A MAGICAL TALK: ESTIMATING AT LEAST SEVEN MEASURES OF QUALITATIVE VARIABLES FROM A SINGLE SAMPLE USING RANDOMIZED RESPONSE TECHNIQUE

Lee, Cheon-Sig, Sedory, Stephen A., and Singh, Sarjinder
Department of Mathematics
Texas A&M University-Kingsville
Kingsville, TX 78363-8202, USA
Email: kuss2008@tamuk.edu

ABSTRACT

A social scientist could be considered to be a tool-less mechanic if he/she does not have the appropriate statistical tools for collecting, analyzing and interpreting a dataset. Good tools are required for a mechanic to make a good vehicle. In the same way, good statistical tools are required for a social scientist to collect, analyze and interpret a dataset. The dependency of a social scientist on statistical tools is in no way less than the dependency of a mechanic on mechanical tools. A mechanic cannot build a vehicle without mechanical tools. A social scientist cannot build a model of a phenomenon for a society without collecting, analyzing and interpreting the views of persons from the same society in an appropriate way. In this talk, like a magician can show several birds flying out of an empty basket, we shall show that at least seven parameters of interest to a social scientist can be estimated from a single sample and one response from each respondent in the sample. A real survey data application is given.

1. Introduction

Warner (1965) proposed an interviewing technique, called Randomized Response, to protect an interviewee's privacy and to reduce a major source of bias (evasive answers or refusing to respond) in estimating the prevalence of sensitive characteristics in surveys of human populations. Warner (1965) designed a randomization device, for example a spinner or a deck of cards that consists of two mutually exclusive outcomes. In the case of carks, each card has one of the following statements: (i) I possess attribute *A*; (i) I do not possess attribute *A*. The maximum likelihood estimator of π, the proportion of respondents in the population possessing the attribute *A*, is given by:

$$\hat{\pi}_w = \frac{(n_1/n)-(1-P)}{2P-1}\ ,\ P \neq 0.5 \tag{1.1}$$

where n_1 is the number of individuals responding "*yes*", n is the number of respondents selected by a simple random and with replacement sample (SRSWR), and P is the probability of the statement "I possess an attribute *A*". The variance of $\hat{\pi}_w$ is given by:

$$V(\hat{\pi}_w) = \frac{\pi(1-\pi)}{n} + \frac{P(1-P)}{n(2P-1)^2} \tag{1.2}$$

Odumade and Singh (2009) suggested another randomized response model (which we refer to as the *OS* model) using two decks of cards. Each deck of cards, designated Deck-I and Deck-II, is the same as in the Warner's model, but with different probabilities. Under the *OS* model, respondents go through the Warner's model twice for a single attribute.

Christofides (2005) developed a new method to estimate the proportion of individuals having two sensitive characteristics which we describe in detail : Assume that in a population Ω some respondents possess either sensitive attribute A or sensitive attribute B, both A and B or none of these. A pictorial representation of such a population is shown in the Venn diagram in Figure 1.2. Let π_A be the population proportion of the people possessing the sensitive attribute A; π_B be the population proportion of the people possessing the sensitive attribute B; π_{AB} be the population proportion of the people possessing the both sensitive attributes $A \cap B$. Note that $(A \cup B)^c \neq \Phi$ and $(A \cup B)^c \cup (A \cup B) = \Omega$. Christofides (2005) developed estimators for π_A, π_B, π_{AB}, and $\pi_{A|B}$. In this study, we developed two different methods for estimating nine different parameters: π_A, π_B, π_{AB}, $\pi_{A|B}$, ρ_{AB}, π_{A-B}, $\pi_{A\cup B}$, π_d, and $RR(A|B)$, where ρ_{AB} stands for correlation coefficient between two the sensitive characteristics, *RR* stands for the relative risk of *A* given *B*, and other parameters are to be defined below. The methods developed here are simpler and more practical than the Christofides' (2005) model. The first and second proposed methods are named the *Simple Model* and the *Crossed Model*, respectively, in the following sections.

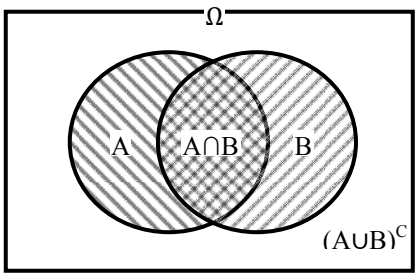

Figure 1.2. Population of interest.

2. Simple Model

We consider selecting a simple random and with replacement sample of *n* respondents from the given population. Two shuffled decks of cards are provided to each respondent in the sample. The decks are marked as Deck-I and Deck-II and each deck is comprised two sorts of cards that indicate whether or not the respondent possesses a particular sensitive characteristic. Two types of cards are present in proportions as shown in Figure 2.1. Each respondent is requested to draw one card from each deck, matches his or her status with the statement on the card drawn from Deck-I and Deck-II, and then reports the result in terms of "Yes or No" without reporting the statement written on the card to the interviewer. The probabilities of (Yes, Yes), (Yes, No), (No, Yes), and (No, No) are denoted as θ_{11}, θ_{10}, θ_{01}, and θ_{00}, respectively. Responses fall into four categories as follows:

$I \in A$ with probability P
$I \in A^c$ with probability $(1-P)$
Deck-I

$I \in B$ with probability T
$I \in B^c$ with probability $(1-T)$
Deck-II

Figure 2.1. Two Decks of Cards

Category 1: (Yes, Yes) response results from four different ways. If the respondent possesses both sensitive characteristics *A* and *B* and draws statements "$I \in A$" and "$I \in B$" from Deck-I and Deck-II, respectively, then the respondent is requested to report (Yes, Yes). If the respondent possesses a sensitive characteristic *A*, but *B*, and draws statements "$I \in A$" and "$I \in B^c$" from each deck of cards, respectively, then the respondent is requested to report (Yes, Yes). If the respondent possesses a sensitive characteristic *B*, but *A*, and draws statements "$I \in A^c$" and "$I \in B$" from each deck of cards, respectively, then the respondent is requested to report (Yes, Yes). If the respondent does not possess either sensitive characteristics *A* and *B* and draws statements "$I \in A^c$" and "$I \in B^c$", then the respondent is requested to report (Yes, Yes). Thus, the response (Yes, Yes) may result whether a respondent belongs to group *A*, A^c, *B*, or B^c and hence, their privacy will be protected. The probability of getting the response (Yes, Yes) is given by:

$$\theta_{11} = (2P-1)(2T-1)\pi_{AB} + (2P-1)(1-T)\pi_A + (1-P)(2T-1)\pi_B + (1-P)(1-T) \quad (2.1)$$

