	0.7	0.50	1.7241	1.7241	1.7241	1.7241	1.7241	1.7241	1.7241	1.7241	1.7241
		0.1	0.9	0.50	1.2195	1.2195	1.2195	1.2195	1.2195	1.2195	1.2195	1.2195	1.2195

表 2 爲 θ 值從 0.18 到 0.82，w_1 值從 0.9 到 0.1 且 w_2 值從 0.1 到 0.9 的部份做效率比較。由表中的 RE_2 值可知，隨著 L 值的增加 RE_2 值也跟著增加。此外，兩種模式比較之 RE_2 值呈現對稱的情形，其對稱係以 $w_1 = w_2 = 0.5$ 爲中心。在效率比較的部分，RE_2 值皆大於 1，表示本節所提之兩分層一般隨機作答模式確實比 Christofides (2003) 所提出的模式效率更佳。

4.3 分層廣義隨機作答模式與分層一般隨機作答模式之效率比較

本節將比較本文所提出的兩種新模式之相對效率，亦即比較在分層抽樣的情形下，分層廣義隨機作答模式與分層一般隨機作答模式之效率比較，因此定義兩模式在母體分爲兩層情形下 $(m=2)$ 其相對效率比較爲

$$RE_3 = \frac{Var\left(\hat{\theta}_I\right)}{Var\left(\hat{\theta}_{II}\right)}$$

由於表格內容過多，且其效率比較在 $L=2$ 到 $L=10$ 之間差異不大的情況下，因此吾人僅摘錄當 $L=2$ 之效率比較如表 3 所示。

由表 3 可看出，當 $L=2$ 、π_i 值大於 0.4 時，RE_3 值皆會大於 1，表示在這樣的條件下，分層一般隨機作答模式的效果會比分層廣義隨機作答模式更好。除此之外，當 π_i 值爲 1 時，分層廣義隨機作答模式之估計量 $\hat{\theta}_I$ 與分層一般隨機作答模式之估計量 $\hat{\theta}_{II}$ 相同且其變異數亦相同，因此其相對效率值皆爲 1。

5. 結論

根據上述的分析結果，加以整理與歸納，可得到下列的結論：

(1) 由於受訪者只需回答出一個「數字」即可，而無需直接回答「是」或「否」，故可降低受訪者的不安全感並減少了不誠實回答的誤差。

(2) 當 $\pi_i = 0$ 或 1 時且 $m=1$，分層廣義隨機作答模式之估計量 $\hat{\theta}_I$ 與 Christofides (2003) 的估計量 $\hat{\theta}_C$ 相同且其變異數亦相同，因此，Christofides (2003) 的模式可視爲分層廣義隨機作答模式之特例。當 $L_i = 2$、$\pi_i = 1$ 且 $m=1$ 時，分層廣義隨機作答模式之估計量 $\hat{\theta}_I$ 與 Warner (1965) 的估計量 $\hat{\theta}_W$ 相同且其變異數亦相同，因此，Warner (1965) 的模式亦可視爲分層廣義隨機作答模式之特例。

(3) 當 $\pi_i = 1$ 且 $m=1$ 時，分層一般隨機作答模式之估計量 $\hat{\theta}_{II}$ 與 Christofides (2003) 的估計量 $\hat{\theta}_C$ 相同且其變異數亦相同，因此，Christofides (2003) 的模式亦可視爲分層一般隨機作答模式之特例。當 $L_i = 2$ 、$\pi_i = 1$ 且 $m=1$ 時，分層一般隨機作答模式之估計量 $\hat{\theta}_{II}$ 與 Warner (1965) 的估計量 $\hat{\theta}_W$ 相同且其變異數亦相同，因此，Warner (1965) 的模式亦可視爲分層一般隨機作