Category 2: In the same manner in the (Yes, Yes) response, the response (Yes, No) may result from four different ways; the probability is given by:

$$\theta_{10} = -(2P-1)(2T-1)\pi_{AB} + (2P-1)T\pi_A + (1-P)(2T-1)\pi_B + (1-P)T \quad (2.2)$$

Category 3: Similarly, the response (No, Yes) results from four different ways and has the probability given by:

$$\theta_{01} = -(2P-1)(2T-1)\pi_{AB} + (2P-1)(1-T)\pi_A + P(2T-1)\pi_B + P(1-T) \tag{2.3}$$

Category 4: Similarly, the response (No, No) results from four different ways, and has the probability given by:

$$\theta_{00} = (2P-1)(2T-1)\pi_{AB} + (2P-1)T\pi_A + P(2T-1)\pi_B + PT \tag{2.4}$$

Responses from n respondents can be classified into the four categories as shown in Table 2.1 and the corresponding true probabilities of these responses are shown in Table 2.2. Note that θ_{11}, θ_{10}, θ_{01}, and θ_{00} are given in the equations (2.1), (2.2), (2.3), and (2.4), respectively and that $\theta_{11}+\theta_{10}+\theta_{01}+\theta_{00}=1$. Note also that $n_{11}+n_{10}+n_{01}+n_{00}=n$. Let $\hat{\theta}_{11}=n_{11}/n$, $\hat{\theta}_{10}=n_{10}/n$, $\hat{\theta}_{01}=n_{01}/n$, and $\hat{\theta}_{00}=n_{00}/n$ be the observed proportions respectively of (Yes, Yes), (Yes, No), (No, Yes), and (No, No) responses.

Responses	Yes	No
Yes	n_{11}	n_{10}
No	n_{01}	n_{00}

Table 2.1. Observed responses

True Probabilities		Deck-II	
		Yes	No
Deck-I	Yes	θ_{11}	θ_{10}
	No	θ_{01}	θ_{00}

Table 2.2. Expected proportions

The least square distance between observed proportions and true proportions is defined as follows:

$$D = \frac{1}{2}\sum_{i=0}^{1}\sum_{j=0}^{1}\left(\theta_{ij} - \hat{\theta}_{ij}\right)^2 \tag{2.5}$$

On setting $\frac{\partial D}{\partial \pi_A} = 0$, $\frac{\partial D}{\partial \pi_B} = 0$, and $\frac{\partial D}{\partial \pi_{AB}} = 0$ to minimize the least squared distance *D*, and by using the method of moments, we have the following theorems:

Theorem 2.1. Unbiased estimators of the population proportions π_A, π_B, and π_{AB} are given by:

$$\hat{\pi}_A = \frac{\hat{\theta}_{11} + \hat{\theta}_{10} - \hat{\theta}_{01} - \hat{\theta}_{00} + (2P-1)}{2(2P-1)} \tag{2.6}$$

$$\hat{\pi}_B = \frac{\hat{\theta}_{11} - \hat{\theta}_{10} + \hat{\theta}_{01} - \hat{\theta}_{00} + (2T-1)}{2(2T-1)} \tag{2.7}$$

and

$$\hat{\pi}_{AB} = \frac{(P+T)\hat{\theta}_{11} + (T-P)\hat{\theta}_{10} + (P-T)\hat{\theta}_{01} + (2-P-T)\hat{\theta}_{00} - T(1-P) - P(1-T)}{2(2P-1)(2T-1)} \tag{2.8}$$

for $T \neq 0.5$ and $P \neq 0.5$

Theorem 2.2. The variance of $\hat{\pi}_A$, $\hat{\pi}_B$, and $\hat{\pi}_{AB}$ are given by:

$$V(\hat{\pi}_A) = \frac{\pi_A(1-\pi_A)}{n} + \frac{P(1-P)}{n(2P-1)^2} \tag{2.9}$$

$$V(\hat{\pi}_B) = \frac{\pi_B(1-\pi_B)}{n} + \frac{T(1-T)}{n(2T-1)^2} \tag{2.10}$$

and

$$V(\hat{\pi}_{AB}) = \frac{\pi_{AB}(1-\pi_{AB})}{n} + \frac{(2P-1)^2T(1-T)\pi_A + P(1-P)(2T-1)^2 + PT(1-P)(1-T)}{n(2P-1)^2(2T-1)^2} \tag{2.11}$$

for $P = T \neq 0.5$

Now, we suggest a natural estimator of the conditional proportion $\pi_{A|B}$ as:

$$\hat{\pi}_{A|B} = \frac{\hat{\pi}_{AB}}{\hat{\pi}_B} \tag{2.12}$$

Then we have the following theorems:

Theorem 2.3. The bias, to the first order of approximation, in the estimator $\hat{\pi}_{A|B}$ is given by:

$$B\left(\hat{\pi}_{A|B}\right)=\pi_{A|B}\left[\frac{V\left(\hat{\pi}_B\right)}{\pi_B^2}-\frac{Cov\left(\hat{\pi}_{AB},\hat{\pi}_B\right)}{\pi_{AB}\pi_B}\right] \tag{2.13}$$

where $Cov\left(\hat{\pi}_{AB},\hat{\pi}_B\right)=\frac{\pi_{AB}(1-\pi_B)}{n}+\frac{T(1-T)\pi_A}{n(2T-1)^2}$

Theorem 2.4. The mean squared error, to the first order of approximation, of the estimator $\hat{\pi}_{A|B}$ is given by:

$$MSE\left(\hat{\pi}_{A|B}\right)=\pi_{A|B}^2\left[\frac{V\left(\hat{\pi}_{AB}\right)}{\pi_{AB}^2}+\frac{V\left(\hat{\pi}_B\right)}{\pi_B^2}-\frac{2Cov\left(\hat{\pi}_{AB},\hat{\pi}_B\right)}{\pi_{AB}\pi_B}\right] \tag{2.14}$$

An unbiased estimator of the proportion of persons possessing exactly one characteristic, say *A* but not *B*, π_{A-B}, is given by

$$\hat{\pi}_{A-B}=\hat{\pi}_A-\hat{\pi}_{AB} \tag{2.15}$$

Theorem 2.5. The variance of the estimator $\hat{\pi}_{A-B}$ is given by

$$V\left(\hat{\pi}_{A-B}\right)=V\left(\hat{\pi}_A\right)+V\left(\hat{\pi}_{AB}\right)-2Cov\left(\hat{\pi}_A,\hat{\pi}_{AB}\right) \tag{2.16}$$

An unbiased estimator of the proportion of persons who possess at least one of the characteristics *A* or *B*, that is, $\hat{\pi}_{A\cup B}$ is given by

$$\hat{\pi}_{A\cup B}=\hat{\pi}_A+\hat{\pi}_B-\hat{\pi}_{AB} \tag{2.17}$$

Theorem 2.6. The variance of the estimator $\hat{\pi}_{A\cup B}$ is given by

$$V\left(\hat{\pi}_{A\cup B}\right)=V\left(\hat{\pi}_A\right)+V\left(\hat{\pi}_B\right)+V\left(\hat{\pi}_{AB}\right)+2Cov\left(\hat{\pi}_A,\hat{\pi}_B\right)-2Cov\left(\hat{\pi}_A,\hat{\pi}_{AB}\right)-2Cov\left(\hat{\pi}_B,\hat{\pi}_{AB}\right) \tag{2.18}$$

where $Cov\left(\hat{\pi}_{AB},\hat{\pi}_A\right)=\frac{\pi_{AB}(1-\pi_A)}{n}+\frac{P(1-P)\pi_B}{n(2P-1)^2}$ and $Cov(\hat{\pi}_A,\hat{\pi}_B)=\frac{\pi_{AB}-\pi_A\pi_B}{n}$

Following Rosner (2006), we define a natural estimator of a relative risk (*RR*) of a respondent belonging to group *B* given that the respondent belongs to group *A* as:

$$\hat{RR}\left(B|A\right)=\frac{\hat{\pi}_{AB}\left(1-\hat{\pi}_A\right)}{\hat{\pi}_A\left(\hat{\pi}_B-\hat{\pi}_{AB}\right)} \tag{2.19}$$

Then we have the following theorems:

Theorem 2.7. The bias in the estimator $\hat{RR}\left(B|A\right)$, to the first order of approximation, is given by:

$$B\left(\hat{RR}\left(B|A\right)\right)=RR\left[\frac{V\left(\hat{\pi}_A\right)}{\pi_A^2\left(1-\pi_A\right)}+\frac{V\left(\hat{\pi}_B\right)}{\left(\pi_B-\pi_{AB}\right)^2}+\frac{\pi_B V\left(\hat{\pi}_{AB}\right)}{\pi_{AB}\left(\pi_B-\pi_{AB}\right)^2}+\frac{Cov\left(\hat{\pi}_A,\hat{\pi}_B\right)}{\pi_A\left(1-\pi_A\right)\left(\pi_B-\pi_{AB}\right)}\right.$$
$$\left.-\frac{\pi_B Cov\left(\hat{\pi}_A,\hat{\pi}_{AB}\right)}{\pi_A\pi_{AB}\left(1-\pi_A\right)\left(\pi_B-\pi_{AB}\right)}-\frac{\left(\pi_B+\pi_{AB}\right)Cov\left(\hat{\pi}_B,\hat{\pi}_{AB}\right)}{\pi_{AB}\left(\pi_B-\pi_{AB}\right)^2}\right] \tag{2.20}$$

Theorem 2.8. The mean squared error of the estimator $\hat{RR}\left(B|A\right)$, to the first order of approximation, is given by:

$$MSE\left(\hat{RR}\left(B|A\right)\right)=RR^2E\left[\frac{\pi_B^2V\left(\hat{\pi}_{AB}\right)}{\pi_{AB}^2\left(\pi_B-\pi_{AB}\right)^2}+\frac{V\left(\hat{\pi}_A\right)}{\pi_A^2\left(1-\pi_A\right)^2}+\frac{V\left(\hat{\pi}_B\right)}{\left(\pi_B-\pi_{AB}\right)^2}+\frac{2\pi_B Cov\left(\hat{\pi}_A,\hat{\pi}_{AB}\right)}{\pi_A\pi_{AB}\left(1-\pi_A\right)\left(\pi_B-\pi_{AB}\right)}\right.$$
$$\left.-\frac{2\pi_B Cov\left(\hat{\pi}_B,\hat{\pi}_{AB}\right)}{\pi_{AB}\left(\pi_B-\pi_{AB}\right)}+\frac{2Cov\left(\hat{\pi}_A,\hat{\pi}_B\right)}{\pi_A\left(1-\pi_A\right)\left(\pi_B-\pi_{AB}\right)}\right] \tag{2.21}$$

Now we consider a usual estimator of the correlation coefficient ρ_{AB} as:

$$\hat{\rho}_{AB} = \frac{\hat{\pi}_{AB} - \hat{\pi}_A \hat{\pi}_B}{\sqrt{\hat{\pi}_A(1-\hat{\pi}_A)}\sqrt{\hat{\pi}_B(1-\hat{\pi}_B)}} \tag{2.22}$$

Then we have the following theorems:

Theorem 2.9. The bias in the estimator $\hat{\rho}_{AB}$, to the first order of approximation, is given by:

$$B(\hat{\rho}_{AB}) = \rho_{AB}\left[F_4\frac{V(\hat{\pi}_A)}{\pi_A^2} + F_5\frac{V(\hat{\pi}_B)}{\pi_B^2} + F_6\frac{Cov(\hat{\pi}_A,\hat{\pi}_B)}{\pi_A\pi_B} + F_7\frac{Cov(\hat{\pi}_A,\hat{\pi}_{AB})}{\pi_A\pi_{AB}} + F_8\frac{Cov(\hat{\pi}_B,\hat{\pi}_{AB})}{\pi_B\pi_{AB}}\right] \tag{2.23}$$