表 **3** 當 $m=2$ 且 $L=2$ 時，兩個新的分層模式之效率比較

θ_1	θ_2	w_1	w_2	π_i	0.1	0.2	0.3	0.4	0.6	0.7	0.8	0.9	1.0
				θ	RE_3	RE_3	RE_3	RE_3	RE_3	RE_3	RE_3	RE_3	RE_3
0.1	0.9	0.9	0.1	0.18	0.0167	0.1205	0.6160	4.3976	9.8081	3.2955	1.8782	1.3050	1.0000
		0.7	0.3	0.34	0.0166	0.1194	0.6088	4.3392	9.6791	3.2576	1.8618	1.2986	1.0000
		0.5	0.5	0.50	0.0164	0.1169	0.5921	4.2046	9.3813	3.1698	1.8234	1.2834	1.0000
		0.3	0.7	0.66	0.0162	0.1145	0.5763	4.0781	9.1014	3.0866	1.7865	1.2685	1.0000
		0.1	0.9	0.82	0.0161	0.1135	0.5700	4.0279	8.9901	3.0534	1.7716	1.2624	1.0000
0.2	0.8	0.9	0.1	0.26	0.0170	0.1208	0.6111	4.3297	9.5938	3.2265	1.8461	1.2916	1.0000
		0.7	0.3	0.38	0.0169	0.1200	0.6058	4.2871	9.5010	3.1994	1.8343	1.2870	1.0000
		0.5	0.5	0.50	0.0167	0.1181	0.5933	4.1879	9.2843	3.1357	1.8065	1.2760	1.0000
		0.3	0.7	0.62	0.0166	0.1163	0.5814	4.0932	9.0774	3.0745	1.7795	1.2652	1.0000
		0.1	0.9	0.74	0.0165	0.1155	0.5766	4.0552	8.9942	3.0499	1.7685	1.2607	1.0000
0.3	0.7	0.9	0.1	0.34	0.0171	0.1207	0.6059	4.2689	9.4173	3.1709	1.8205	1.2811	1.0000
		0.7	0.3	0.42	0.0171	0.1202	0.6024	4.2413	9.3575	3.1534	1.8129	1.2781	1.0000
		0.5	0.5	0.50	0.0170	0.1189	0.5942	4.1761	9.2164	3.1120	1.7948	1.2709	1.0000
		0.3	0.7	0.58	0.0169	0.1177	0.5862	4.1129	9.0795	3.0717	1.7771	1.2639	1.0000
		0.1	0.9	0.66	0.0168	0.1172	0.5830	4.0872	9.0240	3.0553	1.7698	1.2609	1.0000
0.4	0.6	0.9	0.1	0.42	0.0172	0.1203	0.6005	4.2147	9.2746	3.1269	1.8005	1.2730	1.0000
		0.7	0.3	0.46	0.0172	0.1201	0.5988	4.2012	9.2456	3.1184	1.7968	1.2715	1.0000
		0.5	0.5	0.50	0.0171	0.1194	0.5947	4.1690	9.1762	3.0980	1.7880	1.2680	1.0000
		0.3	0.7	0.54	0.0171	0.1188	0.5907	4.1373	9.1079	3.0780	1.7792	1.2645	1.0000
		0.1	0.9	0.58	0.0170	0.1186	0.5890	4.1243	9.0799	3.0698	1.7756	1.2630	1.0000
0.5	0.5	0.9	0.1	0.50	0.0172	0.1196	0.5949	4.1667	9.1629	3.0934	1.7857	1.2670	1.0000
		0.7	0.3	0.50	0.0172	0.1196	0.5949	4.1667	9.1629	3.0934	1.7857	1.2670	1.0000
		0.5	0.5	0.50	0.0172	0.1196	0.5949	4.1667	9.1629	3.0934	1.7857	1.2670	1.0000
		0.3	0.7	0.50	0.0172	0.1196	0.5949	4.1667	9.1629	3.0934	1.7857	1.2670	1.0000
		0.1	0.9	0.50	0.0172	0.1196	0.5949	4.1667	9.1629	3.0934	1.7857	1.2670	1.0000

答模式之特例。

(4) 在效率比較的部分， RE_1 值皆大於 1，表示兩分層之分層廣義隨機作答模式比 Christofides (2003) 所提出的模式效率更佳。而 RE_2 值皆大於 1，表示兩分層的分層一般隨機作答模式比 Christofides (2003) 所提出的模式效率更佳。在不同的參數設計下，由 RE_3 值可以發現分層一般隨機作答模式與分層廣義隨機作答模式等兩模式互有優劣。而當 π_i 值爲 1 時，本文所提出的兩種新隨機作答模式之結果皆爲相同。

參考文獻

王智立、蔡宛容，「應用一般化 Greenberg 無關聯隨機作答模式於敏感性問題之研究」，中國統計學報，第 45 卷第 2 期，民國 96 年，190-207 頁。

Christofides, T. C., “A Generalized Randomized Response Technique,” *Metrika*, Vol. 57, No. 2, 2003, pp. 195-200.

Christofides, T. C., “Randomized Response in Stratified Sampling,” *Journal of Statistical Planning and Inference,* Vol. 128, Iss. 1, 2005, pp. 303-310.

Greenberg, B. G., Abul-Ela, A. A., Simmons, W. R., and Horvitz, D. G., “The Unrelated Question Randomized Response Model：Theoretical Framework,” *Journal of the American Statistical Association*, Vol. 64, No. 326, 1969, pp. 520-539.

Kim, J. M. and Warde, W. D., “A Stratified Warner’s Randomized Response Model,” *Journal of Statistical Planning and Inference*, Vol. 120, Iss. 1-2, 2004, pp. 155-165.

Warner, S. L., “Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias,” *Journal of the American Statistical Association*, Vol. 60, No. 309, 1965, pp. 63-69.