Theorem 2.10. The mean squared error of the estimator $\hat{\rho}_{AB}$, to the first order of approximation, is given by:

$$MSE(\hat{\rho}_{AB}) = \rho_{AB}^2\left[F_1^2\frac{V(\hat{\pi}_{AB})}{\pi_{AB}^2} + F_2^2\frac{V(\hat{\pi}_A)}{\pi_A^2} + F_3^2\frac{V(\hat{\pi}_B)}{\pi_B^2} + \frac{2F_1F_2Cov(\hat{\pi}_A,\hat{\pi}_{AB})}{\pi_A\pi_{AB}} + \frac{2F_1F_3Cov(\hat{\pi}_B,\hat{\pi}_{AB})}{\pi_B\pi_{AB}} + \frac{2F_2F_3Cov(\hat{\pi}_A,\hat{\pi}_B)}{\pi_A\pi_B}\right] \tag{2.24}$$

where $F_1 = \frac{\pi_{AB}}{\pi_{AB}-\pi_A\pi_B}$, $F_2 = \frac{\pi_A\pi_B}{\pi_{AB}-\pi_A\pi_B} + \frac{(1-2\pi_A)}{2(1-\pi_A)}$, $F_3 = \frac{\pi_A\pi_B}{\pi_{AB}-\pi_A\pi_B} + \frac{(1-2\pi_B)}{2(1-\pi_B)}$,

$F_4 = \frac{\pi_A\pi_B(1-2\pi_A)}{2(1-\pi_A)(\pi_{AB}-\pi_A\pi_B)} + \frac{\pi_A}{2(1-\pi_A)} + \frac{3(1-2\pi_A)}{8(1-\pi_A)^2}$, $F_5 = \frac{\pi_A\pi_B(1-2\pi_B)}{2(1-\pi_B)(\pi_{AB}-\pi_A\pi_B)} + \frac{\pi_B}{2(1-\pi_B)} + \frac{3(1-2\pi_B)}{8(1-\pi_B)^2}$,

$F_6 = \frac{1-2\pi_A-2\pi_B+3\pi_A\pi_B}{4(1-\pi_A)(1-\pi_B)} - \frac{\pi_A\pi_B}{\pi_{AB}-\pi_A\pi_B}$, $F_7 = \frac{\pi_{AB}(1-2\pi_A)}{2(1-\pi_A)(\pi_{AB}-\pi_A\pi_B)}$, and $F_8 = \frac{\pi_{AB}(1-2\pi_B)}{2(1-\pi_B)(\pi_{AB}-\pi_A\pi_B)}$

Now we consider an unbiased estimator of the difference between two proportions π_d as:

$$\hat{\pi}_d = \hat{\pi}_A - \hat{\pi}_B \tag{2.25}$$

Then we have the following theorem:

Theorem 2.11. The variance of the estimator $\hat{\pi}_d$ is given by

$$V(\hat{\pi}_d) = V(\hat{\pi}_A) + V(\hat{\pi}_B) - 2Cov(\hat{\pi}_A,\hat{\pi}_B) \tag{2.26}$$

When proportions π_A and π_B are considered independently, variances of $V(\hat{\pi}_A)$ and $V(\hat{\pi}_B)$ remain same as in the Warner (1965). However, the case of independence is of no interest in this study. Thus, the proposed simple model has the advantages of estimating additional parameters, such as $\pi_{A\cap B}$, $\pi_{A|B}$, π_{A-B}, $\pi_{A\cup B}$, $RR(A|B)$, ρ_{AB}, and π_d etc. These make it possible for a social scientist to relate two sensitive characteristics to make a more insightful decision about the prevalence of such characteristics in a society under study. Now a natural question arises: Is it possible to develop a model with two decks of cards which could be more efficient than the Warner model even while estimating only the individual population proportions π_A and π_B as well? This idea motivates to think about a new model in the following section which is labeled as *Crossed Model*.

3. Crossed Model

All assumption and the procedure are the same as in section 2. The method uses two decks of cards as in section 2. The two decks for the *Crossed Model* are shown in Figure 3.1.

$I \in A$ with probability P
$I \in B^c$ with probability $(1-P)$
Deck-I

$I \in B$ with probability T
$I \in A^c$ with probability $(1-T)$
Deck-II

Figure 3.1. Two Decks of Cards

With the crossed model, each response, (Yes, Yes), (Yes, No), (No, Yes), and (No, No), of individuals in population may also occur from four different ways; the probabilities of getting each response are given by:

$$\theta_{11}^* = \pi_{AB}\{PT + (1-P)(1-T)\} - \pi_A(1-P)(1-T) - \pi_B(1-P)(1-T) + (1-P)(1-T) \quad (3.1)$$

$$\theta_{10}^* = -\pi_{AB}\{PT + (1-P)(1-T)\} - \pi_A\{(1-P)T - 1\} - \pi_B(1-P)T + (1-P)T \quad (3.2)$$

$$\theta_{01}^* = -\pi_{AB}\{PT + (1-P)(1-T)\} - \pi_A P(1-T) - \pi_B\{P(1-T) - 1\} + P(1-T) \quad (3.3)$$

and

$$\theta_{00}^* = \pi_{AB}\{PT + (1-P)(1-T)\} - \pi_A PT - \pi_B PT + PT \quad (3.4)$$

Responses from the *n* respondents can be classified into the four categories as before, as shown in Table 3.1, and their corresponding true probabilities of these responses are as show in Table 3.2. θ_{11}^*, θ_{10}^*, θ_{01}^*, and θ_{00}^* are given in the equations (3.1), (3.2), (3.3), and (3.4), respectively. As before, we note that: $\theta_{11}^* + \theta_{10}^* + \theta_{01}^* + \theta_{00}^* = 1$ and $n_{11}^* + n_{10}^* + n_{01}^* + n_{00}^* = n$. Recall that our aim is to estimate the three unknown population proportions π_A, π_B and π_{AB} of the respondents belonging to groups *A*, *B*, and $A \cap B$, respectively. Let $\hat{\theta}_{11}^* = n_{11}^*/n$, $\hat{\theta}_{10}^* = n_{10}^*/n$, $\hat{\theta}_{01}^* = n_{01}^*/n$, and $\hat{\theta}_{00}^* = n_{00}^*/n$ be the observed proportions of (Yes, Yes), (Yes, No), (No, Yes) and (No, No) responses.

Responses	Yes	No
Yes	n_{11}^*	n_{10}^*
No	n_{01}^*	n_{00}^*

Table 3.1. Observed responses

True Probabilities		Deck-II	
		Yes	No
Deck-I	Yes	θ_{11}^*	θ_{10}^*
	No	θ_{01}^*	θ_{00}^*

Table 3.2. Expected proportions

The least square distance between observed proportions and true proportions is defined as follows:

$$D^* = \frac{1}{2}\sum_{i=0}^{1}\sum_{j=0}^{1}(\theta_{ij}^* - \hat{\theta}_{ij}^*)^2 \quad (3.5)$$

On setting $\frac{\partial D^*}{\partial \pi_A} = 0$, $\frac{\partial D^*}{\partial \pi_B} = 0$, and $\frac{\partial D^*}{\partial \pi_{AB}} = 0$ to minimize the least squared distance *D*, and by using the method of moments we have theorems as follows:

Theorem 3.1. Unbiased estimators of the population proportions π_A, π_B, and π_{AB} are given by:

$$\hat{\pi}_A^* = \frac{1}{2} + \frac{(T-P+1)\left(\hat{\theta}_{11}^* - \hat{\theta}_{00}^*\right) + (P+T-1)\left(\hat{\theta}_{10}^* - \hat{\theta}_{01}^*\right)}{2(P+T-1)} \quad (3.6)$$

$$\hat{\pi}_B^* = \frac{1}{2} + \frac{(P-T+1)\left(\hat{\theta}_{11}^* - \hat{\theta}_{00}^*\right) + (P+T-1)\left(\hat{\theta}_{01}^* - \hat{\theta}_{10}^*\right)}{2(P+T-1)} \quad (3.7)$$

and

$$\hat{\pi}_{AB}^* = \frac{PT\hat{\theta}_{11}^* - (1-P)(1-T)\hat{\theta}_{00}^*}{\{PT + (1-P)(1-T)\}(P+T-1)} \quad (3.8)$$

for $P + T \neq 1$.

Theorem 3.2. The variances of $\hat{\pi}_A^*$, $\hat{\pi}_B^*$, and $\hat{\pi}_{AB}^*$ are given by:

$$V(\hat{\pi}_A^*) = \frac{\pi_A(1-\pi_A)}{n} + \frac{(1-P)\left[T\{PT + (1-P)(1-T)\}\left(1-\pi_A-\pi_B+2\pi_{AB}\right)\right]}{n(P+T-1)^2} \quad (3.9)$$

$$V(\hat{\pi}_B^*) = \frac{\pi_B(1-\pi_B)}{n} + \frac{(1-T)\left[P\{PT + (1-P)(1-T)\}\left(1-\pi_A-\pi_B+2\pi_{AB}\right)\right]}{n(P+T-1)^2} \quad (3.10)$$

and

$$V\left(\hat{\pi}^*_{AB}\right)=\frac{\pi_{AB}\left(1-\pi_{AB}\right)}{n}+\frac{1}{n\{PT+(1-P)(1-T)\}(P+T-1)^2}$$
$$\times\left[\pi_{AB}\left\{P^2T^2+(1-P)^2(1-T)^2-\{PT+(1-P)(1-T)\}(P+T-1)^2\right\}\right.$$
$$\left.+PT(1-P)(1-T)(P+T-1)(1-\pi_A-\pi_B)\right] \tag{3.11}$$

The natural estimator of the conditional proportion $\pi_{A|B}$ is given by:

$$\hat{\pi}^*_{A|B}=\frac{\hat{\pi}^*_{AB}}{\hat{\pi}^*_B} \tag{3.12}$$

Now we have the following theorems:

Theorem 3.3. The bias, to the first order of approximation, in the estimator $\hat{\pi}^*_{A|B}$ is give by:

$$B\left(\hat{\pi}^*_{A|B}\right)=\pi_{A|B}\left[\frac{V(\hat{\pi}^*_B)}{\pi_B^2}-\frac{Cov(\hat{\pi}^*_{AB},\hat{\pi}^*_B)}{\pi_{AB}\pi_B}\right] \tag{3.13}$$

where $Cov\left(\hat{\pi}^*_{AB},\hat{\pi}^*_B\right)=\frac{\pi_{AB}\left(1-\pi_B\right)}{n}+\frac{\pi_{AB}P(1-T)(T-P+1)}{n(P+T-1)^2}+\frac{PT(1-P)(1-T)(P-T+1)\left(1-\pi_A-\pi_B\right)}{n\{PT+(1-P)(1-T)\}(P+T-1)^2}$

Theorem 3.4. The mean squared error, to the first order of approximation, of the estimator $\hat{\pi}^*_{A|B}$ is given by

$$MSE\left(\hat{\pi}^*_{A|B}\right)=\pi_{AB}{}^2\left[\frac{V(\hat{\pi}^*_{AB})}{\pi_{AB}^2}+\frac{V(\hat{\pi}^*_B)}{\pi_B^2}-\frac{2Cov(\hat{\pi}^*_B,\hat{\pi}^*_{AB})}{\pi_B\pi_{AB}}\right] \tag{3.14}$$

An unbiased estimator of the proportion of persons possessing exactly one characteristic, say *A* but not *B*, π_{A-B}, is given by

$$\hat{\pi}^*_{A-B}=\hat{\pi}^*_A-\hat{\pi}^*_{AB} \tag{3.15}$$

Theorem 3.5. The variance of the estimator $\hat{\pi}^*_{A-B}$ is given by

$$V\left(\hat{\pi}^*_{A-B}\right)=V\left(\hat{\pi}^*_A\right)+V\left(\hat{\pi}^*_{AB}\right)-2Cov\left(\hat{\pi}^*_A,\hat{\pi}^*_{AB}\right) \tag{3.16}$$

An unbiased estimator of the proportion of persons in the population who possess at least one of the characteristics *A* or *B*, that is, $\hat{\pi}_{A\cup B}$ is given by

$$\hat{\pi}^*_{A\cup B}=\hat{\pi}^*_A+\hat{\pi}^*_B-\hat{\pi}^*_{AB} \tag{3.17}$$

Theorem 3.6. The variance of the estimator $\hat{\pi}^*_{A\cup B}$ is given by

$$V\left(\hat{\pi}^*_{A\cup B}\right)=V\left(\hat{\pi}^*_A\right)+V\left(\hat{\pi}^*_B\right)+V\left(\hat{\pi}^*_{AB}\right)+2Cov\left(\hat{\pi}^*_A,\hat{\pi}^*_B\right)-2Cov\left(\hat{\pi}^*_A,\hat{\pi}^*_{AB}\right)-2Cov\left(\hat{\pi}^*_B,\hat{\pi}^*_{AB}\right) \tag{3.18}$$

Now, we define an estimator of a relative risk (*RR*)of a respondent belonging to group *B* given that the respondent belongs to group *A* as:

$$\hat{RR}^*\left(B|A\right)=\frac{\hat{\pi}^*_{AB}\left(1-\hat{\pi}^*_A\right)}{\hat{\pi}^*_A\left(\hat{\pi}^*_B-\hat{\pi}^*_{AB}\right)} \tag{3.19}$$

Then we have the following theorems:

Theorem 3.7. The bias in the estimator $\hat{RR}^*\left(B|A\right)$, to the first order of approximation, is given by:

$$B\left(\hat{RR}^*\left(B|A\right)\right)=RR\left[\frac{V\left(\hat{\pi}^*_A\right)}{\pi_A^2(1-\pi_A)}+\frac{V\left(\hat{\pi}^*_B\right)}{(\pi_B-\pi_{AB})^2}+\frac{\pi_B V\left(\hat{\pi}^*_{AB}\right)}{\pi_{AB}(\pi_B-\pi_{AB})^2}+\frac{Cov\left(\hat{\pi}^*_A,\hat{\pi}^*_B\right)}{\pi_A(1-\pi_A)(\pi_B-\pi_{AB})}\right.$$
$$\left.-\frac{\pi_B Cov\left(\hat{\pi}^*_A,\hat{\pi}^*_{AB}\right)}{\pi_A\pi_{AB}(1-\pi_A)(\pi_B-\pi_{AB})}-\frac{(\pi_B+\pi_{AB})Cov\left(\hat{\pi}^*_B,\hat{\pi}^*_{AB}\right)}{\pi_{AB}(\pi_B-\pi_{AB})^2}\right] \tag{3.20}$$

where $Cov\left(\hat{\pi}^*_{AB},\hat{\pi}^*_{A}\right)=\dfrac{\pi_{AB}\left(1-\pi_A\right)}{n}+\dfrac{\pi_{AB}T\left(1-P\right)\left(P-T+1\right)}{n\left(P+T-1\right)^2}+\dfrac{PT\left(1-P\right)\left(1-T\right)\left(T-P+1\right)\left(1-\pi_A-\pi_B\right)}{n\left\{PT+(1-P)(1-T)\right\}\left(P+T-1\right)^2}$ and

$$Cov\left(\hat{\pi}^*_{A},\hat{\pi}^*_{B}\right)=\frac{\pi_A\left(1-\pi_A\right)-\pi_{AB}\left\{PT+\left(1-P\right)\left(1-T\right)\right\}}{n}-\frac{\left\{PT+\left(1-P\right)\left(1-T\right)+\left(P-T\right)\left(1-2\pi_A\right)\right\}\left(1-\pi_A-\pi_B\right)}{2n}$$
$$+\frac{\left\{PT+\left(1-P\right)\left(1-T\right)\right\}^2\left\{1-\pi_A-\pi_B+2\pi_{AB}\right\}}{2n\left(P+T-1\right)^2}$$

Theorem 3.8. The mean squared error of the estimator $\hat{RR}^*\left(B|A\right)$, to the first order of approximation, is given by:

$$MSE\left(\hat{RR}^*\left(B|A\right)\right)=RR^2E\left[\frac{\pi_B^2V\left(\hat{\pi}^*_{AB}\right)}{\pi^2_{AB}\left(\pi_B-\pi_{AB}\right)^2}+\frac{V\left(\hat{\pi}^*_{A}\right)}{\pi^2_{A}\left(1-\pi_A\right)^2}+\frac{V\left(\hat{\pi}^*_{B}\right)}{\left(\pi_B-\pi_{AB}\right)^2}+\frac{2\pi_BCov\left(\hat{\pi}^*_{A},\hat{\pi}^*_{AB}\right)}{\pi_A\pi_{AB}\left(1-\pi_A\right)\left(\pi_B-\pi_{AB}\right)}\right.$$
$$\left.-\frac{2\pi_BCov\left(\hat{\pi}^*_{B},\hat{\pi}^*_{AB}\right)}{\pi_{AB}\left(\pi_B-\pi_{AB}\right)}+\frac{2Cov\left(\hat{\pi}^*_{A},\hat{\pi}^*_{B}\right)}{\pi_A\left(1-\pi_A\right)\left(\pi_B-\pi_{AB}\right)}\right] \tag{3.21}$$

The usual estimator of the correlation coefficient between two sensitive characteristics, ρ_{AB},is given by:

$$\hat{\rho}^*_{AB}=\frac{\hat{\pi}^*_{AB}-\hat{\pi}^*_A\hat{\pi}^*_B}{\sqrt{\hat{\pi}^*_A(1-\hat{\pi}^*_A)}\sqrt{\hat{\pi}^*_B(1-\hat{\pi}^*_B)}} \tag{3.22}$$

Then we have the following theorems:

Theorem 3.9. The bias in the estimator $\hat{\rho}^*_{AB}$, to the first order of approximation, is given by:

$$B\left(\hat{\rho}^*_{AB}\right)=\rho_{AB}\left[F_4\frac{V\left(\hat{\pi}^*_{A}\right)}{\pi^2_{A}}+F_5\frac{V\left(\hat{\pi}^*_{B}\right)}{\pi^2_{B}}+F_6\frac{Cov\left(\hat{\pi}^*_{A},\hat{\pi}^*_{B}\right)}{\pi_A\pi_B}+F_7\frac{Cov\left(\hat{\pi}^*_{A},\hat{\pi}^*_{AB}\right)}{\pi_A\pi_{AB}}+F_8\frac{Cov\left(\hat{\pi}^*_{B},\hat{\pi}^*_{AB}\right)}{\pi_B\pi_{AB}}\right] \tag{3.23}$$

Theorem 3.10. The mean squared error of the estimator $\hat{\rho}^*_{AB}$, to the first order of approximation, is given by:

$$MSE(\hat{\rho}^*_{AB})=\rho^2_{AB}\left[F_1{}^2\frac{V\left(\hat{\pi}^*_{AB}\right)}{\pi^2_{AB}}+F_2^2\frac{V\left(\hat{\pi}^*_{A}\right)}{\pi^2_{A}}+F_3^2\frac{V\left(\hat{\pi}^*_{B}\right)}{\pi^2_{B}}+\frac{2F_1F_2Cov\left(\hat{\pi}^*_{A},\hat{\pi}^*_{AB}\right)}{\pi_A\pi_{AB}}\right.$$
$$\left.+\frac{2F_1F_3Cov\left(\pi^*_{B},\hat{\pi}^*_{AB}\right)}{\pi_B\pi_{AB}}+\frac{2F_2F_3Cov\left(\hat{\pi}^*_{A},\hat{\pi}^*_{B}\right)}{\pi_A\pi_B}\right] \tag{3.24}$$

Now we consider an unbiased estimator of the difference between two proportions π_d as:

$$\hat{\pi}^*_d=\hat{\pi}^*_A-\hat{\pi}^*_B \tag{3.25}$$

Then we have the following theorem:

Theorem 3.11. The variance of the estimator $\hat{\pi}^*_d$ is given by

$$V\left(\hat{\pi}^*_d\right)=V\left(\hat{\pi}^*_A\right)+V\left(\hat{\pi}^*_B\right)-2Cov\left(\hat{\pi}^*_A,\hat{\pi}^*_B\right) \tag{3.26}$$

Although we tried to compare the variances of the proposed estimators of the different parameters analytically to develop some theoretical conditions for efficiency of one estimator over another, the expressions are too complicated to reach at any conclusion. Thus, in the next section, we compare the estimators of different parameters through numerical illustrations to suggest the usefulness of the proposed models based on the performance of the estimators in different situations.

4. Comparison of Two Types of Models

We define the relative efficiency of the proposed estimators $\hat{\pi}_A^*$, $\hat{\pi}_B^*$ and $\hat{\pi}_{AB}^*$ with respect to the estimators $\hat{\pi}_A$, $\hat{\pi}_B$ and $\hat{\pi}_{AB}$, respectively, as:

$$RE(\hat{\pi}_A^*,\hat{\pi}_A)=\frac{V(\hat{\pi}_A)}{V(\hat{\pi}_A^*)}\times 100\%,\ RE(\hat{\pi}_B^*,\hat{\pi}_B)=\frac{V(\hat{\pi}_B)}{V(\hat{\pi}_B^*)}\times 100\%,\ \text{and}\ RE(\hat{\pi}_{AB}^*,\hat{\pi}_{AB})=\frac{V(\hat{\pi}_{AB})}{V(\hat{\pi}_{AB}^*)}\times 100\%)$$

We used a FORTRAN program to find values of the percent relative efficiency for different choice of the parameters. We decided to keep the choice of $P=0.7$ and $T=0.7$ in both the simple and crossed model. The values of π_{AB} (which we generally expect small in a real survey) was fixed at 0.05, 0.1 and 0.2, and the values of π_A and π_B were changed from 0.01 to 0.99 with a step of 0.01. Note that the relative efficiency expressions are free from the sample size, but in the FORTRAN code conditions were imposed that $\pi_{AB}\le\pi_A$, $\pi_{AB}\le\pi_B$, and $\pi_A+\pi_B<0.99$. Graphical representation of the percent relative efficiencies is given in Figures 4.1 through Figure 4.3 for $\pi_{AB}=0.05$, and similar results are observed for π_{AB} equal to 0.1 and 0.2.

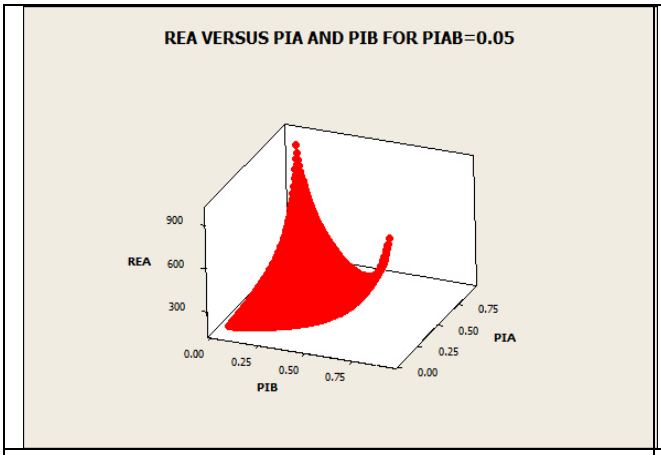

Figure 4.1. *Percent relative efficiency of the estimator* $\hat{\pi}_A^*$ *with respect to the estimator* $\hat{\pi}_A$ *for* $\pi_{AB}=0.05$.

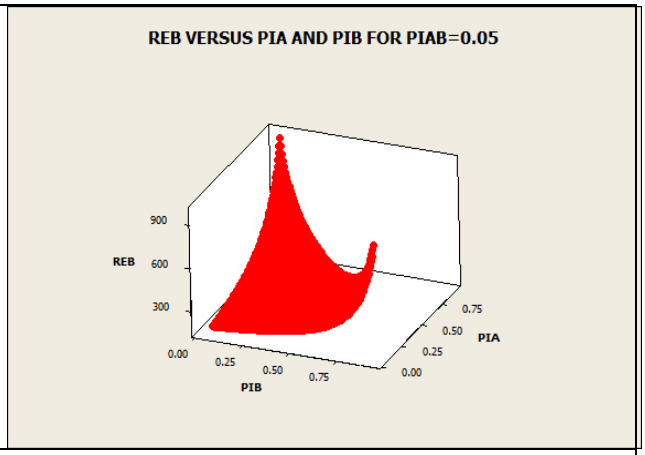

Figure 4.2. *Percent relative efficiency of the estimator* $\hat{\pi}_B^*$ *with respect to the estimator* $\hat{\pi}_B$ *for* $\pi_{AB}=0.05$.

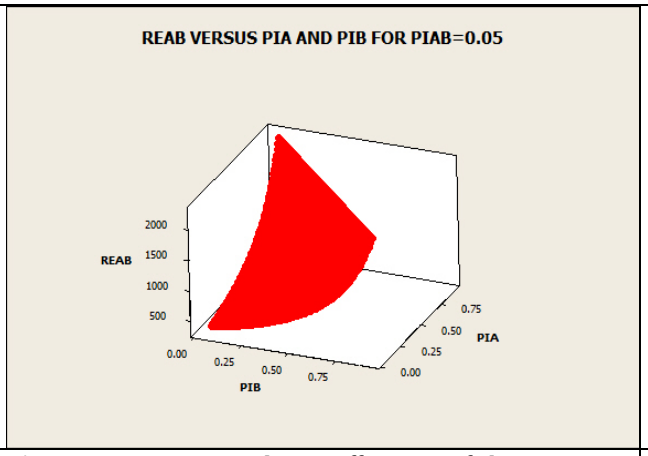

Figure 4.3. *Percent relative efficiency of the estimator* $\hat{\pi}_{AB}^*$ *with respect to the estimator* $\hat{\pi}_{AB}$ *for* $\pi_{AB}=0.05$.

Figure. 4.4. Sarjinder Singh (Left) with a deck of cards and a conference attendee (Right) drawing one card from the deck at the booth STAT-HAWKERS at the Joint Statistical Meeting, Miami Beach, FL.

Figures 4.1-4.3 show that the estimators $\hat{\pi}_A^*$, $\hat{\pi}_B^*$ and $\hat{\pi}_{AB}^*$ remain, respectively, more efficient than the estimators $\hat{\pi}_A$, $\hat{\pi}_B$ and $\hat{\pi}_{AB}$. In the same way, the estimators of the other parameters considered and obtained from the crossed model were found to perform better, from the relative efficiency point of views, than those obtained from the simple model. In short, we found that the proposed crossed model is always more efficient than the proposed simple model for all nine different parameters. Details of these results are available in Lee (2011).

5. Survey Data Application

Sarjinder Singh organized a booth, STAT-HAWKERS, at the Joint Statistical Meeting, Miami Beach, FL during July 31 to Aug 3, 2011, to promote his research among the distinguished statisticians attending the conference (Figure 4.4). At the booth he displayed the 'simple model' and the 'crossed model' using a poster. The problem of estimation of proportions of smokers, drinkers and both was considered using the proposed crossed model. He made two decks of cards: Deck-I, a green deck of cards and Deck-II, a pink deck of cards. Two types of cards bearing two different statements made up the green deck of cards: 56 cards with the statement, "I consider myself a smoker" and 24 cards with the statement, "I do not consider myself a drinker." Two types of cards bearing two different statements made up the pink deck of cards: 56 cards with statement, "I consider myself a drinker" and 24 cards with the statement, "I do not consider myself a smoker." During the three days, a total of 75 conference attendees participated in the survey. The respondents took an interest after being assured of their anonymity. The respondents were cooperative and smiling while drawing cards. Many participants also told that they felt like they were playing a card game. A two-way classification of 75 responses is given in the Table 5.1. By using the proposed crossed model estimators, the estimate of proportion of smokers is 0.240, that of drinkers is 0.360, and that of smokers as well as drinkers is 0.237. It seems that a smoker is likely to be a drinker, but a drinker may not be a smoker. The estimate of correlation between smoking and drinking attitude is 0.733569. The estimate of the relative risk of a drinker to be a smoker is 140.44, which means a smoker is 140.44 times as likely to be a drinker than a non-user of both; whereas the estimate of the relative risk of a smoker to be a drinker is 6.10, which means a drinker is 6.10 times as likely to be a smoker than a non-user of both. This study shows that 63.7% among the conference attendees had neither a drinking nor a smoking habit.

Table 5.1: Responses from the survey.

	Pink Deck-II	
Green Deck-I	**Yes**	**No**
Yes	13	14
No	23	25

6. Discussion

In conclusion, we have created new and more efficient estimators of proportions of people possessing sensitive characteristics in a population and in the process have magically produced seven additional estimators of parameters involving the relationships between the two sensitive characteristics.

Acknowledgement

This work is partially supported by Ministerio de Educación y Cienicia
(contract No. MTM2009-10055)

REFERENCES

Chaudhuri, A. (2011). *Randomized Response and Indirect Questioning Techniques.* Chapman & Hall, CRC, Taylor & Francis Group, FL (USA)

Christofides, T.C. (2005). Randomized response technique for two sensitive characteristics at the same time. *Metrika*, 62, 53-63.

Lee, Cheon-Sig (2011). *Estimation of Parameters for Two sensitive Characteristics Using Two Decks of Cards*. Unpublished master thesis submitted to Department of Mathematics, Texas A&M University-Kingsville.

Rosner, B. (2006). *Fundamentals of Biostatistics.* 6th ed. Thompson: Brooks/Cole, Belmont, CA.

Odumade, O. and Singh, S. (2009). Efficient use of two decks of cards in randomized response sampling. *Commun. Statist.-Theory Meth.* 38, 439-446.

Warner, S.L. (1965): Randomized response: a survey technique for eliminating evasive answer bias. *J. Amer. Stat. Assoc.*, 60, 63-69.